ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 28, Number 3, Fall 1998

A GENERALIZATION OF IFS WITH PROBABILITIES TO INFINITELY MANY MAPS

FRANKLIN MENDIVIL

ABSTRACT. This paper considers the problem of extending the notion of an IFS with probabilities from the case of finitely many maps in the IFS to the case of infinitely many maps. We prove that, under an average contractivity condition, the IFS is contractive in the Monge-Kantorovich metric. We also show that the invariant distribution is continuous with respect to the parameters of the IFS. Furthermore, using results of Lewellen, we obtain a result relating the support of the invariant measure to the attractor of the "geometric" IFS. Finally, we discuss the issue of the convergence of integrals with respect to the invariant measure and estimates on the error of these integrals.

1. Introduction. In his seminal paper [**3**], Hutchinson discusses the notion of self-similarity and introduces some ways to measure or define self-similarity. One such way is to say that a set $A \subset X$ is self-similar if there is come collection of maps $w_i : X \to X$ so that

$$A = \bigcup_i w_i(A).$$

In this way, A is seen to be made up of transformed copies of itself. Given this set of maps, one can define a set-valued map W by

$$W(B) = \bigcup_i w_i(B)$$

and we see that A is self-similar under the w_i's if A is a fixed point of W. While Hutchinson considered only finitely many maps, later Lewellen considered the case of infinitely many maps indexed by some compact metric space [**4**].

Received by the editors on May 4, 1996, and in revised form on October 15, 1996.
This research was supported in part by a Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Grant in the form of a postdoctoral fellowship.

Given a collection of maps w_i on X and a set of probabilities p_i, i.e., $\sum p_i = 1$ and $p_i \geq 0$, we can define an associated Markov operator M on the set of probability measures on X (a so-called IFS with probabilities)

$$M(\mu)(S) = \sum_i p_i \mu(w_i^{-1}(S))$$

for all Borel subsets S. If μ is a fixed point of M, then μ can also be said to be self-similar. In this paper we generalize the results of Hutchinson to the case where there are infinitely many maps. Thus, this work can be seen as a natural complement to Lewellen's work.

2. Main results. Let X and Λ be compact metric spaces. The space Λ is our parameter space. Let $w : \Lambda \times X \to X$ be continuous in both x and λ. Let p be a probability measure on Λ. We denote by $\mathcal{PM}(X)$ the set of regular probability measures on X. Define $T : \mathcal{PM}(\Lambda) \times \mathcal{PM}(X) \to \mathcal{PM}(X)$ by

$$T_p(\mu)(B) = \int_\Lambda \mu(w_\lambda^{-1}(B))\, dp(\lambda) \quad \text{for all Borel sets} \quad B \subset X$$

where $\mu \in \mathcal{PM}(X)$ and $p \in \mathcal{PM}(\Lambda)$. When $p \in \mathcal{PM}(\Lambda)$ is fixed, we leave out the subscript on T_p.

Note that by the Riesz representation theorem (for continuous linear functionals on $C(X)$) an equivalent way of defining $T(\mu)$ would be

$$\int_X f(x)\, dT(\mu)(x) = \int_\Lambda \int_X f(w_\lambda(x))\, d\mu(x)\, dp(\lambda)$$

for all continuous bounded real-valued functions f on X. Formally, this definition works exactly the same as the previous definition.

It is a straightforward calculation to show that T maps $\mathcal{PM}(X)$ to $\mathcal{PM}(X)$, so we leave out the details.

We now recall the definition of the Monge-Kantorovich metric ([**3**] and called the Hutchinson metric in [**1**]). Let μ and ν be two probability measures on X. Then

$$d_H(\mu, \nu) = \sup \left\{ \int_X f(x)\, d\mu(x) - \int_X f(x)\, d\nu(x) : f \in \mathrm{Lip}_1(X) \right\}$$

where we denote by $\mathrm{Lip}_k(X)$ the functions on X with Lipschitz constant at most k. The Monge-Kantorovich metric induces the topology of weak convergence of measures on $\mathcal{PM}(X)$, see [**3, 1**].

We prove now that T is a contraction if w is contractive on average.

Definition 1. We say that w is *contractive on average* if, for all $x, y \in X$,

$$\int_\Lambda d(w_\lambda(x), w_\lambda(y))\, dp(\lambda) \le s d(x,y)$$

with $s < 1$. We call the minimum such s the contraction factor.

