ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 36, Number 1, 2006

APPROXIMATION OF INVARIANT MEASURES FOR RANDOM ITERATIONS

ANDERS ÖBERG

ABSTRACT. In this paper iterated function systems are investigated from the point of view of their invariant measures. Different ideas of how to approximate invariant measures are investigated, but we also discuss necessary and sufficient conditions for uniqueness of an invariant measure. We also consider sufficient conditions for the measure separated open set condition under weak assumptions.

1. Introduction. In this paper we generalize some results obtained in Strichartz et al. [**33**] from the interval to compact subsets of Euclidean spaces. In [**33**], they managed to approximate invariant measures associated to iterated function systems with finitely many maps S_j, that is measures which satisfy an identity of the form

$$\mu = \sum_{j=1}^{m} p_j \mu \circ S_j^{-1}, \quad m \geq 2. \tag{1.1}$$

In the part of [**33**] we are presently interested in, one has one-to-one maps $S_j : [0,1] \to [0,1]$ with strictly positive continuous probability weight functions p_j. In [**33**] a *non-overlapping condition* was also imposed, which was the key to their successful *approximation algorithm* and was also sufficient to establish *uniqueness* of an invariant measure of the form (1.1). If the state space is X, this non-overlapping condition means that the interiors of the images $S_j X$ are disjoint and if $X = K$, a compact subset of a Euclidean space, the so-called *open set condition*, see Hutchinson [**15**], is satisfied with the interior of K as the open set.

A simple condition for uniqueness of an invariant measure is strict contraction of the maps, see for instance [**15**, Theorem 1, p. 733]. There

2000 AMS *Mathematics Subject Classification.* Primary 60B10, Secondary 28A80, 37A30, 60J05.
Key words and phrases. Iterated function system, uniqueness of invariant measure, open set condition.
Received by the editors on July 12, 2003.

are weaker conditions, such as the average contraction conditions of [**3**, Theorem 1, p. 20]. The non-overlapping condition applies in some situations when average contraction, even in a geometric mean sense, is difficult to verify or not even present, for instance, if one map on the unit interval is $S_1(x) = \sqrt{x}/2$ and the other map is $S_2(x) = 1 - \sqrt{x}/2$. For further examples, see [**33**, p. 112]. There they treat inverse branches of polynomial maps $P(z) = z^2 - a$, for a real and $a \geq 2$. In this range the Julia sets lies in the real axis and is the attractor for the IFS given by the two maps $\pm\sqrt{x+a}$. The attractor is included in the interval $[-b, b]$, where b is the larger root of the equation $b^2 = a + b$.

Our goal in this paper was to investigate what could be done in a higher dimensional context using the non-overlapping condition, which is a natural condition when one is interested in iterating the inverse branches of a map.

In a one-dimensional context, the non-overlapping condition was treated also in Dubins and Freedman (1966) in [**10**]; see in particular Theorem 5.4 and its corollary.

Here we also investigate in some detail why, in the context of continuous one-to-one maps, the non-overlapping condition works as a uniqueness condition in the one-dimensional case (Theorem 2, Remark 8, Subsection 4.3), and why it is not sufficient in higher dimensions (Subsection 4.1). In Subsection 4.3 we consider a condition for uniqueness which is weaker than the non-overlapping condition, and which can be used in higher dimensions. This is done in connection with estimates in probability metrics. The Wasserstein metric estimate is known, but in this paper we also consider an approximation in the L^1-metric as a generalization from one to two (and higher) dimensions. In one dimension the Wasserstein metric on probability measures and the L^1-metric for the corresponding distribution functions are equal (not only equivalent), but in two (and higher) dimensions we have complications concerning the definition of an L^1-metric (Subsection 4.2) and we also obtain a different type of estimate (Theorem 3), once a reasonable definition has been made.

We also discuss the case when the weights in (1.1) are not constants, or even probabilities. This means that we have to replace (1.1) with the following identity, for all Borel sets A:

$$\lambda\mu(A) = \sum_{j=1}^{m} \int_{S_j^{-1}A} q_j(x)\, d\mu(x), \quad \lambda > 0, \tag{1.2}$$

which we write shorter as

$$\lambda\mu = \sum_{j=1}^{m} (q_j\mu) \circ S_j^{-1}, \tag{1.3}$$

where we write p_j instead of q_j, whenever $\sum_{j=1}^{m} p_j(x) = 1$ for all x (then we also have $\lambda = 1$). In such a *probabilistic* case we prove (Theorem 1) that the support of all μ satisfying (1.2) is uniquely defined by the IFS parameters (by Stenflo [**29**, Theorem 1], there can be more than one μ), if the maps S_j are continuous and one-to-one on $[0, 1]$. We also give a necessary condition for uniqueness (Proposition 1). The reader is referred to Johansson and Öberg [**18**] and, in the case of probabilistic weights, Stenflo [**31**], for some recent contributions to uniqueness of μ of the form (1.2).

Our main objective is to suggest how to obtain a 'local' approximation of a unique invariant measure (Theorem 2 and Proposition 2) in two and higher dimensions. This can be regarded as an extension of the local approximation in Strichartz et al. [**33**, in particular Algorithm 2.2 and Theorem 2.5, pp. 106–107], where the supremum (Kolmogorov) metric on the distribution functions was used. The supremum metric cannot be used in higher dimensions in this context, for reasons explained in Subsection 4.2. Instead, we estimate μ in (1.1) or (1.2) on small sets (the diameter tends to zero). When the weights are constant probabilities we prove, following either Lemmas 4, 5 or 6, that $\mu_n(A) = \mu(A)$, where A is a set of the form $S_{i_n} \circ \cdots \circ S_{i_1} K$, K a compact subset of $\mathbf{R}^2$, and μ_n the uniform distribution on A of mass $p_{i_1} \cdots p_{i_n}$. (This immediately gives the supremum metric estimate in one dimension.) In Lemma 5, which depends on Lemma 4, and Lemma 6, which does not depend on Lemma 4, we prove the *measure separated open set condition*, i.e., $\mu(S_j K \cap S_k K) = \varnothing$, for $j \neq k$, see Strichartz [**32**]. Lemmas 3–5 are extensions/elaborations of results of Lau and Wang [**20**], in particular of Theorem 2.3. Lemma 6 is our own and it relies on the Jordan curve theorem and is therefore two-dimensional in its nature; it requires that ∂K is a Jordan curve and that the maps do not have common points of prime period 1 or 2.

We extend this local approximation idea to the case of variable weights (Proposition 2) giving a sufficient condition for this approximation to work. This condition, 'asymptotic matching of set partition data,' is quite indirect and it would require further research to obtain predictable criteria.

We end this paper (Subsection 4.3) by comparing with other uniqueness conditions. In Example 3 we give an example where we have a unique invariant measure without having $\mathrm{diam}(S_{i_n} \circ \cdots \circ S_{i_1}[0,1]) \to 0$ for any sequence of maps.

This paper was the first half of the author's thesis [**23**] and has been referred to under this title (in Słomczyński et al. [**27**], Öberg et al. [**24**]) and also under its preprint coordinates: "U.U.D.M. Rep. 6, Dept. of Mathematics, Uppsala University, 1997" (in Fan and Lau [**13**], Lau [**19**], Stenflo [**28**], Stenflo [**30**]).

2. Notation and preliminaries. Throughout this paper X will denote a compact metric space, and if nothing else is stated we are considering a compact metric space X as the ambient space. K denotes a compact subset of $\mathbf{R}^2$. For a set A, A° is the interior of A, ∂A the boundary of A, $\overline{A}$ the closure of A and A^c the complement of A. The underlying metric on X and K will be denoted by d and on K it is assumed that d is the Euclidean metric if nothing else is mentioned. For a set A we will by $\mathrm{diam}\,(A)$ mean $\sup_{x,y\in A} d(x,y)$. We will assume that the probability weight functions are strictly positive and continuous; they are not assumed to be constants unless explicitly stated, but sometimes we have emphasized that the weight functions may be place-dependent by writing $(p_j\mu)$ instead of $p_j\mu$ in the measure-invariance identity. Also, the non-overlapping condition is never assumed without so saying. The maps S_j are always assumed to be continuous and it is tacitly assumed that there are at least two of them (often we denote the number of maps with the letter m). On Euclidean spaces, F_μ denotes the distribution function of the probability measure μ.

We have borrowed the formalism presented below from Barnsley et al. [**2**]. Invariant measures of IFS's can be interpreted as stationary measures to a discrete time Markov process, which is a sequence of random variables $\{Z_n\}$ on a probability space $(\Omega, \mathcal{F}, P)$ with values in $(X, \mathcal{B})$ (where $\mathcal{B}$ are the Borel subsets of the state space X) satisfying

$$P(Z_{n+1} \in A \mid Z_n, \dots, Z_0) = P(Z_{n+1} \in A \mid Z_n) \quad \text{a.s.},$$

for all $n \geq 0$ and all $A \in \mathcal{B}$.

