V-variable fractals:
Fractals with partial self similarity ☆

Michael F. Barnsley [a], John E. Hutchinson [a,*], Örjan Stenflo [b]

[a] *Department of Mathematics, Mathematical Sciences Institute, Australian National University, Canberra, ACT, 0200, Australia*
[b] *Department of Mathematics, Uppsala University, 751 05 Uppsala, Sweden*

Abstract

We establish properties of a new type of fractal which has partial self similarity at all scales. For any collection of iterated functions systems with an associated probability distribution and any positive integer *V* there is a corresponding class of *V*-variable fractal sets or measures. These *V*-variable fractals can also be obtained from the points on the attractor of a single deterministic iterated function system. Existence, uniqueness and approximation results are established under average contractive assumptions. We also obtain extensions of some basic results concerning iterated function systems.
© 2008 Elsevier Inc. All rights reserved.

Keywords: *V*-variable fractal; Superfractal; Fractal; Random fractal; Iterated function system; Chaos game; Markov chain

1. Introduction

A *V*-variable fractal is loosely defined by the fact that it possesses at most *V* distinct local patterns at each level of magnification, where the class of patterns depends on the level, see Definition 18 and Remark 19. Such fractals are useful for modelling purposes and for geometric
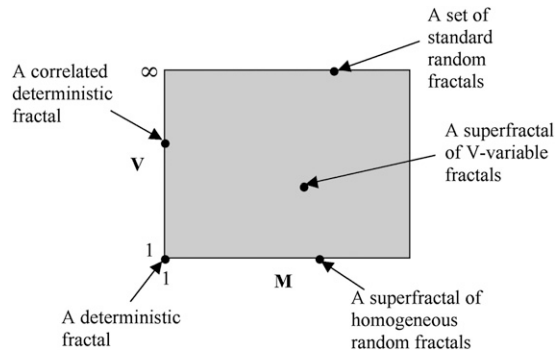
Fig. 1. Sets of V-variable fractals for different V and M. See Remark 55.

applications which require random fractals with a controlled degree of strict self similarity at each scale, see [2, Chapter 5].

Let F be a family of iterated function systems [IFSs] acting on a metric space X, together with an associated probability distribution on F. A construction tree associated to F is a labelled tree with a single root node; corresponding to each node is an IFS from F, and the edges upward from each node are in one-to-one correspondence with the functions in the corresponding IFS. Associated to almost every such construction tree is a fractal set or measure. The construction tree and the corresponding fractals are defined to be V-variable if at each level of the construction tree there are at most V distinct subtrees with base node at that level, see Definition 24. The case $V = 1$ corresponds to homogeneous random fractals. The case $V > 1$ corresponds to a weaker spatial homogeneity requirement that we call V-variability.

An unexpected fact is that the class of V-variable fractal sets or measures associated to F, and an associated natural probability distribution, can be generated by a single deterministic IFS (or Markov chain or "chaos game," depending on one's perspective) operating *not* on the state space X but on the state space $\mathcal{C}(X)^V$ or $\mathcal{M}(X)^V$ of V-tuples of compact subsets of X or probability measures over X, respectively. The projection of the IFS attractor in any of the V coordinate directions gives the class of V-variable fractal sets or measures corresponding to F together with its natural probability distribution in each case. See Theorems 29, 43 and 45, and see Section 8 for a simple example. The full attractor contains further information about the correlation structure of subclasses of these V-variable fractals. The Markov chain converges exponentially, and approximations to its steady state attractor can readily be obtained.

The limit $V \to \infty$ gives standard random fractals, and for this reason the Markov chain for large V provides a fast way of generating classes of standard random fractals together with their probability distributions. Ordinary fractals generated by a single IFS can be seen as special cases of the present construction and this provides new insight into the structure of such fractals, see Remark 55. For the connection with other classes of fractals in the literature see Remarks 39 and 48.

1.1. We now summarise the main notation and results

Let (X, d) be a complete separable metric space. Typically this will be Euclidean space $\mathbb{R}^k$ with the standard metric. For each λ in some index set Λ let F^λ be an IFS acting on (X, d), i.e.

$$F^\lambda = \left(f_1^\lambda, \ldots, f_M^\lambda, w_1^\lambda, \ldots, w_M^\lambda\right), \quad f_m^\lambda : X \to X, \ 0 \leqslant w_m^\lambda \leqslant 1, \ \sum_{m=1}^{M} w_m^\lambda = 1. \tag{1}$$

We will require both the cases where Λ is finite and where Λ is infinite. In order to simplify the exposition we assume that there is only a finite number M of functions in each F^λ and that M does not depend on λ. Let P be a probability distribution on some σ-algebra of subsets of Λ. The given data is then denoted by

$$\boldsymbol{F} = \left\{(X, d), F^\lambda, \lambda \in \Lambda, P\right\}. \tag{2}$$

Let V be a fixed positive integer.

Suppose first that the family Λ is finite and the functions f_m^λ are uniformly contractive. The set $\mathcal{K}_V$ of V-*variable* fractal subsets of X and the set $\mathcal{M}_V$ of V-*variable* fractal measures on X associated to $\boldsymbol{F}$ is then given by Definition 18. There are Markov chains acting on the set $\mathcal{C}(X)^V$ of V-tuples of compact subsets of X and on the set $\mathcal{M}_c(X)^V$ of V-tuples of compactly supported unit mass measures on X, whose stationary distributions project in any of the V coordinate directions to probability distributions $\mathfrak{K}_V$ on $\mathcal{K}_V$ and $\mathfrak{M}_V$ on $\mathcal{M}_V$, respectively. Moreover, these Markov chains are each given by a single deterministic IFS $\mathfrak{F}_V^{\mathcal{C}}$ or $\mathfrak{F}_V^{\mathcal{M}_c}$ constructed from $\boldsymbol{F}$ and acting on $\mathcal{C}(X)^V$ or $\mathcal{M}_c(X)^V$ respectively. The IFS's $\mathfrak{F}_V^{\mathcal{C}}$ and $\mathfrak{F}_V^{\mathcal{M}_c}$ are called *superIFS*'s. The sets $\mathcal{K}_V$ and $\mathcal{M}_V$, and the probability distributions $\mathfrak{K}_V$ and $\mathfrak{M}_V$, are called *superfractals*. See Theorem 43; some of these results were first obtained in [6]. The distributions $\mathfrak{K}_V$ and $\mathfrak{M}_V$ have a complicated correlation structure and differ markedly from other notions of random fractal in the literature. See Remarks 39 and 52.

In many situations one needs an infinite family Λ or needs average contractive conditions, see Example 47. In this case one works with the set $\mathcal{M}_1(X)^V$ of V-tuples of finite first moment unit mass measures on X. The corresponding superIFS $\mathfrak{F}_V^{\mathcal{M}_1}$ is pointwise average contractive by Theorem 45 and one obtains the existence of a corresponding superfractal distribution $\mathfrak{M}_V$. There are technical difficulties in establishing these results, see Remarks 11, 12 and 46.

In Section 2 the properties of the Monge–Kantorovitch and the strong Prokhorov probability metrics are summarised. The strong Prokhorov metric is not widely known in the fractal literature although it is the natural metric to use with uniformly contractive conditions. We include the mass transportation, or equivalently the probabilistic, versions of these metrics, as we need them in Theorems 8, 43, 45 and 49. We work with probability metrics on spaces of measures and such metrics are not always separable. So in Section 2 the extensions required to include non-separable spaces are noted.

In Section 3, particularly Theorem 8 and the following remarks, we summarise and in some cases extend basic results in the literature concerning IFS's, link the measure theoretic and probabilistic approaches and sketch the proofs. We hope this brief survey will be of independent use. In particular, IFS's with a possibly infinite family of functions and pointwise average contractive conditions are considered. The law of large numbers for the corresponding Markov process starting from an arbitrary point, also known as the convergence theorem for the "chaos game" algorithm, is extended to the case when the IFS acts on a non locally compact state space. This situation typically occurs when the state space is a function space or space of measures, and here it is required for the superIFS $\mathfrak{F}_V^{\mathcal{M}_1}$ in Theorem 45. The strong Prokhorov metric is used in the case of uniform contractions. In Section 4 we summarise some of the basic properties of standard random fractals generated by a family of IFS's.