Theorem 1. *If w is contractive on average, then T is contractive in the Monge-Kantorovich metric.*

Proof. Let $f \in \mathrm{Lip}\,(X)$. We calculate for μ and ν in $\mathcal{PM}(X)$,

$$\begin{aligned}
&\int_X f(x)\, d(T(\mu) - T(\nu)))\, dx \\
&\qquad = \int_X f(x)\, d\bigg(\int_\Lambda (\mu(w_\lambda^{-1}(x)) - \nu(w_\lambda^{-1}(x)))\, dp(\lambda)\bigg) \\
&\qquad = \int_\Lambda \int_X f(x)\, d(\mu(w_\lambda^{-1}(x)) - \nu(w_\lambda^{-1}(x)))\, dp(\lambda) \\
&\qquad = \int_\Lambda \int_X f(w_\lambda(y))\, d(\mu(y) - \nu(y))\, dp(\lambda) \\
&\qquad = \int_X \bigg(\int_\Lambda f(w_\lambda(y))\, dp(\lambda)\bigg)\, d(\mu(x) - \nu(x)) \\
&\qquad = s \int_X \phi(y)\, d(\mu(y) - \nu(y))
\end{aligned}$$

where $\phi(y) = s^{-1} \int_\Lambda f(w_\lambda(y))\, dp(\lambda) \in \mathrm{Lip}\ (X)$ by definition of s. Taking the supremum, we get

$$d_H(T(\mu), T(\nu)) \le s d_H(\mu, \nu)$$

and the result follows. □

Notice that any probability measure on X is the attractor of an infinite IFS with probabilities in a trivial way. If we wish to obtain the measure μ, we simply take $\Lambda = X$ and $w_x(y) = x$ for each $y \in X$ and $p = \mu$. It is easy to see that $T(\nu) = \mu$ for any $\nu \in \mathcal{PM}(X)$.

Now we prove a result about continuous dependence of the invariant measure with respect to the "parameters" of the IFS (the probability measure p on Λ). For a fixed probability measure, p, we denote by μ_p the invariant measure of the operator T_p.

Theorem 2. *Suppose that $p^{(n)}$ is a sequence of probability measures in $\mathcal{PM}(\Lambda)$ which converges to $p \in \mathcal{PM}(\Lambda)$ in the Monge-Kantorovich metric* (*weak convergence of measures*). *Then $\mu_{p^{(n)}} \Rightarrow \mu_p$ in the Monge-Kantorovich metric.*

Proof. Let f be a bounded continuous function on X. We calculate that

$$\begin{aligned} \int_X f(x)\, d\biggl(\int_\Lambda \mu(w_\lambda^{-1}(x))\, dp^{(n)}(\lambda)\biggr) &= \int_\Lambda \int_X f(x)\, d\mu(w_\lambda^{-1}(x))\, dp^{(n)}(\lambda) \\ &= \int_\Lambda \int_X f(w_\lambda(x))\, d\mu(x)\, dp^{(n)}(\lambda). \end{aligned}$$

Let $\phi(\lambda) = \int_X f(w_\lambda(x))\, d\mu(x)$. Clearly ϕ is bounded since f is bounded and μ is a probability measure. Let $\varepsilon > 0$ be given. Now both f and w are uniformly continuous in X and Λ. Thus, there is a $\delta > 0$ so that if $d(\lambda, \lambda') < \delta$ then $|f(w_\lambda(x)) - f(w_{\lambda'}(x))| \le \varepsilon$ for all $x \in X$. Therefore, for $\lambda, \lambda' \in \Lambda$ with $d(\lambda, \lambda') < \delta$ we have

$$\begin{aligned} |\phi(\lambda) - \phi(\lambda')| &\le \int_X |f(w_\lambda(x)) - f(w_{\lambda'}(x))|\, d\mu(x) \\ &\le \int_X \varepsilon\, d\mu(x) = \varepsilon \end{aligned}$$

so $\phi \in C^*(\Lambda)$ and

$$\int_\Lambda \phi(\lambda)\, dp^{(n)}(\lambda) \longrightarrow \int_\Lambda \phi(\lambda)\, dp(\lambda).$$

Since this is true for all $f \in C^*(X)$, we know that T_p is a continuous function of p (the distribution on Λ). To get continuity of μ_p (the fixed point of T_p) as a function of p, we need to have a uniform bound for the contraction factor of the family $T_{p^{(n)}}$. However, it suffices to get a bound for sufficiently large n. For fixed $x, y \in X$, we have that $d(w_\lambda(x), w_\lambda(y))$ is a continuous function of λ, so we know that

$$\int_\Lambda d(w_\lambda(x), w_\lambda(y))\, dp^{(n)}(\lambda) \longrightarrow \int_\Lambda d(w_\lambda(x), w_\lambda(y))\, dp(\lambda)$$

and thus the contraction factor of $T_{p^{(n)}}$ converges to the contraction factor of T_p. This gives us our uniform bound s on the contraction factors. Now, by the estimate

$$d_H(\mu_p, \mu_q) \leq \frac{d_H(T_p(\mu_p), T_q(\mu_p))}{1-s}$$

we know that $\mu_{p^{(n)}} \Rightarrow \mu_p$ in the Monge-Kantorovich metric. □