A transition probability function $p(x, B)$, or a *stochastic kernel*, is a version of $P(Z_{n+1} \in B \mid Z_n = x)$ for all Borel sets B (we are just considering cases when $p(x, B)$ is independent of n), and is defined for a given x and a given Borel subset $B \subseteq K$ by

$$p(x, B) = \sum_{j=1}^{m} p_j(x) \delta_{S_j(x)}(B) = \sum_{j=1}^{m} p_j(x) 1_B(S_j(x))$$

where δ_x denotes the Dirac measure concentrated at x and 1_B denotes the indicator function of B.

We will make use of an operator $T^\star : \mathcal{P}(X) \to \mathcal{P}(X)$. This operator is the restriction to $\mathcal{P}(X)$ of the operator $T^\star$ taking $\mathcal{M}(X)$, the space of finite signed Borel measures on X, into itself, and this is the adjoint operator to the Markov operator T defined by

$$(Tf)(x) = \int f(y) p(x, dy) = \sum_{j=1}^{m} p_j(x) f(S_j(x))$$

for $f \in C(X)$.

Let $\nu \in \mathcal{P}(X)$, and let $\{Z_n^\nu,\ n = 0, 1, 2, \dots\}$ be the Markov process having initial distribution ν. The operator $T^\star$ describes how the probability distribution ν on X is transformed in one step of the Markov process $\{Z_n^\nu,\ n = 0, 1, 2, \dots\}$, i.e., $(T^\star \nu)(B) = P(Z_1^\nu \in B)$, where B is a Borel subset of X. It is also easy to see that our operator $T^\star : \mathcal{P}(X) \to \mathcal{P}(X)$ is given by

$$(T^\star \mu)(B) = \int p(x, B)\, \mu(dx) = \sum_{j=1}^{m} (p_j \mu)(S_j^{-1}(B)).$$

From this it is easy to see that an invariant measure is a fixed point to the $T^\star$-operator. Also,

$$(T^{\star n}\mu)(B) = \int p^{(n)}(x,B)\,\mu(dx),$$

where $p^{(n)}(x,B)$ is the n-step transition probability, the probability after n iterations of landing in B when starting at x.

The space of invariant probability measures, $\mathcal{P}_{\text{inv}}(X)$, to an IFS is non-empty (as follows by applying the Schauder-Tychonoff theorem to the operator $T^\star$, since the maps are continuous.

Let $\Omega = [m]^{\mathbf{N}} = \{\mathbf{i} = (i_1, i_2, \dots) : 1 \le i_k \le m\}$ and let P_x, for a fixed x, be the probability measure on Ω defined by

$$\begin{aligned} &P_x\{\mathbf{i} : j\text{th coordinate of } \mathbf{i} \text{ is } i_j,\ j = 1, \dots, n\} \\ &\qquad\qquad = p_{i_1}(x)p_{i_2}(S_{i_1}x)\cdots p_{i_n}(S_{i_{n-1}} \circ \cdots \circ S_{i_1}x). \end{aligned}$$

The process $\{Z_n^{\delta_x}\}$ can be realized on (Ω, P_x) as

$$Z_n^{\delta_x}(\mathbf{i}) = S_{i_n} \circ \cdots \circ S_{i_1}x.$$

We will also use the concept of mathematical expectation, which will be denoted by the symbol $\mathbf{E}$.

3. Approximation in one dimension and uniqueness.

3.1 The general construction. We will now present an algorithm by which, in the non-overlapping case with one-to-one maps, we will be able to construct approximating measures to a unique invariant measure. We will require for this construction of measures that either K is the closure of a bounded open subset of $\mathbf{R}^2$, or that $K = [0,1]$.

First we construct *set partition data*, (which we call *interval partition data* when dealing with $[0,1]$), $(\mathcal{S}_n, \nu_n)$, successively by defining a family of subsets with disjoint interiors in the following manner:

$$\begin{aligned} \mathcal{S}_0 &= \{K\}, \\ \mathcal{S}_1 &= \{S_jK : j = 1, \dots, m\}, \\ \mathcal{S}_2 &= \{S_iS_jK : i,\, j = 1, \dots, m\}, \qquad \text{etc.} \end{aligned}$$

Sometimes we refer to the sets in $\mathcal{S}_n$, for a fixed n, as a *generation* of images.

From Brouwer's domain theorem (Vick [**35**, p. 38]) it follows that the interior of K is mapped onto the interior of S_jK for all j and that the boundary of K is mapped onto the boundary of S_jK. Since this argument holds also for the next generation of images, we will have non-overlapping sets also in that generation and these will be included in the sets of the former generation. Also, the sets S_jK will be the closures of bounded open sets.

When the weight functions are all constants, we define set functions as follows

$$\begin{aligned} \nu_0(K) &= 1, \\ \nu_1(S_jK) &= p_j, \qquad \text{for all } j, \\ \nu_2(S_iS_jK) &= p_ip_j, \quad \text{for all } i,\, j, \qquad \text{etc.} \end{aligned}$$

Then we construct a unique sequence of probability measures $\{\mu_n\}$ by letting

$$\mu_n(A) = \nu_n(A) \quad \text{for all } A \in \mathcal{S}_n$$

with μ_n uniformly distributed in the interior of every set in every set partition $\mathcal{S}_n$.

3.2 The one-dimensional case. From the construction in Subsection 3.1 above, we see that in the one-dimensional case we get piecewise linear continuous distribution functions corresponding to the constructed measures. In Strichartz et al. [**33**, p. 105] it turned out that in the one-dimensional case, given the non-overlapping condition for one-to-one maps (monotonicity is sufficient), every invariant measure is atom-free and this implies that the distribution function of any invariant measure is continuous, which of course is necessary if we want to approximate such a function uniformly by continuous functions.

Definition 1. Let $(\mathcal{S}, \nu)$ denote set partition data in the sense that $\mathcal{S}$ is a finite collection of subsets A of K and ν is a set function satisfying the probability conditions $0 \le \nu(A) \le 1$ and $\sum_{A\in\mathcal{S}} \nu(A) = 1$. A measure $\mu \in \mathcal{P}(K)$ *matches the set partition data* $(\mathcal{S}, \nu)$ *exactly* if $\mu(A) = \nu(A)$ for all $A \in \mathcal{S}$.

In the one-dimensional case Strichartz et al. observed the following approximation in the uniform metric when the maps are continuous, one-to-one and non-overlapping and the weight functions are all constants:

$$\begin{aligned}\sup_{x\in[0,1]} |F_n(x) - F(x)| &= \sup_{x\in[0,1]} |\mu_n[0,x] - \mu[0,x]| \\ &= \sup_{\substack{x\in[a,b] \\ [a,b]\in\mathcal{S}_n}} |\mu_n[a,x] - \mu[a,x]| \\ &\le \sup_{[a,b]\in\mathcal{S}_n} \nu_n([a,b]) \le p_{\max}^n \to 0, \quad n\to\infty,\end{aligned}$$

where $p_{\max}$ denotes the largest weight. The error in each step n cannot exceed the value the set function assigns to the interval where the error stays, in a particular partition, and this is due to the exact matching of the interval partition data $(\mathcal{S}_n, \nu_n)$, for all n, of the invariant measure. This follows since the non-overlapping condition gives, see [**33**], $\mu(S_{i_n} \circ \cdots \circ S_{i_1} A) = p_{i_1} \cdots p_{i_n} \mu(A)$, for all $n \ge 1$, all sequences $i_1, \ldots i_n$ and all Borel subsets A of $[0,1]$.

They also proved the following related result (Strichartz et al. [**33**, Theorem 2.5]):

$$d_{\sup}(T^\star \mu_1, T^\star \mu_2) \le p_{\max}\, d_{\sup}(\mu_1, \mu_2),$$

where μ_1 and μ_2 are arbitrary members of $\mathcal{P}([0,1])$ and $d_{\sup}$ denotes the uniform metric with respect to the distribution functions corresponding to these probability measures.

The same result was obtained already in Dubins and Freedman [**10**, Theorem 5.4 and its corollary] for iteration of monotone i.i.d. maps. They considered a more general hypothesis, allowing all maps but two to be overlapping ('splitting with positive probability' was their terminology). They also got the same result when requiring at least two maps to be non-overlapping after a finite number of steps. However, the error in each step will, with their hypothesis, of course be larger than with the non-overlapping condition used by Strichartz et al. If we have the general non-overlapping after one iteration step, we get the error $(1 - p_{\min})^n$ in the nth step, since there is one probability weight to exclude, and a priori we may only exclude the one with least weight in each step.