In Section 5 the representation of V-variable fractals in the space Ω_V of tree codes and in the space $\mathcal{A}_V^\infty$ of addresses is developed, and the connection between the two spaces is discussed. The space Ω_V is of the type used for realisations of general random fractals and consists of trees with each node labelled in effect by an IFS, see the comment following Definition 13 and see Definition 18. The space $\mathcal{A}_V^\infty$ is of the type used to address points on a single deterministic fractal, and here consists of infinite sequences of $V \times (M + 1)$ matrices each of which defines a map from the set of V-tuples of sets or measures to itself, see Definition 30. Also see Figs. 2 and 4.

In Section 6 the existence, uniqueness and convergence results for V-variable fractals and superfractals are proved, some examples are given, and the connection with graph directed IFSs is discussed. In Section 7 we establish the rate at which the probability distributions $\mathfrak{K}_V$ and $\mathfrak{M}_V$ converge to the corresponding distributions on standard random fractal sets and measures respectively as $V \to \infty$. In Section 8 a simple example of a super IFS and the associated Markov chain is given. In Section 9 we make some concluding remarks including the relationship with other types of fractals, extensions of the results, and some motivation for the method of construction of V-variable fractals.

The reader may find it easier to begin with Section 4 and refer back to Sections 2 and 3 as needed, particularly in the proofs of Theorems 43 and 45.

2. Preliminaries

Throughout the paper (X, d) denotes a complete separable metric space, except where mentioned otherwise.

Definition 1. The collection of nonempty compact subsets of X with the Hausdorff metric is denoted by $(\mathcal{C}(X), d_\mathcal{H})$. The collection of nonempty bounded closed subsets of X with the Hausdorff metric is denoted by $(\mathcal{BC}(X), d_\mathcal{H})$.

For $A \subset X$ let $A^\epsilon = \{x: d(x, A) \leqslant \epsilon\}$ be the *closed ϵ-neighbourhood* of A.

Both spaces $(\mathcal{C}(X), d_\mathcal{H})$ and $(\mathcal{BC}(X), d_\mathcal{H})$ are complete and separable if (X, d) is complete and separable. Both spaces are complete if (X, d) is just assumed to be complete.

Definition 2 (*Prokhorov metric*). The collection of unit mass Borel (i.e. probability) measures on the Borel subsets of X with the topology of weak convergence is denoted by $\mathcal{M}(X)$. Weak convergence of $\nu_n \to \nu$ means $\int \phi \, d\nu_n \to \int \phi \, d\nu$ for all bounded continuous ϕ. The (standard) *Prokhorov metric* ρ on $\mathcal{M}(X)$ is defined by

$$\rho(\mu, \nu) := \inf\{\epsilon > 0: \ \mu(A) \leqslant \nu(A^\epsilon) + \epsilon \text{ for all Borel sets } A \subset X \}.$$

The Prokhorov metric ρ is complete and separable and induces the topology of weak convergence. Moreover, ρ is complete if (X, d) is just assumed to be complete. See [7, pp. 72, 73]. We will not use the Prokhorov metric but mention it for comparison with the strong Prokhorov metric in Definition 4.

The *Dirac measure* δ_a concentrated at a is defined by $\delta_a(E) = 1$ if $a \in E$ and otherwise $\delta_a(E) = 0$.

If $f : X \to X$ or $f : X \to \mathbb{R}$, the *Lipschitz constant* for f is denoted by Lip f and is defined to be the least L such that $d(f(x), f(y)) \leqslant L d(x, y)$ for all $x, y \in X$.

Definition 3 *(Monge–Kantorovitch metric).* The collection of those $\mu \in \mathcal{M}(X)$ with finite first moment, i.e. those μ such that

$$\int d(a, x) \, d\mu(x) < \infty \tag{3}$$

for some and hence any $a \in X$, is denoted by $\mathcal{M}_1(X)$. The *Monge–Kantorovitch metric* d_{MK} on $\mathcal{M}_1(X)$ is defined in any of the three following equivalent ways:

$$
\begin{aligned}
d_{MK}(\mu, \mu') &:= \sup\left\{ \int f \, d\mu - \int f \, d\mu' \colon \text{Lip } f \leqslant 1 \right\} \\
&= \inf\left\{ \int d(x, y) \, d\gamma(x, y) \colon \gamma \text{ a Borel meas. on } X \times X, \ \pi_1(\gamma) = \mu, \ \pi_2(\gamma) = \mu' \right\} \\
&= \inf\left\{ \mathbb{E} \, d(W, W') \colon \text{dist } W = \mu, \ \text{dist } W' = \mu' \right\}.
\end{aligned}
\tag{4}
$$

The maps $\pi_1, \pi_2 : X \times X \to X$ are the projections onto the first and second coordinates, and so μ and μ' are the marginals for γ. In the third version the infimum is taken over X-valued random variables W and W' with distribution μ and μ' respectively but otherwise unspecified joint distribution.

Here and elsewhere, "dist" denotes the probability distribution on the associated random variable.

The metric space $(\mathcal{M}_1(X), d_{MK})$ is complete and separable. The moment restriction (3) is automatically satisfied if (X, d) is bounded. The second equality in (4) requires proof, see [12, §11.8], but the third form of the definition is just a rewording of the second. The connection between d_{MK} convergence in $\mathcal{M}_1(X)$ and weak convergence is given by

$$\nu_n \xrightarrow{\ d_{MK}\ } \nu \quad \text{iff} \quad \nu_n \to \nu \text{ weakly and } \int d(x, a) \, d\nu_n(x) \to \int d(x, a) \, d\nu(x)$$

for some and hence any $a \in X$. See [37, Section 7.2].

Suppose (X, d) is only assumed to be complete. If measures μ in $\mathcal{M}_1(X)$ are also required to satisfy the condition $\mu(X \setminus \mathrm{spt}\, \mu) = 0$, then $(\mathcal{M}_1(X), d_{MK})$ is complete, see [26] and [16, §2.1.16]. This condition is satisfied for all finite Borel measures μ if X has a dense subset whose cardinality is an Ulam number, and in particular if (X, d) is separable. The requirement that the cardinality of X be an Ulam number is not very restrictive, see [16, §2.2.16].

It is often more natural to use the following strong Prokhorov metric rather than the Monge–Kantorovitch or standard Prokhorov metrics in the case of a uniformly contractive IFS.

Definition 4 *(Strong Prokhorov metric).* The set of compact support, or bounded support, unit mass Borel measures on X is denoted by $\mathcal{M}_c(X)$ or $\mathcal{M}_b(X)$, respectively. The *strong Prokhorov metric* d_P is defined on $\mathcal{M}_b(X)$ in any of the following equivalent ways:

$$d_P(\mu, \mu') := \inf\{\epsilon > 0 \colon \mu(A) \leqslant \mu'(A^\epsilon) \text{ for all Borel sets } A \subset X\}$$

$$= \inf\{\text{ess sup}_\gamma \, d(x, y) \colon \gamma \text{ a meas. on } X \times X, \ \pi_1(\gamma) = \mu, \ \pi_2(\gamma) = \mu'\}$$

$$= \inf\{\text{ess sup} \, d(W, W') \colon \text{dist } W = \mu, \ \text{dist } W' = \mu'\}, \tag{5}$$

where the notation is as in the paragraph following (4).

Note that

$$\mathcal{M}_c(X) \subset \mathcal{M}_b(X) \subset \mathcal{M}_1(X) \subset \mathcal{M}(X).$$

The first definition in (5) is symmetric in μ and μ' by a standard argument, see [12, proof of Theorem 11.3.1]. For discussion and proof of the second equality see [33, Eq. (7.4.15), p. 160] and the other references mentioned there. The third version is a probabilistic reformulation of the second.

Proposition 5. *$(\mathcal{M}_c(X), d_P)$ and $(\mathcal{M}_b(X), d_P)$ are complete. If $\nu, \nu' \in \mathcal{M}_b(X)$ then*

$$d_{\mathcal{H}}(\text{spt } \nu, \text{spt } \nu') \leqslant d_P(\nu, \nu'), \qquad d_{MK}(\nu, \nu') \leqslant d_P(\nu, \nu').$$

In particular, $\nu_k \to \nu$ in the d_P metric implies $\nu_k \to \nu$ in the d_{MK} metric and $\text{spt } \nu_k \to \text{spt } \nu$ in the $d_{\mathcal{H}}$ metric.

Proof. The first inequality follows directly from the definition of the Hausdorff metric and the second from the final characterisations in (4) and (5).