The next result concerns the support of the invariant measure of T_p. We need to assume that w_λ is contractive for all $\lambda \in \Lambda$. For a given $p \in \mathcal{PM}(\Lambda)$, we let $\Omega_p \subset \Lambda$ be the support of the measure p. Then there exists a compact set $A_p \subset X$ which is invariant with respect to $\{w_\lambda | \lambda \in \Omega_p\}$, see [**4**, Theorem 3.2]. Invariant means

$$A_p = \bigcup_{\lambda \in \Omega_p} w_\lambda(A_p).$$

We call A_p the *attractor* of the (infinite) IFS $\{w_\lambda | \lambda \in \Omega_p\}$.

Theorem 3. *If w_λ is contractive for all $\lambda \in \Omega_p$, the support of μ_p is equal to A_p.*

Proof. This proof is a modification of the proof of the finite case in [**1**]. Let B be the support of μ_p so that $B \subset X$ is compact. When we consider $T_p|B : \mathcal{PM}(B) \to \mathcal{PM}(B)$ we get the same fixed point μ_p. Thus the support of μ_p must lie in A_p so $B \subset A_p$.

For the other inclusion, let $a \in A_p$, and let O be an open neighborhood of a. Then, by Corollary 2.8 in [**4**], for any $\theta \in \Sigma = \Pi\Omega_p$ (the space of addresses) which is an address of a, we have

$$w_{\theta_1} \circ w_{\theta_2} \circ \cdots \circ w_{\theta_n}(A) \subset O$$

for large enough n. Since w is continuous as a function of λ, $\bar{w}(\bar{\sigma})() := w_{\sigma_1} \circ w_{\sigma_2} \circ \cdots \circ w_{\sigma_n}()$ is continuous as a function of $\bar{\sigma} \in \Omega_p^n$. Thus, there is a neighborhood U of $\bar{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$ in Ω_p^n so that $\bar{w}(\bar{\sigma})(A) \subset O$ for all $\bar{\sigma} \in U$. Let $\bar{p}$ be the product measure on Ω_p^n given by the measure p on each coordinate. Then

$$\mu_p(O) \geq \bar{p}(U) > 0$$

($\bar{p}(U) > 0$ since Ω_p is the support of μ_p). Since this is true for any such O, a is in the support of μ_p. Thus, A_p is the support of μ_p as claimed. $\square$

We mention that this theorem combined with Theorem 2 does not imply that the support of $\mu_{p^{(n)}}$ converges to the support of μ_p in the Hausdorff metric. The following simple example illustrates this.

Let $\Lambda = \{1, 2\}$ and $w_1(x) = 1/2x$ and $w_2(x) = 1/2x + 1/2$, and let $p^{(n)}(\{1\}) = 1 - 1/n$ and $p^{(n)}(\{2\}) = 1/n$ so p must be the point mass at 1. Then the support of $\mu_{p^{(n)}}$ is all of $[0, 1]$. However, the support of μ_p is just 0.

By Theorem 5.1 in [**4**], if $\operatorname{supp}(p^{(n)}) \to \operatorname{supp}(p)$ in the Hausdorff distance on Λ, then $\operatorname{supp}(\mu_{p^{(n)}}) \to \operatorname{supp}(\mu_p)$ in the Hausdorff distance on X. Conditions on $p^{(n)}$ to insure the convergence of the supports in the Hausdorff distance seem to be unknown.

3. Approximation of integrals. We now turn to the question of calculating integrals with respect to the invariant measure of the operator T. We start with a discussion of the finite case to make the analogy clear.

For f a bounded continuous real-valued function on X and μ the invariant measure of M, we have

$$\begin{aligned}\int_X f(x)\, d\mu(x) &= \lim_n \int_X f(x)\, dM^n(\delta_{x_0})(x) \\ &= \lim_n \sum p_{i_1} p_{i_2} \cdots p_{i_n} f(w_{i_1}(w_{i_2}(\cdots(w_{i_n}(x_0))\cdots)\end{aligned}$$

for any $x_0 \in X$, where the sum is over all possible sequences of length n using the symbols $\{1, 2, \ldots, N\}$. Thus we can approximate the integral of f with respect to μ by enumerating the leaves of an N-ary tree and calculating the sum. If we let n be large enough, we have an approximation to the true integral. In the special case that f is Lipschitz, then we can even have an error bound in terms of the contraction factor of M on $\mathcal{PM}(X)$.

In the infinite case, we have to modify this construction slightly since the image of a point mass under T is not a finite sum of point masses in general. We approximate p and μ simultaneously and take a "diagonal" sequence to approximate the integrals we wish. In the special case where w and f are Lipschitz, we get an error bound on the integral.