It should also be mentioned that a broad extension of the theory of Dubins and Freedman for iteration of monotone maps were later given in Bhattacharya and Lee [**5**].

3.3 Ergodic decompositions. We now discuss ergodic decompositions of a compact metric state space. Here the theory diverges somewhat from the orthodox theory of Markov chains on a general state space, since in that theory one generally does not take topological concepts into consideration, which is natural in our context.

Definition 2. A Borel set $A \neq \varnothing$ is *stochastically closed* if $p(x, A) = 1$ for all $x \in A$.

Remark 1. In our case this means that $\cup_{j=1}^{m} S_j A \subseteq A$.

Definition 3. A non-empty subset F of X is an *ergodic kernel* if it is stochastically and topologically closed, and if it has no stochastically and topologically closed proper subsets.

Remark 2. It is an easy consequence of Zorn's lemma, see, e.g., Norman [**22**, p. 52], that every stochastically and topologically closed set contains a minimal stochastically and topologically closed set, i.e., an ergodic kernel. In some texts an ergodic kernel is referred to as a *minimal set*. From the minimality it follows that if E is an ergodic kernel, then $\cup_{j=1}^{m} S_j E = E$.

Remark 3. There always exists an invariant measure on any topologically and stochastically closed subset of X (as is easily seen by considering the iterates of a point in such a set).

It is also easy to see that the support $\operatorname{supp}\mu$ of an invariant measure μ, i.e., the smallest topologically closed set A which satisfies $\mu(A) = 1$) is stochastically closed. If $U := X \setminus \operatorname{supp}\mu$, then $0 = \mu(U) = \sum_j (p_j\mu)(S_j^{-1}U)$, which implies that $\mu(S_j^{-1}U) = 0$, for all j, and we have $S_j^{-1}U \subseteq U$, for all j. We conclude that for all j, $S_j(\operatorname{supp}\mu) \subseteq \operatorname{supp}\mu$. Hence $\operatorname{supp}\mu$ contains an ergodic kernel. In particular, if F is an ergodic kernel and μ is an invariant measure such that $\operatorname{supp}\mu \subseteq F$, then $\operatorname{supp}\mu = F$.

Lemma 1. *There exists a unique invariant measure only if there are no two distinct ergodic kernels. Any two distinct ergodic kernels are disjoint.*

Proof. Suppose we have two distinct ergodic kernels F_1 and F_2 and suppose that $F_1 \cap F_2 \neq \varnothing$. Let $F = F_1 \cap F_2$. Then F is both topologically and stochastically closed, implying that $F = F_1 = F_2$. That is, distinct ergodic kernels are disjoint.

It is now rather trivial to show that if we have two disjoint ergodic kernels, then we cannot have a unique invariant measure and this was also observed in Jamison [**16**, Theorem 2.4]. $\square$

Proposition 1. *A necessary and sufficient condition for uniqueness of an ergodic kernel is that for all x and y and all $\delta > 0$ there exist integers n and m and sequences $\{i_k\}_{k=1}^n$ and $\{j_l\}_{l=1}^r$ such that*

$$d(S_{i_n} \circ \cdots \circ S_{i_1} x, S_{j_r} \circ \cdots \circ S_{j_1} y) < \delta.$$

So this is also, by Lemma 1, a necessary condition for the existence of a unique invariant measure.

Proof. Suppose we have two distinct ergodic kernels, K_1 and K_2. Then we know from Lemma 1 that there exists a $\delta > 0$ such that $\delta = \min\{d(x,y) : x \in K_1,\, y \in K_2\}$. So, if for all x and y in X and all $\delta > 0$, there exist sequences $i_1, i_2, \ldots, i_n$ and $j_1, j_2, \ldots, j_r$ such that $d(S_{i_n} \circ \cdots \circ S_{i_1} x, S_{j_r} \circ \cdots \circ S_{j_1} y) < \delta$, then there can be only one ergodic kernel.

For the converse part we observe that if it is *not* true that for all x and y and all $\delta > 0$ there exist sequences $\{i_k\}_{k=1}^n$ and $\{j_l\}_{l=1}^r$ such that $d(S_{i_n} \circ \cdots \circ S_{i_1} x, S_{j_r} \circ \cdots \circ S_{j_1} y) < \delta$, then there are points $x, y \in X$ and a $\delta > 0$ such that $d(S_{i_n} \circ \cdots \circ S_{i_1} x, S_{j_r} \circ \cdots \circ S_{j_1} y) \geq \delta > 0$, for all sequences $\{i_k\}_{k=1}^n$ and $\{j_l\}_{l=1}^r$, all $r, n \geq 1$. Let $\mathbf{I}_x$ be the union of all iterates of x. Let A be the (topological) closure of $\mathbf{I}_x$ and B the closure of $\mathbf{I}_y$. Then $A \cap B = \varnothing$, and $p(x, A) = 1$ for all x in A and $p(y, B) = 1$

for all y in B, by continuous extension. Thus by Lemma 1 there exist at least two invariant measures. $\square$

Remark 4. This condition, as sufficient for a unique ergodic kernel, appeared in Jamison [**17**, p. 464], but no indication was given there that this is necessary for uniqueness of an invariant measure, a fact we here prove.

One would of course like to know when uniqueness of an ergodic kernel gives uniqueness of an invariant measure. One such situation, see Jamison [**17**, pp. 452–455, 463–464] is that the family $\{1/n \sum_{k=1}^{n} T^k f\}$ is equicontinuous, which for instance happens when $d(S_j x, S_j y) \leq d(x,y)$ for all $x, y \in X$, for all j, and the weight functions are constants, since then $\{T^n f\}$ is an equicontinuous family. But in general this is a difficult problem if not all maps are non-expanding or the weight functions are not all constants. If the non-overlapping condition is satisfied we can in the one-dimensional case prove a somewhat weaker result, Theorem 1, but first we will prove a useful lemma.

Lemma 2. *Let $S_j : X \to X$ and suppose μ is an invariant measure and that for a non-empty Borel set A we have $S_j A \subseteq A$ for all j. Then the measures $1_A\mu$ and $1_{X\setminus A}\mu$ ($1_A\mu(B) := \mu(A \cap B)$) are invariant.*

Proof. Let $\mu_A := 1_A\mu$. For all Borel sets B we have, since $A \cap S_j^{-1}(B) \subseteq S_j^{-1}(A \cap B)$, that

$$\begin{aligned}\mu_A(B) = \mu(A \cap B) &= \sum_j (p_j\mu)(S_j^{-1}(A \cap B)) \\ &\geq \sum_j (p_j\mu)(A \cap S_j^{-1}(B)) \\ &= \sum_j (p_j\mu_A)(S_j^{-1}(B)) = T^\star\mu_A(B).\end{aligned}$$

In the same way we get $\mu_A(X \setminus B) \geq T^\star\mu_A(X \setminus B)$. But

$$\begin{aligned}\mu_A(B) + \mu_A(X \setminus B) = \mu_A(X) &= \sum_j (p_j\mu_A)(X) \\ &= T^\star\mu_A(X) = T^\star\mu_A(B) + T^\star\mu_A(X \setminus B),\end{aligned}$$

so we have in fact equality. $\square$

Theorem 1. *If* $S_j : [0,1] \to [0,1]$ *are non-overlapping one-to-one maps, then all invariant measures of the* IFS *have the same support.*

Proof. The necessary and sufficient condition in Proposition 1 for a unique ergodic kernel is satisfied, since the total length of all intervals in each set partition is less than or equal to one. Since there exists a unique ergodic kernel E, the supports of all invariant measures must intersect, with E as their common intersection. There exists at least one invariant measure μ which has E as its support. We will prove that all invariant measures have E as their support. Suppose now that there exists an extremal, see Section 2, invariant measure μ' which does not have E as its support. Then it follows that $\mu'(E) = 0$, since by Lemma 2 all extremal invariant measures assigns measure 0 or 1 to a set A with $S_j A \subseteq A$ for all j.

Then $U := E^c$ satisfies $\mu'(U) = 1$. We will now prove that there exist an $r \geq 1$ and a sequence $j_1, \dots, i_r$, such that $S_{j_r} \circ \cdots \circ S_{j_1}[0,1] \subset U$, which contradicts the fact that E is stochastically closed.

There exists a finite union of disjoint closed intervals $\cup_{i=1}^{N} I_i \subset U$ such that $\mu'(\cup_{i=1}^{N} I_i) > 1/2$. It is sufficient that some image in some generation is included in the interior of one of these intervals. By the non-overlapping condition it is sufficient that at least one of the intervals I_i is hit by at least five images in some generation. Suppose this is not true. Then, for every fixed generation n, we have that $\mu'(\cup_{i=1}^{N} I_i) \leq 4N \cdot p_{\max}^n$, where $p_{\max} := \max_j \max_x p_j(x) < 1$, which contradicts $\mu'(\cup_{i=1}^{N} I_i) > 1/2$, as n becomes sufficiently large.