Completeness of $\mathcal{M}_b(X)$ can be shown as follows and this argument carries across to $\mathcal{M}_c(X)$. Completeness of $\mathcal{M}_c(X)$ is also shown in [15, Theorem 9.1].

Suppose $(\nu_k)_{k \geqslant 1} \subseteq (\mathcal{M}_b(X), d_P)$ is d_P-Cauchy. It follows that $(\nu_k)_{k \geqslant 1}$ is d_{MK}-Cauchy and hence converges to some measure ν in the d_{MK} sense and in particular weakly. Moreover, $\text{spt}(\nu_k)_{k \geqslant 1}$ converges to some bounded closed set K in the Hausdorff sense, hence $\text{spt } \nu \subset K$ using weak convergence, and so $\nu \in \mathcal{M}_b(X)$. Suppose $\epsilon > 0$ and using the fact $(\nu_k)_{k \geqslant 1}$ is d_P-Cauchy choose J so $k, j \geqslant J$ implies $\nu_k(A) \leqslant \nu_j(A^\epsilon)$ for all Borel $A \subset X$. By weak convergence and because A^ϵ is closed, $\limsup_{j \to \infty} \nu_j(A^\epsilon) \leqslant \nu(A^\epsilon)$ and so $\nu_k(A) \leqslant \nu(A^\epsilon)$ if $k \geqslant J$. Hence $\nu_k \to \nu$ in the d_P sense. $\quad\square$

If (X, d) is only assumed to be complete, but measures in $\mathcal{M}_c(X)$ and $\mathcal{M}_b(X)$ are also required to satisfy the condition $\mu(X \setminus \text{spt } \mu) = 0$ as discussed following Definition 3, then the same proof shows that Proposition 5 is still valid. The main point is that one still has completeness of $(\mathcal{M}_1(X), d_{MK})$.

Remark 6 *(Strong and standard Prokhorov metrics).* Convergence in the strong Prokhorov metric is a much stronger requirement than convergence in the standard Prokhorov metric or the Monge–Kantorovitch metric. A simple example is given by $X = [0, 1]$ and $\nu_n = (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_1$. Then $\nu_n \to \delta_0$ weakly and in the d_{MK} and ρ metrics, but $d_P(\nu_n, \delta_0) = 1$ for all n.

The strong Prokhorov metric is normally not separable. For example, if $\mu_x = x\delta_0 + (1 - x)\delta_1$ for $0 < x < 1$ then $d_P(\mu_x, \mu_y) = 1$ for $x \neq y$. So there is no countable dense subset.

If $f : X \to X$ is Borel measurable then the *pushforward measure* $f(\nu)$ is defined by $f(\nu)(A) = \nu(f^{-1}(A))$ for Borel sets A. The *scaling property* for Lipschitz functions f, namely

$$d_P\big(f(\mu), f(\nu)\big) \leqslant \text{Lip } f \ d_P(\mu, \nu), \tag{6}$$

follows from the definition of d_P. Similar properties are well known and easily established for the Hausdorff and Monge–Kantorovitch metrics.

3. Iterated function systems

Definition 7. An *iterated functions system* [IFS] $F = (X, f_\theta, \theta \in \Theta, W)$ is a set of maps $f_\theta : X \to X$ for $\theta \in \Theta$, where (X, d) is a *complete separable* metric space and W is a probability measure on some σ-algebra of subsets of Θ. The map $(x, \theta) \mapsto f_\theta(x) : X \times \Theta \to X$ is measurable with respect to the product σ-algebra on $X \times \Theta$, using the Borel σ-algebra on X. If $\Theta = \{1, \ldots, M\}$ is finite and $W(m) = w_m$ then one writes $F = (X, f_1, \ldots, f_M, w_1, \ldots, w_M)$.

It follows f is measurable in θ for fixed x and in x for fixed θ. Notation such as $\mathbb{E}_\theta$ is used to denote taking the expectation, i.e. integrating, over the variable θ with respect to W.

Sometimes we will need to work with an IFS on a nonseparable metric space. The properties which still hold in this case will be noted explicitly. See Remark 12 and Theorem 43.

The IFS F acts on subsets of X and Borel measures over X, for finite and infinite Θ respectively, by

$$F(E) = \bigcup_{m=1}^{M} f_m(E), \qquad F(\nu) = \sum_{m=1}^{M} w_m f_m(\nu),$$

$$F(E) = \bigcup_{\theta} f_\theta(E), \qquad F(\nu) = \int dW(\theta) f_\theta(\nu). \tag{7}$$

We put aside measurability matters and interpret the integral formally as the measure which operates on any Borel set $A \subset X$ to give $\int dW(\theta)(f_\theta(\nu))(A)$. The latter is thought of as a weighted sum via $W(\theta)$ of the measures $f_\theta(\nu)$. The precise definition in the cases we need for infinite Θ is given by (9).

If $F(E) = E$ or $F(\nu) = \nu$ then E or ν respectively is said to be *invariant* under the IFS F.

In the study of fractal geometry one is usually interested in the case of an IFS with a finite family Θ of maps. If $X = \mathbb{R}^2$ then compact subsets of X are often identified with black and white images, while measures are identified with greyscale images. Images are generated by iterating the map F to approximate $\lim_{k \to \infty} F^k(E_0)$ or $\lim_{k \to \infty} F^k(\nu_0)$. As seen in the following theorem, under natural conditions the limits exist and are independent of the starting set E_0 or measure ν_0.

In the study of Markov chains on an arbitrary state space X via iterations of random functions on X, it is usually more natural to consider the case of an infinite family Θ. One is concerned with a random process Z_n^x, in fact a Markov chain, with initial state $x \in X$, and

$$Z_0^x(\boldsymbol{i}) = x, \qquad Z_n^x(\boldsymbol{i}) := f_{i_n}\big(Z_{n-1}^x(\boldsymbol{i})\big) = f_{i_n} \circ \cdots \circ f_{i_1}(x) \quad \text{if } n \geqslant 1, \tag{8}$$

where the $i_n \in \Theta$ are independently and identically distributed [iid] with probability distribution W and $\boldsymbol{i} = i_1 i_2 \dots$. The *induced probability measure on the set of codes* $\boldsymbol{i}$ is also denoted by W.

Note that the probability $P(x, B)$ of going from $x \in X$ into $B \subset X$ in one iteration is $W\{\theta: f_\theta(x) \in B\}$, and $P(x, \cdot) = \operatorname{dist} Z_1^x$. More generally, if the starting state is given by a random variable X_0 independent of $\boldsymbol{i}$ with $\operatorname{dist} X_0 = \nu$, then one defines the random variable $Z_n^\nu(\boldsymbol{i}) = Z_n^{X_0}(\boldsymbol{i})$. The sequence $(Z_n^\nu(\boldsymbol{i}))_{n \geqslant 0}$ forms a Markov chain starting according to ν. We define $F(\nu) = \operatorname{dist} Z_1^\nu$ and in summary we have

$$\nu := \operatorname{dist} Z_0^\nu, \qquad F(\nu) := \operatorname{dist} Z_1^\nu, \qquad F^n(\nu) = \operatorname{dist} Z_n^\nu. \tag{9}$$

The operator F can be applied to bounded continuous functions $\phi : X \to \mathbb{R}$ via any of the following equivalent definitions:

$$\big(F(\phi)\big)(x) = \int \phi\big(f_\theta(x)\big) \, dW(\theta) \quad \bigg(\text{or } \sum_m w_m \phi\big(f_m(x)\big)\bigg)$$

$$= \mathbb{E}_\theta \, \phi\big(f_\theta(x)\big) = \mathbb{E} \, \phi\big(Z_1^x\big). \tag{10}$$

In the context of Markov chains, the operator F acting on functions is called the *transfer operator*. It follows from the definitions that $\int F(\phi) \, d\mu = \int \phi \, d(F\mu)$, which is the expected value of ϕ after one time step starting with the initial distribution μ. If one assumes $F(\phi)$ is continuous (it is automatically bounded) then F acting on measures is the *adjoint* of F acting on bounded continuous functions. Such F are said to satisfy the *weak Feller property*—this is the case if all f_θ are continuous by the dominated convergence theorem, or if the pointwise average contractive condition is satisfied, see [35].