We assume that $w_\lambda(\cdot)$ is contractive for all $\lambda \in \Lambda$.

For each $n \in \mathbf{N}$, we generate a measure $p^{(n)} = \sum_i a_i^n \delta_{\lambda_i}$ on Λ with $d_H(p, p^n) < 1/n$. To do this, cover Λ with finitely many disjoint sets A_i^n, each of diameter $< 1/n$ and choose $\lambda_i^n \in A_i^n$. Set

$$p^n = \sum_i p(A_i^n)\delta_{\lambda_i^n}.$$

For each $f \in \mathrm{Lip}\,(\Lambda)$ we have $|f(x) - f(\lambda_i^n)| < 1/n$ for all $x \in A_i^n$, thus (integrating over X) we get $d_H(p, p^n) < 1/n$ as claimed. By Theorem 2, $\mu_{p^{(n)}} \Rightarrow \mu_p$ in the Monge-Kantorovich metric. Notice that if $w_\lambda(\cdot)$ were just contractive on average, then one would have to choose the points λ_i^n more carefully in order to guarantee that $T_{p^{(n)}}$ would be contractive.

Now, choose any $x_0 \in X$. Then $T^k_{p^{(n)}}(\delta_{x_0}) \Rightarrow \mu_{p^{(n)}}$ in the Monge-Kantorovich metric as $k \to \infty$, and this convergence is uniform over n because of the contractivity of $w_\lambda(\cdot)$. Thus, considered as a double sequence, the sequence $T^k_{p^{(n)}}(\delta_{x_0})$ converges so the diagonal sequence converges as well, see [**2**] for a nice discussion of double sequences. Thus $T^n_{p^{(n)}}(\delta_{x_0}) \Rightarrow \mu_p$ so, for all continuous bounded f on X, we have

$$\int_X f(x)\, d(T^m_{p^{(n)}}(\delta_{x_0}))(x) \longrightarrow \int_X f(x)\, d\mu_p(x).$$

For each n, $T^n_{p^{(n)}}(\delta_{x_0})$ is a finite sum of point masses. Thus, we approximate the integral in terms of finite sums.

If we wish to have error bounds on these approximations, we need some further hypothesis on f and w. So, suppose that $f \in \mathrm{Lip}_M(X)$

and $w(\cdot, x) \in \mathrm{Lip}_K(\Lambda)$ for each $x \in X$. Then, for $p, q \in \mathcal{PM}(\Lambda)$ and fixed $\nu \in \mathcal{PM}(X)$, we have

$$\begin{aligned}\int_X f(x)\, d(T_p(\nu) - T_q(\nu)) \\ &= \int_\Lambda \left(\int_X f(w_\lambda(y))\, d\nu(y) \right) d(p(\lambda) - q(\lambda)) \\ &= \int_\Lambda \phi(\lambda)\, d(p(\lambda) - q(\lambda))\end{aligned}$$

where (as before) $\phi(\lambda) = \int_X f(w_\lambda(y))\, d\nu(y)$. For $\lambda, \lambda' \in \Lambda$, we have

$$\begin{aligned}|\phi(\lambda) - \phi(\lambda')| &\le \int_X |f(w_\lambda(y)) - f(w_{\lambda'}(y))|\, d\nu(y) \\ &\le KM\, d(\lambda, \lambda').\end{aligned}$$

Thus, $\phi \in \mathrm{Lip}_{KM}(\Lambda)$. This means that

$$d_H(T_p(\nu), T_q(\nu)) \le KM\, d(p, q).$$

Using this estimate we obtain the estimate

$$d_H(\mu_p, \mu_q) \le \frac{KM}{1-s} d_H(p, q)$$

where s is the maximum of the contraction factors of T_p and t_q. This gives us our desired estimate on the integrals. We have proved the following theorem.

Theorem 4. *if $w(\cdot, x) \in \mathrm{Lip}_K(\Lambda)$ for each $x \in X$ and $f \in \mathrm{Lip}_M(X)$, then*

$$\left| \int_X f(x)\, d(\mu_p - \mu_q) \right| \le \frac{KM}{1-s}\, d_H(p, q).$$

REFERENCES

1. Michael Barnsley, *Fractals everywhere*, Academic Press, New York, 1988.

2. Robert G. Bartle, *The elements of real analysis*, Wiley, New York, 1976.

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

4. Gary B. Lewellen, *Self-similarity*, Rocky Mountain J. Math. **23** (1993), 1023–1040.

School of Mathematics, Georgia Tech., Atlanta, GA 30332

Current address: Applied Mathematics, University of Waterloo, Waterloo, ONT N2L 3G1, Canada
E-mail address: mendivil@augusta.uwaterloo.ca