Thus E is the support of all invariant measures. $\square$

Remark 5. The conclusion of Theorem 1 does not mean that we have a unique invariant measure, since of course two measures can be mutually singular and yet have the same support. In fact it was proved in Stenflo [**29**, Theorem 1], following the construction in Bramson and Kalikow [**8**], that this situation may arise if we have an IFS with finitely many non-overlapping strict contractions together with strictly positive continuous probability weight functions.

4. Approximation in two dimensions.

4.1 One-to-one maps. To ensure uniqueness of an invariant measure in higher dimensions we cannot use only the non-overlapping condition together with injectivity and continuity of the maps, not even in the place-independent case, unless we make some additional assumptions. It is easy to find a counterexample: consider a triangle with vertices at $(0,0)$, $(0,1)$ and $(1,0)$ and let the IFS be the affine maps

$$S_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/2 & 0 \\ 1/2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and

$$S_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

with probability weights $p_1 = p_2 = 1/2$. Then there is an invariant unit point mass at $(0,0)$. There are, however, infinitely many invariant measures to this IFS, as is easily seen by studying lines parallel with the hypotenuse.

It is not even possible to generalize Algorithm 2.2 in Strichartz et al. [**33**, p. 106] to generate approximating measures of *some* invariant measure, because this algorithm depends on the assumption that there is no mass on the common boundaries of intersecting images in each generation $\mathcal{S}_n$, and this assumption is not automatically satisfied in two dimensions; this can also be seen from the counterexample above.

We will now generalize the ideas in Strichartz et al. [**33**, pp. 106–107] which we discussed in Subsection 3.2. We will use the same 'matching idea,' but as already mentioned, Strichartz et al. took advantage of the simplifications that are special to one dimension. So we have to face some higher dimensional problems on the way. We have already dealt with one such generalization in Subsection 3.1, when using Brouwer's domain theorem to get new set partitions containing non-overlapping sets.

As in the one-dimensional case we will be able to determine the local value of a unique invariant measure if we assume a very weak additional hypothesis. However, we will not be able to approximate the invariant measure using a probability metric; the uniform metric falls short for reasons explained in Subsection 4.2. In Subsection 4.3 we will approximate a unique invariant measure in the Wasserstein metric and in a generalized L^1-metric, but these are just 'average metrics,' which do not take the local behavior into consideration.

Lemma 3. *Suppose μ is any invariant measure to an* IFS *consisting of one-to-one maps $S_j : K \to K$. Then if μ satisfies $\mu(K^\circ) > 0$ and $\mu(\partial K) > 0$, we can decompose it into invariant components concentrated on ∂K and K°, respectively. Also, if the maps are non-overlapping and K is the closure of a bounded open set, we always have that*

$$\mu\left(\bigcup_{j=1}^{m} S_j \partial K \setminus \partial K\right) = 0.$$

Proof. From Brouwer's domain theorem we have that $\cup_{j=1}^{m} S_j K^\circ \subseteq K^\circ$. By Lemma 2 any invariant measure can be decomposed into an invariant part on the boundary of K and an invariant part in the interior of K. To prove the second assertion, we first show that

$$\partial K \cap \operatorname{supp} \mu \subseteq \bigcup_{j=1}^{m} S_j \partial K.$$

Let x be a member of the set $\partial K \cap \operatorname{supp} \mu$. Then $x \in \operatorname{supp} \mu$ implies that $x \in S_j K$ for some j. But Brouwer's domain theorem applies and we get $x \in S_j \partial K$ for some j. From the non-overlapping condition and the assumption that K is the closure of a bounded open set it follows (by applying the measure invariance formula) that $\mu(\cup_{j=1}^{m} S_j K^\circ) = \mu(K^\circ)$ and $\mu(\cup_{j=1}^{m} S_j \partial K) = \mu(\partial K)$, and thus we have

$$0 = \mu\left(\bigcup_{j=1}^{m} S_j \partial K\right) - \mu(\partial K) = \mu\left(\bigcup_{j=1}^{m} S_j \partial K \setminus \partial K\right). \qquad \square$$

The observation to make at this point is that if we move (with some map S_j) some part of the boundary of K from the boundary into the interior of K, then this part will not carry μ-mass.

Lemma 4. *Suppose we have an* IFS *with one-to-one maps* $S_j : K \to K$ *and that there exist* n *and a sequence* $i_1, \dots, i_n$ *such that*

$$S_{i_n} \circ \cdots \circ S_{i_1} K \subseteq K^\circ$$

Then, for every invariant measure μ*, we have* $\mu(K^\circ) = 1$.

Proof. If there exist an n and a sequence $i_1, \dots, i_n$ such that

$$S_{i_n} \circ \cdots \circ S_{i_1} K \subseteq K^\circ,$$

then, since $S_j K^\circ \subseteq K^\circ$ for all j, and since there exists a $\delta > 0$ such that $\inf_{x \in K} p_j(x) \geq \delta$ for all j, we have for all $x \in K$ that $P_x(Z_{mn}^{\delta_x} \in K^\circ) \geq 1 - (1 - \delta^n)^m \to 1$, as $m \to \infty$. Take any invariant measure μ. Then we have for all $m \geq 1$

$$\mu(K^\circ) = (T^{\star(mn)}\mu)(K^\circ) = \int p^{(mn)}(x, K^\circ)\, \mu(dx),$$

which tends to one, as m tends to infinity, and thus $\mu(K^\circ) = 1$. □

Remark 6. The hypothesis of Lemma 4 is satisfied if $K = [0,1]$ and the maps S_j are non-overlapping, continuous and one-to-one. It is enough to take $n = 2$, since by the non-overlapping condition it is then impossible that $S_{i_2} \circ S_{i_1}[0,1] \cap \{0,1\} \neq \varnothing$ for all i_1, i_2, since there are at least four intervals $S_{i_2} \circ S_{i_1}[0,1]$.

Lemma 5. *Suppose we have a non-overlapping* IFS *consisting of one-to-one maps* $S_j : K \to K$*, where* K *is the closure of a bounded open set and suppose that the hypothesis of Lemma* 4 *is satisfied. Then for every invariant measure* μ *we have no* μ*-mass on the overlaps of the images in each generation of the iteration process. And if the associated weight functions are all constants, we will, for all* $n \geq 1$ *and all sequences* $i_1, \dots, i_n$*, have that* $\mu(S_{i_n} \circ \cdots \circ S_{i_1} K) = p_{i_1} \cdots p_{i_n}$*, i.e.,* μ *matches the set partition data* $(\mathcal{S}_n, \nu_n)$ *exactly for all* n.

Proof. The non-overlapping condition, the assumption that K is the closure of a bounded open set and Brouwer's domain theorem imply that, with $j \neq k$,

$$\begin{aligned}\mu(S_jK \cap S_kK) &= \mu(\partial S_jK \cap (S_kK)^\circ) + \mu(\partial S_kK \cap (S_jK)^\circ) \\ &\quad + \mu(\partial S_jK \cap \partial S_kK) \\ &= \mu(\partial S_jK \cap \partial S_kK) \\ &= \mu(S_j\partial K \cap S_k\partial K).\end{aligned}$$

The second part of Lemma 3 gives

$$\mu(S_j\partial K \cap S_k\partial K) = \mu(\partial K \cap S_j\partial K \cap S_k\partial K).$$

Lemma 4 implies $\mu(\partial K) = 0$, which means that $\mu(S_jK \cap S_kK) = 0$.

In the case of constant weights we use this to prove that $\mu(S_{i_n} \circ \cdots \circ S_{i_1}K) = p_{i_1} \cdots p_{i_n}$, for all $n \geq 1$ and all sequences $i_1, \ldots, i_n$. Since

$$\begin{aligned}&\mu(S_{i_n} \circ \cdots \circ S_{i_1}K) \\ &\quad = \sum_{j=1}^{m} p_j \mu(S_j^{-1}(S_{i_n} \circ \cdots \circ S_{i_1}K)) \\ &\quad = p_{i_n}\mu(S_{i_{n-1}} \circ \cdots \circ S_{i_1}K) + \sum_{j \neq i_n} p_j \mu(S_j^{-1}(S_{i_n} \circ \cdots \circ S_{i_1}K)),\end{aligned}$$

this is equivalent to showing that $\mu(S_j^{-1}(S_{i_k} \circ \cdots \circ S_{i_1}K)) = 0$ for $j \neq i_k, k = 1, \ldots, n$. But $\mu(S_j^{-1}(S_{i_k} \circ \cdots \circ S_{i_1}K)) = (\mu \circ S_j^{-1})(S_jK \cap S_{i_k} \circ \cdots \circ S_{i_1}K)$, and this means that it is sufficient to show that $\mu(S_jK \cap S_{i_k}K) = 0$ for $j \neq i_k$, because $\mu \circ S_j^{-1}$ is absolutely continuous with respect to μ for all j, and $S_{i_k} \circ \cdots \circ S_{i_1}K \subseteq S_{i_k}K$.