We will need to apply the maps in (8) in the reverse order. Define

$$\widehat{Z}_0^x(\boldsymbol{i}) = x, \qquad \widehat{Z}_n^x(\boldsymbol{i}) = f_{i_1} \circ \cdots \circ f_{i_n}(x) \quad \text{if } n \geqslant 1. \tag{11}$$

Then from the iid property of the i_n it follows

$$F^n(\nu) = \operatorname{dist} Z_n^\nu = \operatorname{dist} \widehat{Z}_n^\nu, \qquad F^n(\phi)(x) = \mathbb{E} \, \phi\big(Z_n^x\big) = \mathbb{E} \, \phi\big(\widehat{Z}_n^x\big). \tag{12}$$

However, the pathwise behaviour of the processes Z_n^x and $\widehat{Z}_n^x$ are very different. Under suitable conditions the former is ergodic and the latter is a.s. convergent, see Theorem 8(c) and (a) respectively, and the discussion in [9].

The following Theorem 8 is known with perhaps two exceptions: the lack of a local compactness requirement in (c) and the use of the strong Prokhorov metric in (d).

The strong Prokhorov metric, first used in the setting of random fractals in [15], is a more natural metric than the Monge–Kantorovitch metric when dealing with uniformly contractive maps and fractal measures in compact spaces. Either it or variants may be useful in image compression matters. In Theorem 43 we use it to strengthen the convergence results in [6].

The pointwise ergodic theorems for Markov chains for *any* starting point as established in [5,8,13,14,31] require compactness or local compactness, see Remark 11. We remove this restriction in Theorem 8. The result is needed in Theorem 45 where we consider an IFS operating

on the space $(\mathcal{M}_1(X)^V, d_{MK})$ of V-tuples of probability measures. Like most spaces of functions or measures, this space is not locally compact even if $X = \mathbb{R}^k$. See Remark 11 and also Remark 47.

We assume a pointwise average contractive condition, see Remark 9. We need this in Theorem 45, see Remark 46.

The parts of the theorem have a long history. In the Markov chain literature the contraction conditions (13) and (19) were introduced in [24] and [11] respectively in order to establish ergodicity. In the fractal geometry literature, following [27,28], the existence and uniqueness of attractors, their properties, and the Markov chain approach to generating fractals, were introduced in [3,4,10,13,14,20]. See [9,34,35] for further developments and more on the history.

Theorem 8. *Let $F = (X, f_\theta, \theta \in \Theta, W)$ be an IFS on a complete separable metric space (X, d). Suppose F satisfies the pointwise average contractive and average boundedness conditions*

$$\mathbb{E}_\theta \, d\big(f_\theta(x), f_\theta(y)\big) \leqslant r d(x, y) \quad and \quad L := \mathbb{E}_\theta \, d\big(f_\theta(a), a\big) < \infty \tag{13}$$

for some fixed $0 < r < 1$, all $x, y \in X$, and some $a \in X$.

(a) *For some function Π, all $x, y \in X$ and all n, we have*

$$\mathbb{E} \, d\big(Z_n^x(\boldsymbol{i}), Z_n^y(\boldsymbol{i})\big) = \mathbb{E} \, d\big(\widehat{Z}_n^x(\boldsymbol{i}), \widehat{Z}_n^y(\boldsymbol{i})\big) \leqslant r^n d(x, y),$$
$$\mathbb{E} \, d\big(\widehat{Z}_n^x(\boldsymbol{i}), \Pi(\boldsymbol{i})\big) \leqslant \gamma_x r^n, \tag{14}$$

where $\gamma_x = \mathbb{E}_\theta \, d(x, f_\theta(x)) \leqslant d(x, a) + L/(1 - r)$ and $\Pi(\boldsymbol{i})$ is independent of x. The map Π is called the address map *from code space into X. Note that Π is defined only a.e. If $r < s < 1$ then for all $x, y \in X$ for a.e. $\boldsymbol{i} = i_1 i_2 \ldots i_n \ldots$ there exists $n_0 = n_0(\boldsymbol{i}, s)$ such that*

$$d\big(Z_n^x(\boldsymbol{i}), Z_n^y(\boldsymbol{i})\big) \leqslant s^n d(x, y), \qquad d\big(\widehat{Z}_n^x(\boldsymbol{i}), \widehat{Z}_n^y(\boldsymbol{i})\big) \leqslant s^n d(x, y),$$
$$d\big(\widehat{Z}_n^x(\boldsymbol{i}), \Pi(\boldsymbol{i})\big) \leqslant \gamma_x s^n, \tag{15}$$

for $n \geqslant n_0$.

(b) *If ν is a unit mass Borel measure then $F^n(\nu) \to \mu$ weakly where $\mu := \Pi(W)$ is the projection of the measure W on code space onto X via Π. Equivalently, μ is the distribution of Π regarded as a random variable. In particular,*

$$F(\mu) = \mu$$

and μ is the unique invariant unit mass measure.
The map F is a contraction on $(\mathcal{M}_1(X), d_{MK})$ with Lipschitz constant r. Moreover, $\mu \in \mathcal{M}_1(X)$ and

$$d_{MK}\big(F^n(\nu), \mu\big) \leqslant r^n d_{MK}(\nu, \mu) \tag{16}$$

for every $\nu \in \mathcal{M}_1(X)$.

(c) *For all $x \in X$ and for a.e. $\boldsymbol{i}$, the empirical measure (probability distribution)*

$$\mu_n^x(\boldsymbol{i}) := \frac{1}{n} \sum_{k=1}^{n} \delta_{Z_k^x(\boldsymbol{i})} \to \mu \tag{17}$$

weakly. Moreover, if A is the support of μ then there exists $n_0 = n_0(\boldsymbol{i}, x, \epsilon)$ such that

$$Z_n^x(\boldsymbol{i}) \subset A^{\epsilon} \quad \text{if } n \geqslant n_0. \tag{18}$$

(d) *Suppose F satisfies the uniform contractive and uniform boundedness conditions*

$$\sup_\theta d\big(f_\theta(x), f_\theta(y)\big) \leqslant r\, d(x, y) \quad \text{and} \quad L := \sup_\theta d\big(f_\theta(a), a\big) < \infty \tag{19}$$

for some $r < 1$, all $x, y \in X$, and some $a \in X$. Then

$$d\big(Z_n^x(\boldsymbol{i}), Z_n^y(\boldsymbol{i})\big) \leqslant r^n\, d(x, y), \qquad d\big(\widehat{Z}_n^x(\boldsymbol{i}), \widehat{Z}_n^y(\boldsymbol{i})\big) \leqslant r^n\, d(x, y),$$
$$d\big(\widehat{Z}_n^x(\boldsymbol{i}), \Pi(\boldsymbol{i})\big) \leqslant \gamma_x r^n, \tag{20}$$

for all $x, y \in X$ and all $\boldsymbol{i}$. The address map Π is everywhere defined and is continuous with respect to the product topology defined on code space and induced from the discrete metric on Θ. Moreover, $\mu = \Pi(W) \in \mathcal{M}_c$ and for any $\nu \in \mathcal{M}_b$,

$$d_P\big(F^n(\nu), \mu\big) \leqslant r^n d_P(\nu, \mu). \tag{21}$$

Suppose in addition Θ is finite and $W(\{\theta\}) > 0$ for $\theta \in \Theta$. Then A is compact and for any closed bounded E

$$d_{\mathcal{H}}\big(\big(F^n(E), A\big) \leqslant r^n d_{\mathcal{H}}(E, A). \tag{22}$$

Moreover, $F(A) = A$ and A is the unique closed bounded invariant set.

Proof. (a) The first inequality in (14) follows from (8), (11) and contractivity.