Thus, if the weight functions are all constants, we have for all n and all sequences $i_1, \ldots, i_n$ that $\mu(S_{i_n} \circ \cdots \circ S_{i_1}K) = p_{i_1} \cdots p_{i_n}$. □

Lemma 6. *Suppose the maps S_j are one-to-one and non-overlapping. Suppose that ∂K is a Jordan curve, i.e., it is homeomorphic to a circle. Then if there are no common periodic points to all maps S_j with prime period one or two* (*the common periodic points do not necessarily have the same prime period for every map*), *we can draw the same conclusion as in Lemma* 5 *without assuming the hypothesis of Lemma* 4.

Proof. We know that if ∂K is a Jordan curve, then, since the maps S_j are continuous and one-to-one, $\partial S_j K = S_j \partial K$ are Jordan curves. From the non-overlapping condition and the Jordan curve theorem, see, e.g., Apostol [**1**], it follows that $S_j K^\circ \subset \mathbf{R}^2 \setminus S_k K$ and $S_k K^\circ \subset \mathbf{R}^2 \setminus S_j K$. This implies that $S_j K^\circ \cap S_k K = \varnothing$ and $S_k K^\circ \cap S_j K = \varnothing$. So we have that, with $j \neq k$, $S_j K \cap S_k K = \partial S_j K \cap \partial S_k K = S_j \partial K \cap S_k \partial K$ (the last equality again by Brouwer's domain theorem).

As in the proof of the first assertion of Lemma 5, Lemma 3 now gives that $\mu(S_j K \cap S_k K) = \mu(\partial K \cap \partial S_j K \cap \partial S_k K)$. Every discrete extremal invariant measure is uniformly distributed on a finite set of points (Dubins and Freedman [**10**, p. 840]), which is permuted by every map S_j. If we can prove that the set $\partial K \cap \partial S_j K \cap \partial S_k K$ contains at most two points, then by hypothesis $\mu(\partial K \cap \partial S_j K \cap \partial S_k K) = 0$, for any invariant measure μ.

Suppose that there exist at least *three* points a, b and c in the intersection $\partial K \cap \partial S_j K \cap \partial S_k K$ and that ∂K is a circle, which is possible by a topological conjugation argument. From a corollary of the Jordan curve theorem (Apostol [**1**, p. 184]) it now follows that we can join an arbitrary point in $S_j K^\circ$ with a, b and c on the circle with arcs which, except for the endpoints, lie entirely in $S_j K^\circ$. It follows from the observation that K is topologically conjugated to a disc that these arcs can be chosen to be Jordan arcs, i.e., simple arcs—arcs which do not intersect themselves.

But then we have three connected disjoint open domains, by the Jordan curve theorem. The domains are bounded by ∂K and the three Jordan arcs joining a, b and c with the arbitrary point. $S_k K^\circ$ cannot meet the Jordan arcs and thus has to be contained in one of the disjoint open domains. But then it is not possible to join an arbitrary point in $S_k K^\circ$ with all three points a, b and c, with arcs lying in $S_k K^\circ$ except for the endpoints. Thus, we can have at most two points in $\partial K \cap \partial S_j K \cap \partial S_k K$ and the conclusion follows. $\square$

Remark 7. In Lemma 6 we used a typically two-dimensional characterization, a Jordan curve. It is not known to the author if it is possible to state a similar result in three (and higher) dimensions.

We are now prepared to give our first uniqueness result. The method of proof is preparing for future attempts in two dimensions to deal with non-constant weight functions, possibly not even summing to one at all points, see Proposition 2 and Remark 10 below.

Theorem 2. *Suppose that we have constant weight functions and that either the hypotheses of Lemma* 5 *or those of Lemma* 6 *are satisfied. If* $\operatorname{diam}(S_{i_n} \circ \cdots \circ S_{i_1} K) \to 0$ *in probability, we have a unique invariant measure* μ *and* $\mu_n \to \mu$ *in the weak***-topology, where* μ_n *are the approximating measures constructed in Subsection* 3.1*. Also,* $\mu_n(A) = \mu(A)$ *for all* $A \in \mathcal{S}_n$*.*

Proof. From Lemma 5 or Lemma 6 and the construction in Subsection 3.1 it follows that for any invariant measure μ, $\mu_n(A) = \mu(A)$ for all $A \in \mathcal{S}_n$. We now show that when $\operatorname{diam}(S_{i_n} \circ \cdots \circ S_{i_1} K) \to 0$ in probability, we have a unique invariant measure μ with this property.

Let μ be an invariant measure. Let G be an open subset of K, and let $\varepsilon > 0$ be given. By regularity of the probability measures on K, there exists a compact subset C of G such that $\mu(G) - \varepsilon < \mu(C)$. It is easy to see that $\operatorname{diam}(S_{i_n} \circ \cdots \circ S_{i_1} K) \to 0$ in probability implies that $\operatorname{diam}(S_{i_1} \circ \cdots \circ S_{i_n} K) \to 0$ in probability and by monotonicity also almost surely. Then, since $\operatorname{diam}(S_{i_1} \circ \cdots \circ S_{i_n} K) \to 0$ almost surely, we have that for all $\varepsilon' > 0$, there exists, by Egoroff's theorem, a measurable subset Γ of Ω (see Section 2 for the definition of Ω) such that $P(\Omega \setminus \Gamma) \leq \varepsilon'$ and $\operatorname{diam}(S_{i_1} \circ \cdots \circ S_{i_n} K) \to 0$ uniformly on Γ. Then, on Γ, we can choose $N > 0$ so that $n \geq N$ implies $\operatorname{diam}(S_{i_1} \circ \cdots \circ S_{i_n} K) < (1/2)\operatorname{dist}(C, G^c) := (1/2)\min\{d(x,y) : x \in C, y \in G^c\}$. Since, for every invariant measure μ, $\operatorname{supp}\mu$ is covered by the union of the sets in $\mathcal{S}_n$ for a fixed n, we can choose a large n and $A_i^{(n)} \in \mathcal{S}_n$, $i = 1, \dots, r$, where $\cup_{i \leq r} A_i^{(n)} \subset G$, such that (with $\hat{Z}_n$ denoting the process induced by reversed iterations)

$$\mu\Big(\bigcup_{i \leq r} A_i^{(n)}\Big) = P\Big(\hat{Z}_n^\mu(\mathbf{i}) \in \bigcup_{i \leq r} A_i^{(n)}\Big) \geq \mu\Bigg(\bigcup_{\substack{A^{(n)} \in \mathcal{S}_n \\ A^{(n)} \cap C \neq \varnothing}} A^{(n)}\Bigg) - \varepsilon'.$$

We then have

$$\mu(G)-\varepsilon < \mu\Bigg(C\cap \bigcup_{A^{(n)}\in\mathcal{S}_n} A^{(n)}\Bigg) \le \mu\Bigg(\bigcup_{\substack{A^{(n)}\in\mathcal{S}_n\\ A^{(n)}\cap C\neq\varnothing}} A^{(n)}\Bigg) \le \mu\Bigg(\bigcup_{i\le r} A_i^{(n)}\Bigg)+\varepsilon'.$$

Since, by Lemma 5 or Lemma 6, μ matches set partition data exactly, we have $\mu(G)-\varepsilon < \mu_n(\cup_{i\le r}A_i^{(n)})+\varepsilon' \le \mu_n(G)+\epsilon'$. Thus, since ε and ε' can be chosen arbitrarily small, we have that $\liminf_{n\to\infty}\mu_n(G)\ge\mu(G)$, and we have convergence in the weak*-topology of the sequence $\{\mu_n\}$ to μ, which consequently is a unique invariant measure. □

Remark 8. The theorem generalizes some of the results of Strichartz et al. [**33**, pp. 106–107], since their conditions (non-overlapping and one-to-one maps) imply that $\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}[0,1])\to 0$ almost surely, since by the first part of Borel-Cantelli's lemma, we have for all $\varepsilon>0$ that

$$\sum_{n=1}^{\infty} P\{\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}[0,1])>\varepsilon\} \le \sum_{n=1}^{\infty}\frac{1}{\varepsilon}\,p_{\max}^n<\infty,$$

where $p_{\max}:=\max_j p_j$, and this implies that for all $\varepsilon>0$, $P\{\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}[0,1])>\varepsilon$ infinitely often$\}=0$.

Remark 9. It is possible to use the same method of proof as in Theorem 2 when considering non-compact Euclidean state spaces. Then we use the hypothesis that $\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}K)\to 0$ in probability for all compact subsets K of $\mathbf{R}^2$.