Next fix x. Since

$$\mathbb{E}\, d\big(\widehat{Z}_{n+1}^x(\boldsymbol{i}), \widehat{Z}_n^x(\boldsymbol{i})\big) \leqslant r\, \mathbb{E}\, d\big(\widehat{Z}_n^x(\boldsymbol{i}), \widehat{Z}_{n-1}^x(\boldsymbol{i})\big),$$
$$\mathbb{E}\, d\big(\widehat{Z}_{n+1}^a(\boldsymbol{i}), \widehat{Z}_n^a(\boldsymbol{i})\big) \leqslant r\, \mathbb{E}\, d\big(\widehat{Z}_n^a(\boldsymbol{i}), \widehat{Z}_{n-1}^a(\boldsymbol{i})\big) \tag{23}$$

for all n, it follows that $\widehat{Z}_n^x(\boldsymbol{i})$ and $\widehat{Z}_n^a(\boldsymbol{i})$ a.s. converge exponentially fast to the same limit $\Pi(\boldsymbol{i})$ (say) by the first inequality in (14). It also follows that (23) is simultaneously true with $\boldsymbol{i}$ replaced by $i_{k+1}i_{k+2}i_{k+3}, \ldots$ for every k. It then follows from (11) that

$$\Pi(\boldsymbol{i}) = f_{i_1} \circ \cdots \circ f_{i_k}\big(\Pi(i_{k+1}i_{k+2}i_{k+3}\ldots)\big).$$

Again using (11),

$$\mathbb{E}\,d\big(\widehat{Z}_n^x(\boldsymbol{i}),\,\Pi(\boldsymbol{i})\big) \leqslant r^n\,\mathbb{E}\,d\big(x,\,\Pi(i_{n+1}i_{n+2}i_{n+3}\ldots)\big)$$
$$= r^n\,\mathbb{E}\,d\big(x,\,\Pi(\boldsymbol{i})\big) \leqslant r^n\big(d(x,a)+\mathbb{E}\,d\big(a,\,\Pi(\boldsymbol{i})\big)\big).$$

But

$$\mathbb{E}\,d\big(a,\,\Pi(\boldsymbol{i})\big) \leqslant \mathbb{E}\,d\big(a,\,f_{i_1}(a)\big) + \sum_{n\geqslant 1}\mathbb{E}\,d\big(f_{i_1}\circ\cdots\circ f_{i_n}(a),\,f_{i_1}\circ\cdots\circ f_{i_{n+1}}(a)\big)$$

$$\leqslant \sum_{n\geqslant 0} r^n L = \frac{L}{1-r}.$$

This gives the second inequality in (14). See [34] for details.

The estimates in (15) are the standard consequence that exponential convergence in mean implies a.s. exponential convergence.

(b) Suppose $\phi \in \mathcal{BC}(X,d)$, the set of bounded continuous functions on X. Let ν be any unit mass measure. Since for a.e. $\boldsymbol{i}$, $\widehat{Z}_n^x(\boldsymbol{i}) \to \Pi(\boldsymbol{i})$ for every x, using the continuity of ϕ and dominated convergence,

$$\int \phi\,d\big(F^n\nu\big) = \int \phi\,d\big(\text{dist}\,\widehat{Z}_n^\nu\big)\text{(by (12))} = \int \phi\big(\widehat{Z}_n^x(\boldsymbol{i})\big)\,dW(\boldsymbol{i})\,d\nu(x)$$
$$\to \int \phi\big(\Pi(\boldsymbol{i})\,dW(\boldsymbol{i})\,d\nu = \int \phi\,d\mu,$$

by the definition of μ for the last equality. Thus $F^n(\nu) \to \mu$ weakly. The invariance of μ and the fact μ is the unique invariant unit measure follow from the weak Feller property.

One can verify that $F : \mathcal{M}_1(X) \to \mathcal{M}_1(X)$ and F is a contraction map with Lipschitz constant r in the d_{MK} metric. It is easiest to use the second or third form of (4) for this. The rest of (b) now follows.

(c) The main difficulty here is that (X,d) may not be locally compact and so the space $\mathcal{BC}(X,d)$ need not be separable, see Remark 11. We adapt an idea of Varadhan, see [12, Theorem 11.4.1, p. 399].

There is a totally bounded and hence separable, but not usually complete, metric e on X such that (X,e) and (X,d) have the same topology, see [12, Theorem 2.8.2, p. 72]. Moreover, as the proof there shows, $e(x,y) \leqslant d(x,y)$. Because the topology is preserved, weak convergence of measures on (X,d) is the same as weak convergence on (X,e).

Let $\mathcal{BL}(X,e)$ denote the set of bounded Lipschitz functions over (X,e). Then $\mathcal{BL}(X,e)$ is separable in the *sup norm* from the total boundedness of e. Suppose $\phi \in \mathcal{BL}(X,e)$. By the ergodic theorem, since μ is the unique invariant measure for F,

$$\int \phi\,d\mu_n^y(\boldsymbol{i}) = \frac{1}{n}\sum_{k=1}^{n}\phi\big(Z_k^y(\boldsymbol{i})\big) \to \int \phi\,d\mu \tag{24}$$

for a.e. $\boldsymbol{i}$ and μ a.e. y.

Suppose $x \in X$ and choose $y \in X$ such that (24) is true. Using (a), for a.e. $\boldsymbol{i}$

$$e\big(Z_n^x(\boldsymbol{i}),\,Z_n^y(\boldsymbol{i})\big) \leqslant d\big(Z_n^x(\boldsymbol{i}),\,Z_n^y(\boldsymbol{i})\big) \to 0.$$

It follows from (24) and the uniform continuity of ϕ in the e metric that for a.e. $\boldsymbol{i}$,

$$\int \phi \, d\mu_n^x(\boldsymbol{i}) = \frac{1}{n} \sum_{k=1}^n \phi\big(Z_k^x(\boldsymbol{i})\big) \to \int \phi \, d\mu. \tag{25}$$

Let $\mathcal{S}$ be a countable dense subset of $\mathcal{BL}(X, e)$ in the sup norm. One can ensure that (25) is simultaneously true for all $\phi \in \mathcal{S}$. By an approximation argument it follows (25) is simultaneously true for all $\phi \in \mathcal{BL}(X, e)$.

Since (X, e) is separable, weak convergence of measures $\nu_n \to \nu$ is equivalent to $\int \phi \, d\nu_n \to \int \phi \, d\nu$ for all $\phi \in \mathcal{BL}(X, e)$, see [12, Theorem 11.3.3, p. 395]. Completeness is not needed for this. It follows that $\mu_n^x(\boldsymbol{i}) \to \mu$ weakly as required.

The result (18) follows from the third inequality in (15).

(d) The three inequalities in (20) are straightforward as are the claims concerning Π.

It follows readily from the definitions that each of $\mathcal{M}_c(X)$, $\mathcal{M}_b(X)$, $\mathcal{C}(X)$ and $\mathcal{BC}(X)$ are closed under F, and that F is a contraction map with respect to d_P in the first two cases and $d_{\mathcal{H}}$ in the second two cases. The remaining results all follow easily. $\quad\square$

Remark 9 *(Contractivity conditions).* The *global* average contractive condition $\mathbb{E}_\theta r_\theta := \int r_\theta \, dW(\theta) \leqslant r$ where $r_\theta := \mathrm{Lip}\, f_\theta$, implies the *pointwise* average contractive condition. Although the global condition is frequently assumed, for our purposes the weaker pointwise assumption is necessary, see Remark 46.

In some papers, for example [9,38], parts of Theorem 8 are established or used under the *global log average contractive* and average boundedness conditions

$$\mathbb{E}_\theta \log r_\theta < 0, \qquad \mathbb{E}_\theta r_\theta^q < \infty, \qquad \mathbb{E}_\theta d^q\big(a, f_\theta(a)\big) < \infty, \tag{26}$$

for some $q > 0$ and some $a \in X$. However, since d^q is a metric for $0 < q < 1$ and since $(\mathbb{E}_\theta g^q(\theta))^{1/q} \downarrow \exp(\mathbb{E}_\theta \log g(\theta))$ as $q \downarrow 0$ for $g \geqslant 0$, such results follow from Theorem 8. In the main Theorem 5.2 of [9] the last two conditions are replaced by the equivalent algebraic tail condition. One can even obtain in this way similar consequences under the yet weaker *pointwise log average conditions*. See also [13].

Pointwise average contractivity is a much weaker requirement than global average contractivity. A simple example in which f_θ is discontinuous with positive probability is given by $0 < \epsilon < 1$, $X = [0, 1]$ with the standard metric d, $\Theta = [0, 1]$, $W\{0\} = W\{1\} = \epsilon/2$, and otherwise W is uniformly distributed over $(0, 1)$ according to $W\{(a, b)\} = (1 - \epsilon)(b - a)$ for $0 < a < b < 1$. Let $f_\theta = \mathcal{X}_{[\theta, 1]}$ be the characteristic function of $[\theta, 1]$ for $0 \leqslant \theta < 1$ and let $f_1 \equiv 0$ be the zero function. Then $\mathbb{E}_\theta \, d(f_\theta(x), f_\theta(y)) \leqslant (1 - \epsilon)d(x, y)$ for all x and y. The unique invariant measure is of course $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Uniqueness fails if $\epsilon = 0$. A simple example where f_θ is discontinuous with probability one is $f_\theta = \mathcal{X}_{\{1\}}$ and θ is chosen uniformly on the unit interval. See [34,35] for further examples and discussion.

Remark 10 *(Alternative starting configurations).* One can extend (14) by allowing the starting point x to be distributed according to a distribution ν and considering the corresponding random variables $\widehat{Z}_n^\nu(\boldsymbol{i})$. Then (16) follows directly by using the random variables $\widehat{Z}_n^\nu(\boldsymbol{i})$ and $\widehat{Z}_n^\mu(\boldsymbol{i})$ in the third form of (4). Analogous remarks apply in deducing the distributional convergence results in Theorem 8(d) from the pointwise convergence results, via the third form of (5).

Remark 11 *(Local compactness issues).* Versions of Theorem 8(c) for locally compact (X, d) were established in [5,8,13,14,31]. In that case one first proves vague convergence in (17). By $\nu_n \to \nu$ vaguely one means $\int \phi \, d\nu_n \to \int \phi \, d\nu$ for all $\phi \in \mathcal{C}_c(X)$ where $\mathcal{C}_c(X)$ is the set of compactly supported continuous functions $\phi : X \to \mathbb{R}$. The proof of vague convergence is straightforward from the ergodic theorem since $\mathcal{C}_c(X)$ is separable. Moreover, in a locally compact space, vague convergence of probability measures to a probability measure implies weak convergence. That this is not true for more general spaces is a consequence of the following discussion.

The extension of Theorem 8(c) to nonlocally compact spaces is needed in Section 6 and Theorem 45. In order to study images in $\mathbb{R}^k$ we consider IFS's whose component functions act on the space $(\mathcal{M}_1(\mathbb{R}^k)^V, d_{MK})$ of V-tuples of unit mass measures over $\mathbb{R}^k$, where V is a natural number. Difficulties already arise in proving that the chaos game converges a.s. from every initial V-tuple of sets even for $V = k = 1$.

To see this suppose $\nu_0 \in \mathcal{M}_1(\mathbb{R})$ and $\epsilon > 0$. Then $B_\epsilon(\nu_0) := \{\nu : d_{MK}(\nu, \nu_0) \leqslant \epsilon\}$ is not sequentially compact in the d_{MK} metric and so $(\mathcal{M}_1(\mathbb{R}), d_{MK})$ is not locally compact. To show sequential compactness does not hold let $\nu_n = (1 - \frac{\epsilon}{n})\nu_0 + \frac{\epsilon}{n}\tau_n \nu_0$, where $\tau_n(x) = x + n$ is translation by n units in the x-direction. Then clearly $\nu_n \to \nu_0$ weakly. Setting $f(x) = x$ in (4),

$$d_{MK}(\nu_n, \nu_0) \geqslant \int x \, d\nu_n - \int x \, d\nu_0 = \left(1 - \frac{\epsilon}{n}\right) \int x \, d\nu_0 + \frac{\epsilon}{n} \int (x + n) \, d\nu_0 - \int x \, d\nu_0 = \epsilon.$$

On the other hand, let W be a random measure with dist $W = \nu_0$. Independently of the value of W let $W' = W$ with probability $1 - \frac{\epsilon}{n}$ and $W' = \tau_n W$ with probability $\frac{\epsilon}{n}$. Then again from (4),

$$d_{MK}(\nu_n, \nu_0) \leqslant \mathbb{E} \, d_{MK}(W, W') = \left(1 - \frac{\epsilon}{n}\right) \times 0 + \frac{\epsilon}{n} n = \epsilon.$$

It follows that $\nu_n \in B_\epsilon(\nu_0)$ and $\nu_n \nrightarrow \nu$ in the d_{MK} metric, nor does any subsequence. Since d_{MK} implies weak convergence it follows that $(\nu_n)_{n \geqslant 1}$ has no convergent subsequence in the d_{MK} metric.

It follows that $\mathcal{C}_c(\mathcal{M}_1(\mathbb{R}), d_{MK})$ contains only the zero function. Hence vague convergence in this setting is a vacuous notion and gives no information about weak convergence.

Finally, we note that although (c) is proved here assuming the pointwise average contractive condition, it is clear that weaker hypotheses concerning the stability of trajectories will suffice to extend known results from the locally compact setting.

Remark 12 *(Separability and measurability issues).* If (X, d) is separable then the class of Borel sets for the product topology on $X \times X$ is the product σ-algebra of the class of Borel sets on X with itself, see [7, p. 244]. It follows that $\theta \mapsto d(f_\theta(x), f_\theta(y))$ is measurable for each $x, y \in X$ and so the quantities in (13) are well defined.

Separability is not required for the uniform contractive and uniform boundedness conditions in (19) and the conclusions in (d) are still valid with essentially the same proofs. The spaces $\mathcal{M}_c(X), \mathcal{M}_b(X)$ and $\mathcal{M}_1(X)$ need to be restricted to separable measures as discussed following Definition (3) and Proposition 5.

Separability is also used in the proof of (c). If one drops this condition and assumes the uniform contractive and boundedness conditions (19) then a weaker version of (c) holds. Namely, for every $x \in X$ and every bounded continuous $\phi \in \mathcal{BC}(X)$, for a.e. $\boldsymbol{i}$

$$\int \phi \, d\mu_n^x(\boldsymbol{i}) \to \int \phi \, d\mu. \tag{27}$$

The point is that unlike the situation in (c) under the hypothesis of separability, the set of such $\boldsymbol{i}$ might depend on the function ϕ.

In Theorem 43 we apply Theorem 8 to an IFS whose component functions operate on $(\mathcal{M}_c(X)^V, d_P)$ where V is a natural number. Even in the case $V = 1$ this space is not separable, see Example 6.

4. Tree codes and standard random fractals

Let $\boldsymbol{F} = \{X, F^\lambda, \lambda \in \Lambda, P\}$ be a family of IFSs as in (1) and (2). We assume the IFSs F^λ are *uniformly contractive* and *uniformly bounded*, i.e. for some $0 < r < 1$,

$$\sup_\lambda \max_m d\big(f_m^\lambda(x), f_m^\lambda(y)\big) \leqslant r \, d(x, y),$$
$$L := \sup_\lambda \max_m d\big(f_m^\lambda(a), a\big) < \infty \tag{28}$$

for all $x, y \in X$ and some $a \in X$. More general conditions are assumed in Section 6.

We often use $*$ to indicate concatenation of sequences, either finite or infinite.

Definition 13 *(Tree codes).* The tree T is the set of all finite sequences from $\{1, \ldots, M\}$, including the empty sequence $\emptyset$. If $\sigma = \sigma_1 \ldots \sigma_k \in T$ then the *length* of σ is $|\sigma| = k$ and $|\emptyset| = 0$.

A *tree code* ω is a map $\omega : T \to \Lambda$. The metric space (Ω, d) of all tree codes is defined by

$$\Omega = \{\omega \mid \omega : T \to \Lambda\}, \quad d(\omega, \omega') = \frac{1}{M^k} \tag{29}$$

if $\omega(\sigma) = \omega'(\sigma)$ for all σ with $|\sigma| < k$ and $\omega(\sigma) \neq \omega'(\sigma)$ for some σ with $|\sigma| = k$. A *finite tree code of height k* is a map $\omega : \{\sigma \in T : |\sigma| \leqslant k\} \to \Lambda$.

If $\omega \in \Omega$ and $\tau \in T$ then the tree code $\omega \lfloor \tau$ is defined by $(\omega \lfloor \tau)(\sigma) := \omega(\tau * \sigma)$. It is the tree code obtained from ω by starting at the node τ. One similarly defines $\omega \lfloor \tau$ if ω is a finite tree code of height k and $|\tau| \leqslant k$.

If $\omega \in \Omega$ and k is a natural number then the finite tree code $\omega \lfloor k$ is defined by $(\omega \lfloor k)(\sigma) = \omega(\sigma)$ for $|\sigma| \leqslant k$. It is obtained by truncating ω at the level k.

The space (Ω, d) is complete and bounded. If Λ is finite then (Ω, d) is compact.

The tree code ω associates to each node $\sigma \in T$ the IFS $F^{\omega(\sigma)}$. It also associates to each $\sigma \neq \emptyset$ the function $f_m^{\omega(\sigma')}$ where $\sigma = \sigma' * m$. The M^k components of K_k^ω in (30) are then obtained by beginning with the set K_0 and iteratively applying the functions associated to the k nodes along each of the M^k branches of depth k obtained from ω. A similar remark applies to μ_k^ω.