Then if there exists a (non-empty) bounded open set U, such that $S_jU\subseteq U$ for all j and $S_jU\cap S_kU=\varnothing$, $j\neq k$, we can proceed as before, since all invariant measures are concentrated on some subset of $\overline{U}$; if μ is an invariant measure and B_k is a sequence of bounded open sets containing $\overline{U}$ such that $\overline{B}_{k+1}\subset\overline{B}_k$ and $\bigcap\overline{B}_k=\overline{U}$, we have

$$\mu(\overline{B}_k)=\int \mathbf{E}\,1_{\overline{B}_k}(S_{i_n}\circ\cdots\circ S_{i_1}x)\,\mu(dx),$$

which tends to one, as n tends to infinity. Thus $\mu(\overline{U})=\mu(\cap\overline{B}_k)=1$.

If we drop the requirement that U should be bounded, we probably have to assume that the process $\{Z_n^{\delta_x}\}$ is *tight* for some x, and then for all x, meaning that for all $\varepsilon > 0$ there exists a compact set K_ε, such that with $\mu_n(A) = P(Z_n^{\delta_x} \in A)$ for all Borel subsets A of $\mathbf{R}^2$, we have $\mu_n(K_\varepsilon) > 1 - \varepsilon$ for all n. It is sufficient (Barnsley et al. [**3**, p. 380]) to assume that there exists a finite constant M such that $\mathbf{E}\, d(x, Z_n^{\delta_x}) \leq M$ for all n.

The following proposition adapts the situation in Theorem 2 to the place-dependent case, as we shall see in Remark 10 below.

Proposition 2. *Assume that the maps $S_j : K \to K$ are one-to-one and non-overlapping. Suppose we have an invariant measure μ matching the sequence of set partition data $\{(\mathcal{S}_n, \mu_n)\}_{n\geq 1}$ asymptotically in the sense that*

$$\sum_{A\in\mathcal{S}_n} |\mu_n(A) - \mu(A)| \leq \varepsilon_n, \quad \textit{where } \varepsilon_n \to 0, \quad \textit{as } n \to \infty.$$

Then if, for all index sequences $\operatorname{diam}(S_{i_n} \circ \cdots \circ S_{i_1} K) \to 0$*, we have a unique invariant measure μ and $\mu_n \to \mu$ in the weak*-topology.*

Proof. We modify the proof of Theorem 2 slightly by noting that since, for all index sequences $\{i_j\}$ $\operatorname{diam}(S_{i_n} \circ \cdots \circ S_{i_1} K) \to 0$, we can choose $N > 0$ such that $n \geq N$ implies $\operatorname{diam}(S_{i_n} \circ \cdots \circ S_{i_1} K) < (1/2)\operatorname{dist}(C, G^c)$. As before we can, given $\varepsilon > 0$, by regularity take a compact subset C of G, such that if n is chosen large enough, we have

$$\mu(G) - \varepsilon < \mu(C) \leq \mu\biggl(\bigcup_{i\leq m} A_i^{(n)}\biggr),$$

where $A_i^{(n)}, i = 1, \ldots, m$, is some collection of sets in $\mathcal{S}_n$ also included in the given open set G. Since $\sum_{A\in\mathcal{S}_n} |\mu_n(A) - \mu(A)| \leq \varepsilon_n$, we have

$$\mu(G) - \varepsilon < \mu_n\biggl(\bigcup_{i\leq m} A_i^{(n)}\biggr) + \varepsilon_n.$$

Since $\varepsilon_n \to 0$ as $n \to \infty$, it follows that $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ and thus $\mu_n \to \mu$ in the weak*-topology and μ is the unique invariant measure. $\square$

Remark 10. The notion of asymptotic matching of a sequence of set partition data in Proposition 2 above becomes natural if one attempts to approximate unique invariant measures associated to IFS with place-dependent weight functions. This was what Strichartz et al. [**33**, Section 3, pp. 109–112] were trying to do on $[0,1]$ with some partial success. They did not have any criteria ensuring uniqueness of an invariant measure in the place-dependent case, but they provided a convincing approximation model of invariant measures, as follows. Suppose we are given the general situation of

$$\lambda\mu = \sum_{j=1}^{m} (q_j \mu) \circ S_j^{-1},$$

where the q_js are projectivized weight functions, not necessarily satisfying a probability condition, and $\lambda > 0$.

Now let $q_{jk}^{+}(J) = \sup_{x\in J} q_j(x)$ and $q_{jk}^{-}(J) = \inf_{x\in J} q_j(x)$, where J is an interval belonging to the usual interval partition $\mathcal{S}_k$. It is obvious that $q_{jk}^{+}(J) < \infty$ and that $q_{jk}^{-}(J) > 0$. By the integral mean value theorem there exists a constant $q_{jk}(J)$, with $q_{jk}^{-}(J) \le q_{jk}(J) \le q_{jk}^{+}(J)$, so that every invariant measure μ satisfies

$$\lambda\mu(J') = \sum_{S_j J \subseteq J'} q_{jk}(J)\mu(J).$$

Then one approximates $\mu(J')$ by solving the system

$$\lambda_k^{\pm}\mu_k^{\pm}(J') = \sum_{S_j J \subseteq J'} q_{jk}^{\pm}(J)\mu_k^{\pm}(J)$$

for $J \in \mathcal{S}_k$, where the $\mu_k^{\pm}$'s are approximating measures. It follows (Strichartz et al. [**33**, Theorem 3.2]) that $\lambda_k^{\pm} \to \lambda$, which consequently is unique.

On $[0,1]$, we obtain the following estimate in the uniform metric:

$$\sup_{x\in[0,1]} |\mu_k^{\pm}[0,x] - \mu[0,x]| \le \sup_{J\in\mathcal{S}_k} \mu_k^{\pm}(J) + \sum_{J'\in\mathcal{S}_k} |\mu_k^{\pm}(J') - \mu(J')|.$$

This estimate can, under the same additional assumptions, be used to give a criterium for uniqueness of an invariant measure. As in

Proposition 2 one has to give criteria for the matching error sum to converge to zero. In the case of strictly positive continuous probability weight functions, we have in Lemmas 3, 4, 5 and 6 given adequate conditions for an investigation of this sort.

Under suitable conditions there are two things one would like to verify:

1° $\sum_{A\in\mathcal{S}_n} |\mu_n^+(A) - \mu_n^-(A)| \to 0$, as $n \to \infty$, implies that

$$\sum_{A\in\mathcal{S}_n} |\mu_n^{\pm}(A) - \mu_n(A)| \to 0, \quad \text{as } n \to \infty.$$

2° $\sum_{A\in\mathcal{S}_n} |\mu_n^+(A) - \mu_n^-(A)| \to 0$, as $n \to \infty$.

The assumption that for all sequences $\{i_j\}$, $\operatorname{diam}(S_{i_n} \circ \cdots \circ S_{i_1} K) \to 0$, is natural in the case of place-dependent IFS if we do not assume that the weight functions necessarily sum to 1.

4.2 Probability metrics. Before we continue to approximate invariant measures in the place-independent case, we will make ourselves acquainted with some probability metrics.

The Wasserstein (or Monge-Kantorovich, sometimes Hutchinson) distance is defined by

$$d_W(\mu_1, \mu_2) = \sup_{\mathrm{Lip}\,(f)\le 1} \left| \int_X f\,d\mu_1 - \int_X f\,d\mu_2 \right|,$$

where $\{\mathrm{Lip}\,(f) \le 1\} = \{f : X \to \mathbf{R}, |f(x) - f(y)| \le d(x,y)\}$. The Wasserstein metric possesses very nice properties in higher Euclidean dimensions and generalizes to complete metric spaces (on compact metric spaces it is the same as the Lipschitz metric used by Dudley [**11**]) but it is often difficult to grasp the geometrical meaning of it. Since X is a compact metric space, convergence in the Wasserstein metric will be equivalent to weak*-convergence, and this means that we can approximate any probability measure μ on X with finite linear combinations of Dirac measures. (It was shown in Gadde [**14**, Theorem 3.38], that such an approximation is possible in the d_W-metric also if X no longer is compact.)

The L^1-metric is usually only defined for one-dimensional distribution functions as follows:

$$d_{L^1}(\mu, \nu) = \int |F_\mu(x) - F_\nu(x)|\, dx.$$

In one dimension, the Wasserstein distance can be calculated by means of the d_{L^1}-metric, i.e., $d_W(\mu, \nu) = d_{L^1}(\mu, \nu)$. This was the main result of Vallender in [**34**].

However, it seems difficult to obtain an L^1-metric with which it is easy to work. For instance, generalized L^1-metrics are not rotationally invariant.