Definition 14 *(Fractal sets and measures).* If $K_0 \in \mathcal{C}(X)$ and $\mu_0 \in \mathcal{M}_c(X)$ then the *prefractal sets* K_k^ω, the *prefractal measures* μ_k^ω, the *fractal set* K^ω and the *fractal measure* μ^ω, are given by

$$K_k^\omega = \bigcup_{\sigma \in T, |\sigma|=k} f_{\sigma_1}^{\omega(\emptyset)} \circ f_{\sigma_2}^{\omega(\sigma_1)} \circ f_{\sigma_3}^{\omega(\sigma_1\sigma_2)} \circ \cdots \circ f_{\sigma_k}^{\omega(\sigma_1\ldots\sigma_{k-1})}(K_0),$$

$$\mu_k^\omega = \sum_{\sigma \in T, |\sigma|=k} w_{\sigma_1}^{\omega(\emptyset)} w_{\sigma_2}^{\omega(\sigma_1)} \cdot \cdots \cdot w_{\sigma_k}^{\omega(\sigma_1\ldots\sigma_{k-1})} f_{\sigma_1}^{\omega(\emptyset)} \circ f_{\sigma_2}^{\omega(\sigma_1)} \circ \cdots \circ f_{\sigma_k}^{\omega(\sigma_1\ldots\sigma_{k-1})}(\mu_0),$$

$$K^\omega = \lim_{k\to\infty} K_k^\omega, \qquad \mu^\omega = \lim_{k\to\infty} \mu_k^\omega. \tag{30}$$

It follows from uniform contractivity that for *all* ω one has convergence in the Hausdorff and strong Prokhorov metrics respectively, and that K^ω and μ^ω are independent of K_0 and μ_0.

The collections of all such fractals sets and measures for fixed $\{F^\lambda\}_{\lambda \in \Lambda}$ are denoted by

$$\mathcal{K}_\infty = \{K^\omega : \omega \in \Omega\}, \qquad \mathcal{M}_\infty = \{\mu^\omega : \omega \in \Omega\}. \tag{31}$$

For each k one has

$$K^\omega = \bigcup_{|\sigma|=k} K_\sigma^\omega, \quad K_\sigma^\omega := f_{\sigma_1}^{\omega(\emptyset)} \circ f_{\sigma_2}^{\omega(\sigma_1)} \circ f_{\sigma_3}^{\omega(\sigma_1\sigma_2)} \circ \cdots \circ f_{\sigma_k}^{\omega(\sigma_1\ldots\sigma_{k-1})}(K^{\omega \rfloor \sigma}). \tag{32}$$

The M^k sets K_σ^ω are called the *subfractals* of K^ω at level k.

The maps $\omega \mapsto K^\omega$ and $\omega \mapsto \mu^\omega$ are Hölder continuous. More precisely:

Proposition 15. *With L and r as in* (28)*,*

$$d_\mathcal{H}(K^\omega, K^{\omega'}) \leqslant \frac{2L}{1-r} d^\alpha(\omega, \omega') \quad and \quad d_P(\mu^\omega, \mu^{\omega'}) \leqslant \frac{2L}{1-r} d^\alpha(\omega, \omega'), \tag{33}$$

where $\alpha = \log(1/r)/\log M$.

Proof. Applying (30) with K_0 replaced by $\{a\}$, and using (28) and repeated applications of the triangle inequality, it follows that $d_\mathcal{H}(K^\omega, a) \leqslant (1 + r + r^2 + \cdots)L = L/(1-r)$ and so $d_\mathcal{H}(K^\omega, K^{\omega'}) \leqslant 2L/(1-r)$ for any ω and ω'. If $d(\omega, \omega') = M^{-k}$ then $\omega(\sigma) = \omega'(\sigma)$ for $|\sigma| < k$, and since $d_\mathcal{H}(K^{\omega \rfloor \sigma}, K^{\omega' \rfloor \sigma}) \leqslant 2L/(1-r)$, it follows from (32) and contractivity that $d_\mathcal{H}(K^\omega, K^{\omega'}) \leqslant \frac{2Lr^k}{1-r}$. Since $r^k = M^{-k\alpha} = d^\alpha(\omega, \omega')$, the result for sets follows.

The proof for measures is essentially identical; one replaces μ_0 by δ_a in (30). $\quad\square$

Example 16 (*Random Sierpinski triangles and tree codes*). The relation between a tree code ω and the corresponding fractal set K^ω can readily be seen in Fig. 2. The IFSs $F = (f_1, f_2, f_3)$ and $G = (g_1, g_2, g_3)$ act on $\mathbb{R}^2$, and the f_m and g_m are similitudes with contraction ratios $1/2$ and $1/3$ respectively and fixed point m. If a node is labelled F, then reading from left to right the three main branches of the subtree associated with that node correspond to f_1, f_2, f_3 respectively. Similar remarks apply if the node is labelled G.

If the three functions in F and the three functions in G each are given weights equal to $1/3$ then the measure μ^ω is distributed over K^ω in such a way that $1/3$ of the mass is in each of the top level triangles, $1/9$ in each of the next level triangles, etc.
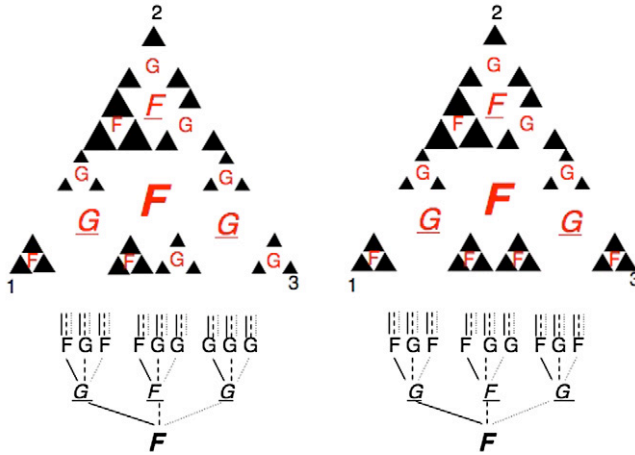
Fig. 2. Random Sierpinski triangles and tree codes.

The sets K^ω and measures μ^ω in Definition 14 are not normally self similar in any natural sense. However, there is an associated notion of statistical self similarity. For this we need the following definition. The reason for the notation ρ_∞ in the definition can be seen from Theorem 49.

Definition 17 *(Standard random fractals).* The *probability distribution* ρ_∞ *on* Ω is defined by choosing $\omega(\sigma) \in \Lambda$ for each $\sigma \in T$ in an iid manner according to P. The *random set* $\mathbb{K} = \omega \mapsto K^\omega$ and the *random measure* $\mathbb{M} = \omega \mapsto \mu^\omega$, each defined by choosing $\omega \in \Omega$ according to ρ_∞, are called *standard random fractals* or *random recursive fractals*. The induced *probability distributions on* $\mathcal{K}_\infty$ *and* $\mathcal{M}_\infty$ respectively are defined by $\mathfrak{K}_\infty = \operatorname{dist} \mathbb{K}$ and $\mathfrak{M}_\infty = \operatorname{dist} \mathbb{M}$.

In the case of random fractal measures one can replace the uniform assumptions in (28) by average or expected type assumptions and the map $\omega \mapsto \mu^\omega$ is then defined a.s., see [21,22]. In Section 6 we do the analogue of this for the case of V-variable random fractal measures.

It follows from the definitions that $\mathbb{K}$ and $\mathbb{M}$ are *statistically self similar* in the sense that

$$\operatorname{dist} \mathbb{K} = \operatorname{dist} \mathbb{F}(\mathbb{K}_1, \ldots, \mathbb{K}_M), \qquad \operatorname{dist} \mathbb{M} = \operatorname{dist} \mathbb{F}(\mathbb{M}_1, \ldots, \mathbb{M}_M), \tag{34}$$

where $\mathbb{F}$ is a random IFS chosen from $(F^\lambda)_{\lambda \in \Lambda}$ according to P, $\mathbb{K}_1, \ldots, \mathbb{K}_M$ are iid copies of $\mathbb{K}$ which are independent of F, and $\mathbb{M}_1, \ldots, \mathbb{M}_M$ are iid copies of $\mathbb{M}$ which are independent of $\mathbb{F}$.