The uniform metric $d_{\sup}$, which is sometimes called the Kolmogorov metric, can be defined on every Euclidean space $\mathbf{R}^k$ as

$$d_{\sup}(\mu, \nu) = \sup_{x \in \mathbf{R}^k} |F_\mu(x) - F_\nu(x)|.$$

However, in our approximation context it does not work very well for dimensions greater than one, since its lack of rotation invariance causes more of a problem than in the case of generalized L^1-metrics. It is easy to find examples where we first have a good approximation and then, after having rotated the plane (if we are in two dimensions) one millionth of a degree, we have the constant error one. Also, there can be more than one set in each partition contributing to the error, and their 'weight sum' does not necessarily drop off as in the one-dimensional case.

Strichartz et al. (Lemma 2.3, Corollary 2.4 and Theorem 2.5 of [**33**]) used the d_{L^1}-metric to approximate invariant measures on $[0, 1]$, and we will generalize some of their results to higher dimensions, but first we will discuss how we should define a d_{L^1}-metric in higher dimensions and see what, if any, relation such a metric could have with the Wasserstein metric. Vallender ([**34**, Remark 1]) claimed that if instead of considering the usual Euclidean distance on $\mathbf{R}^k$, $k \geq 2$, we use the l^1-metric in the definition of the Wasserstein metric, i.e., if the d-metric above is the l^1-metric, then we can still calculate the distance in the Wasserstein metric with the L^1-metric. This is however not true, since if we consider the distance between two distinct Dirac measures δ_a and δ_b in the d_{L^1} sense on the plane, this distance will be infinite, but the

Wasserstein distance will of course be finite ($= d(a, b)$). There will also be a problem of calculating the distance in the L^1-metric between two distinct Dirac measures if we integrate over a compact subset K of $\mathbf{R}^2$, if we for example have a vertical line as part of the boundary with the interior K° on the left. Then the distance between δ_a and δ_b, $a \neq b$, both placed on this vertical line, is zero.

So in the higher dimensional case ($k \geq 2$) we have to be very careful when defining an L^1-metric. For a given compact state space $K \subset \mathbf{R}^k$ we have to integrate over some larger domain containing K in order to define a proper L^1-metric. Thus in the following we will by the L^1-metric mean

$$d_{L^1}(\mu_1, \mu_2) = \int_{\mathcal{O}} |F_{\mu_1}(x) - F_{\mu_2}(x)| \, dx,$$

where $\mathcal{O}$ is some appropriately chosen bounded open set containing K. Then the Wasserstein and d_{L^1}-metrics will be equivalent, due to the fact that $\mathcal{O}$ is bounded, both metrizing the weak*-topology. So there is no general characterization of the relation between the magnitudes of the d_{L^1}- and d_W-metrics; distances in the d_{L^1}-metric depends on the size of $\mathcal{O}$ and that does not affect distances in the d_W-metric.

In Rachev [**25**, p. 74] a weight function in the integrand is used in the defining expression of an L^1-metric and one is integrating over all of $\mathbf{R}^k$. This is analogous to our definition, but instead of integrating over a somewhat arbitrary space, Rachev places an arbitrary function in the integrand. He then obtains equality between his L^1-metric and the Wasserstein metric when some additional assumptions are made. These additional assumptions are very difficult to check; one has to verify properties of the derivatives of the distribution functions involved.

4.3 Mean approximation and uniqueness. It is well known, see for instance Bhattacharya and Majumdar [**6**, especially p. 5, condition (a)], that the existence of a unique invariant probability measure follows for continuous i.i.d. maps S_j on a compact metric space X, if

$$\mathbf{E} \operatorname{diam}(S_{i_n} \circ \cdots \circ S_{i_1} X) \to 0.$$

This is due to the fact that

$$d_W(T^{\star n}\nu, \mu) \leq \mathbf{E} \operatorname{diam}(S_{i_n} \circ \cdots \circ S_{i_1} X),$$

for every invariant probability measure μ and any probability measure ν. This can be seen by iterating Dirac measures and then approximate with finite linear combinations of such measures.

The same method applies if we consider the L^1-metric we defined in Subsection 4.2:

$$\begin{aligned} d_{L^1}(T^{\star n}\delta_a, T^{\star n}\delta_b) &= \int_{\mathcal{O}} |\mathbf{E}F_{\delta_{S_{i_n}\circ\cdots\circ S_{i_1}a}}(x) - \mathbf{E}F_{\delta_{S_{i_n}\circ\cdots\circ S_{i_1}b}}(x)|\,dx \\ &\leq \mathbf{E}\Big\{\int_{\mathcal{O}} |F_{\delta_{S_{i_n}\circ\cdots\circ S_{i_1}a}}(x) - F_{\delta_{S_{i_n}\circ\cdots\circ S_{i_1}b}}(x)|\,dx\Big\} \end{aligned}$$

and integrating over the 'cross' bounded by the vertical and horizontal lines through a and b and by $\mathcal{O}$. The last quantity is dominated by

$$\begin{aligned} &\mathbf{E}\{2\text{ diam}\,(\mathcal{O})\text{ diam}\,(S_{i_n}\circ\cdots\circ S_{i_1}K)\} \\ &\qquad = 2\text{ diam}\,(\mathcal{O})\,\mathbf{E}\text{ diam}\,(S_{i_n}\circ\cdots\circ S_{i_1}K). \end{aligned}$$

Thus we have the following result:

Theorem 3. *If $X = K$, a compact subset of $\mathbf{R}^2$, we will for all $\nu \in \mathcal{P}(K)$ have the estimate*

$$d_{L^1}(T^{\star n}\nu, \mu) \leq 2\text{ diam}\,(\mathcal{O})\,\mathbf{E}\text{ diam}\,(S_{i_n}\circ\cdots\circ S_{i_1}K).$$

Remark 11. We have a straightforward generalization to higher dimensions, the estimate of the L^1-distance being

$$d_{L^1}(T^{\star n}\nu, \mu) \leq k\text{ diam}\,(\mathcal{O})\,\mathbf{E}\{\text{diam}\,(S_{i_n}\circ\cdots\circ S_{i_1}K)\}^{k-1},$$

for all $\nu \in \mathcal{P}(K)$, where k is the dimension of the ambient Euclidean space.

Remark 12. In Barnsley and Elton [**3**] an average contractive condition was introduced for Lipschitz continuous functions in the following way:

$$\sum_{j=1}^{m} p_j \log\frac{d(S_jx, S_jy)}{d(x,y)} < 0 \quad \text{uniformly in } x \text{ and } y.$$

It can be seen by elaborating Lemma 1 of Barnsley and Elton [**3**] that these kinds of average contractions imply that $\mathbf{E}\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}X)\to 0$. In fact it can be seen from Silvestrov and Stenflo [**26**, p. 12] that the average contractive conditions above imply $\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}X)\to 0$ almost surely.

The condition $\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}X)\to 0$ in probability is strictly weaker than $\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}X)\to 0$ almost surely, which can be seen by taking the IFS $S_1x=(1/2)x$; $S_2x=2x$, $0\le x\le 1/2$, $S_2x=1$, $x>1/2$, on $[0,1]$ with the weights $p_1=p_2=1/2$. Then $-\log\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}[0,1])\to\infty$ in probability, but visits 0 infinitely often with probability one.

Example 1 shows that we cannot always expect convergence with geometric rate to a unique invariant measure.

Example 1. Consider the function $Sx=x/(1+x)$ on $[0,1]$. Clearly $S^nx\to 0$ for all $x\in[0,1]$, so under iteration of the $T^\star$-operator, every probability measure on $[0,1]$ converges to δ_0 in the weak*-topology. But if we start at $x_0=1$, we get $x_1=Sx_0=1/2$, $x_2=S^2x_0=1/3$, $x_n=S^nx_0=1/(n+1)$, and this does not converge to 0 with geometric rate. It is now easy to construct an IFS with the same property. Let $S_1x=Sx=x/(1+x)$ and $S_2x=x$ and let $p_1=p_2=1/2$.

We also see that this IFS is not average contractive, since (with the notation as in Remark 12) for every n and for every sequence $i_1,\dots,i_n$, we have $\|S_{i_n}\circ\cdots\circ S_{i_1}\|=1$, where $\|S_j\|=\sup_{x\neq y}(d(S_jx,S_jy))/(d(x,y))$. But for all $\varepsilon>0$: $P\{\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}[0,1])>\varepsilon \text{ infinitely often}\}=0$, which is equivalent to $\operatorname{diam}(S_{i_n}\circ\cdots\circ S_{i_1}[0,1])\to 0$ almost surely.

Example 2 illustrates that we cannot expect that uniform convergence diam$\to 0$ is always present if we have non-overlapping maps.

Example 2. Consider the IFS $S_1x=((3/2)x)/(1+x)$, $S_2x=(1+(1/2)x)/(1+x)$, $p_1=p_2=1/2$, on $[0,1]$. This IFS is non-overlapping, but the diameter of the iterates does not tend to zero uniformly: if we iterate S_1 always, the diameter will be bounded away from zero. However, the results in Strichartz et al. [**33**, Theorem

2.5] apply and we also know, from Remark 8, that diam $(S_{i_n} \circ \cdots \circ S_{i_1}[0,1]) \to 0$ almost surely.

Example 3 shows that it is not necessary for uniqueness of an invariant measure that diam $(S_{i_n} \circ \cdots \circ S_{i_1} X) \to 0$ in probability.

Example 3. We can have a unique invariant measure, although the expected diameter does not tend to 0, if we iterate two continuous one-to-one maps on $[0,1]$. Let one of the two maps be $g_1(x) = 1 - x^2$. This map has precisely two periodic points with prime period two at $x = 0$ and $x = 1$ and a fixed point placed at $(\sqrt{5}-1)/2$. The periodic points with prime period two are attracting, in fact they are super-attracting, since $(g_1^2)' = 0$, and the fixed point is easily seen to be repelling, just compute $|g_1'|$. The point is now to choose g_2 to be a small perturbation of g_1, so that the fixed points are not the same, but so that the periodic points with prime period two are identical.

We now take the function $g_2(x) = 1 - x^2 - \varepsilon x(1-x)$ where we let $\varepsilon > 0$ be small. (The author was led to this choice of the function g_2, by the heuristics that the logistic function $f(x) = \lambda x(1-x)$ does not have any interesting dynamics for small values of the parameter λ.) It has precisely two attracting points with prime period two at $x = 0$ and $x = 1$ and a repelling fixed point at $(\sqrt{\varepsilon^2 - 2\varepsilon + 5} - 1 - \varepsilon)/(2(1-\varepsilon))$ which does not coincide with that of g_1 for any $\varepsilon > 0$. In fact, the graph of g_2 always lies under the graph of g_1. Consider now $g_1^2 = 1 - (1-x^2)^2 = 2x^2 - x^4$. It has three fixed points, one in the interior of $[0,1]$, the same as that for g_1, and two at the endpoints of $[0,1]$. To the left of the fixed point $(\sqrt{5}-1)/2$, g_1^2 is below $f(x) = x$ and to the right of $(\sqrt{5}-1)/2$, g_1^2 is above $f(x) = x$. Of course there are no other periodic points of g_1^2. Graphical analysis gives that, except for the fixed point $(\sqrt{5}-1)/2$, all points in the interior of $[0,1]$ tend to the boundary under iteration of g_1^2. More precisely, the points in the interior to the left of $(\sqrt{5}-1)/2$ are iterated to 0, and those to the right of $(\sqrt{5}-1)/2$ are iterated to 1. This shows that if we have an IFS consisting of g_1 and g_2 (both chosen with positive probability) we have, for a given $\delta > 0$, a positive probability of leaving the subinterval $[\delta, 1-\delta]$, when starting at some point there.

We now consider the explicit choice $\varepsilon = 1/3$, so that $g_2(x) = 1-(2/3)x^2-(1/3)x$, and observe that for all $\delta \leq 1/4$, $[0,1]\backslash[\delta, 1-(1/2)\delta]$ is stochastically closed (invariant) under the two maps g_1 and g_2. Then we cannot have an invariant measure except for $(1/2)(\delta_0+\delta_1)$.

REFERENCES

1. T.M. Apostol, *Mathematical analysis*, Addison-Wesley, New York, 1957.

2. M.F. Barnsley, S.G. Demko, J.H. Elton and J.S. Geronimo, *Invariant measures for Markov processes arising from place-dependent probabilities*, Ann. Inst. H. Poincaré Probab. Statist. **24** (1988), 367–394.

3. M.F. Barnsley and J.H. Elton, *A new class of Markov processes for image encoding*, Adv. in Appl. Probab. **20** (1988), 14–32.

4. M.F. Barnsley, J.H. Elton and D.P. Hardin, *Recurrent iterated function systems*, Constr. Approx. **5** (1989), 3–31.

5. R. Bhattacharya and O. Lee, *Asymptotics of a class of Markov processes which are not in general irreducible*, Ann. Probab. **16** (1988), 1333–1347. Correction, ibid. **25** (1997), 1541–1543.

6. R. Bhattacharya and M. Majumdar, *Dynamical systems subject to random schocks*: *An introduction*, and *Random dynamical systems*: *A review*, Economic Theory **23** (2004), 1–12, 13–38.

7. P. Billingsley, *Convergence of probability measures*, Wiley, New York, 1968.

8. M. Bramson and S. Kalikow, *Nonuniqueness in g-functions*, Israel J. Math. **84** (1993), 153–160.

9. P. Diaconis and D. Freedman, *Iterated random functions*, SIAM Rev. **41** (1999), 45–76.

10. L.E. Dubins and D.A. Freedman, *Invariant probabilities for certain Markov processes*, Ann. Statist. **37** (1966), 837–848.

11. R.M. Dudley, *Convergence of Baire measures*, Studia Math. **27** (1966), 251–268.

12. J.H. Elton and Z. Yan, *Approximation of measures*, Constr. Approx. **5** (1989), 69–87.

13. A.H. Fan and K.-S. Lau, *Iterated function system and Ruelle operator*, J. Math. Anal. Appl. **231** (1999), 319–344.

14. E. Gadde, *Stable iterated function systems*, Ph.D. Thesis, Department of Mathematics, Umeå University, 1992.

15. J.E. Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J. **30** (1981), 713–747.

16. B. Jamison, *Asymptotic behavior of successive iterates of continuous functions under a Markov operator*, J. Math. Anal. Appl. **9** (1964), 203–214.

17. ———, *Ergodic decomposition induced by certain Markov operators*, Trans. Amer. Math. Soc. **117** (1965), 451–468.

18. A. Johansson and A. Öberg, *Square summability of variations of g-functions and uniqueness of g-measures*, Math. Res. Lett. **10** (2003), 587–601.

19. K.-S. Lau, *Iterated function systems with overlaps and multifractal structure*, in *Trends in probability and related analysis*, (Taipei, 1998), World Sci. Publ., River Edge, NJ, 1999, pp. 35–76.

20. K.-S. Lau and J. Wang, *Mean quadratic variations and Fourier asymptotics of self-similar measures*, Monatsh. Math. **115** (1993), 99–132.

21. G. Letac, *A contraction principle for certain Markov chains and its applications*, in *Random matrices and their applications*, (Brunswick, Maine, 1984), Contemp. Math., vol. 50, Amer. Math. Soc., Providence, RI, 1986, pp. 263–273.

22. M.F. Norman, *Markov processes and learning models*, Academic Press, New York, 1972.

23. A. Öberg, *Approximation of invariant measures for iterated function systems*, Ph.D. Thesis, Department of Mathematics, Umeå University, 1998.

24. A. Öberg, R.S. Strichartz and A.Q. Yingst, *Level sets of harmonic functions on the Sierpiński gasket*, Ark. Mat. **40** (2002), 335–362.

25. S.T. Rachev, *Probability metrics*, Wiley, New York, 1991.

26. D.S. Silvestrov and Ö. Stenflo, *Ergodic theorems for iterated function systems controlled by regenerative sequences*, J. Theoret. Probab. **11** (1998), 589–608.

27. W. Słomczyński, J. Kwapień and K. Zyczkowski, *Entropy computing via integration over fractal measures*, Chaos **10** (2000), 180–188.

28. Ö. Stenflo, *Ergodic theorems for iterated function systems with time dependent probabilities*, Theory Stoch. Process. **19** (1997), 436–446.

29. ———, *A note on a theorem of Karlin*, Statist. Probab. Lett. **54** (2001), 183–187.

30. ———, *Uniqueness of invariant measures for place-dependent random iterations of functions*, Fractals in Multimedia (Minneapolis, MN, 2001), IMA Vol. Math. Appl., vol. 132, Springer, New York, 2002, pp. 13–32.

31. ———, *Uniqueness in g-measures*, Nonlinearity **16** (2003), 403–410.

32. R.S. Strichartz, *Self-similar measures and their Fourier transforms* III, Indiana Univ. Math. J. **42** (1993), 367–411.

33. R.S. Strichartz, A. Taylor and T. Zhang, *Densities of self-similar measures on the line*, Experiment. Math. **4** (1995), 101–128.

34. S.S. Vallender, *Calculation of the Wasserstein distance between two probability distributions on the line*, Theory Probab. Appl. **18** (1973), 784–786.

35. J.W. Vick, *Homology theory*, Academic Press, New York, 1973.

Department of Mathematics, Uppsala University, P.O. Box 480, S-751 06 Uppsala, Sweden
E-mail address: anders@math.uu.se