Here, and in the following sections, an IFS F acts on M-tuples of subsets $K_1, \ldots, K_M$ of X and measures $\mu_1, \ldots, \mu_M$ over X by

$$F(K_1, \ldots, K_M) = \bigcup_{m=1}^{M} f_m(K_m), \qquad F(\mu_1, \ldots, \mu_M) = \sum_{m=1}^{M} w_m f_m(\mu_m). \tag{35}$$

This extends in a pointwise manner to random IFSs acting on random sets and random measures as in (34).

We use the terminology "standard" to distinguish the class of random fractals given by Definition 17 and discussed in [15,17,21,22,30] from other classes of random fractals in the literature.

5. V-variable tree codes

5.1. Overview

We continue with the assumptions that $\boldsymbol{F} = \{X, F^\lambda, \lambda \in \Lambda, P\}$ is a family of IFSs as in (1) and (2), and that $\{F^\lambda\}_{\lambda \in \Lambda}$ satisfies the uniform contractive and uniform bounded conditions (28). In Theorem 45 and Example 47 the uniformity conditions are replaced by pointwise average conditions.

In Section 5.2, Definition 18, we define the set $\Omega_V \subset \Omega$ of V-variable tree codes, where Ω is the set of tree codes in Definition 13. Since the $\{F^\lambda\}_{\lambda \in \Lambda}$ are uniformly contractive this leads directly to the class $\mathcal{K}_V$ of V-variable fractal sets and the class $\mathcal{M}_V$ of V-variable fractal measures.

In Section 5.3, Ω_V is alternatively obtained from an IFS $\boldsymbol{\Phi}_V = (\Omega^V, \Phi^a, a \in \mathcal{A}_V)$ acting on Ω^V. More precisely, the attractor $\Omega_V^* \subset \Omega^V$ of $\boldsymbol{\Phi}_V$ projects in any of the V-coordinate directions to Ω_V. However, $\Omega_V^* \neq (\Omega_V)^V$ and in fact there is a high degree of dependence between the coordinates of any $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_V) \in \Omega_V^*$.

If

$$\boldsymbol{\omega} = \lim_{k \to \infty} \Phi^{a_0} \circ \Phi^{a_1} \circ \cdots \circ \Phi^{a_{k-1}} \big(\omega_1^0, \ldots, \omega_V^0\big)$$

we say $\boldsymbol{\omega}$ has address $a_0 a_1 \ldots a_k \ldots$. The limit is independent of $(\omega_1^0, \ldots, \omega_V^0)$.

In Section 5.4 a formalism is developed for finding the V-tuple of tree codes $(\omega_1, \ldots, \omega_V)$ from the address $a_0 a_1 \ldots a_k \ldots$, see Proposition 33 and Example 34. Conversely, given a tree code ω one can find all possible addresses $a_0 a_1 \ldots a_k \ldots$ of V-tuples $(\omega_1, \ldots, \omega_V) \in \Omega_V^*$ for which $\omega_1 = \omega$.

In Section 5.5 the probability distribution ρ_V on the set Ω_V of V-variable tree codes is defined and discussed. The probability P on Λ first leads to a natural probability P_V on the index set $\mathcal{A}_V$ for the IFS $\boldsymbol{\Phi}_V$, see Definition 35. This then turns $\boldsymbol{\Phi}_V$ into an IFS $(\Omega^V, \Phi^a, a \in \mathcal{A}_V, P_V)$ with weights whose measure attractor ρ_V^* is a probability distribution on its set attractor Ω_V^*. The projection of ρ_V^* in any coordinate direction is the same, is supported on Ω_V and is denoted by ρ_V, see Theorem 38.

5.2. V-variability

Definition 18 (*V-variable tree codes and fractals*). A tree code $\omega \in \Omega$ is *V-variable* if for each positive integer k there are at most V distinct tree codes of the form $\omega \rfloor \tau$ with $|\tau| = k$. The set of V-variable tree codes is denoted by Ω_V.

Similarly a finite tree code ω of height p is *V-variable* if for each $k < p$ there are at most V distinct finite subtree codes $\omega \rfloor \tau$ with $|\tau| = k$.

For a uniformly contractive family $\{F^\lambda\}_{\lambda \in \Lambda}$ of IFSs, if ω is V-variable then the fractal set K^ω and fractal measure μ^ω in (30) are said to be *V-variable*. The collections of all V-variable sets and measures corresponding to $\{F^\lambda\}_{\lambda \in \Lambda}$ are denoted by

$$\mathcal{K}_V = \big\{K^\omega \colon \omega \in \Omega_V\big\}, \qquad \mathcal{M}_V = \big\{\mu^\omega \colon \omega \in \Omega_V\big\} \tag{36}$$

respectively, cf. (31).

If $V = 1$ then ω is V-variable if and only if $|\sigma| = |\sigma'|$ implies $\omega(\sigma) = \omega(\sigma')$, i.e. if and only if for each k all values of $\omega(\sigma)$ at level k are equal. In the case $V > 1$, if ω is V-variable then for each k there are at most V distinct values of $\omega(\sigma)$ at level $k = |\sigma|$, but this is not sufficient to imply V-variability.

Remark 19 (*V-variable terminology*). The motivation for the terminology "V-variable fractal" is as follows. Suppose all functions f_m^λ belong to the same group G of transformations. For example, if $X = \mathbb{R}^n$ then G might be the group of invertible similitudes, invertible affine transformations or invertible projective transformations. Two sets A and B are said to be *equivalent modulo G* if $A = g(B)$ for some $g \in G$. If K^ω is V-variable and k is a positive integer, then there are at most V distinct trees of the form $\omega\rfloor\sigma$ such that $|\sigma| = k$. If $|\sigma| = |\sigma'| = k$ and $\omega\rfloor\sigma = \omega\rfloor\sigma'$, then from (32)

$$K_\sigma^\omega = g\left(K_{\sigma'}^\omega\right),$$

$$g = f_{\sigma_1}^{\omega(\emptyset)} \circ \cdots \circ f_{\sigma_k}^{\omega(\sigma_1 \ldots \sigma_{k-1})} \circ \left(f_{\sigma_k'}^{\omega(\sigma_1' \ldots \sigma_{k-1}')}\right)^{-1} \circ \cdots \circ \left(f_{\sigma_1'}^{\omega(\emptyset)}\right)^{-1} K_{\sigma'}^\omega. \tag{37}$$

In particular, K_σ^ω and $K_{\sigma'}^\omega$ are equivalent modulo G.

Thus the subfractals of K^ω at level k form at most V distinct equivalence classes modulo G. However, the actual equivalence classes depend upon the level. Similar remarks apply to V-variable fractal measures.

Proposition 20. *A tree code ω is V-variable iff for every positive integer k the finite tree codes $\omega\lfloor k$ are V-variable.*

Proof. If ω is V-variable the same is true for every finite tree code of the form $\omega\lfloor k$.

If ω is not V-variable then for some k there are at least $V + 1$ distinct subtree codes $\omega\rfloor\tau$ with $|\tau| = k$. But then for some p the $V + 1$ corresponding finite tree codes $(\omega\rfloor\tau)\lfloor p$ must also be distinct. It follows $\omega\lfloor(k + p)$ is not V-variable. $\quad\square$

Example 21 (*V-variable Sierpinski triangles*). The first tree code in Fig. 2 is an initial segment of a 3-variable tree code but not of a 2-variable tree code, while the second tree is an initial segment of a 2-variable tree code but not of a 1-variable tree code. The corresponding Sierpinski type triangles are, to the level of approximation shown, 3-variable and 2-variable respectively.

Theorem 22. *The Ω_V are closed and nowhere dense in Ω, and*

$$\Omega_V \subset \Omega_{V+1}, \qquad d_{\mathcal{H}}(\Omega_V, \Omega) < \frac{1}{V}, \qquad \bigcup_{V \geqslant 1} \Omega_V \subsetneqq \overline{\bigcup_{V \geqslant 1} \Omega_V} = \Omega,$$

where the bar denotes closure in the metric d.

Proof. For the inequality suppose $\omega \in \Omega$ and define k by $M^k \leqslant V < M^{k+1}$. Then if ω' is chosen so $\omega'(\sigma) = \omega(\sigma)$ for $|\sigma| \leqslant k$ and $\omega'(\sigma)$ is constant for $|\sigma| > k$, it follows $\omega' \in \Omega_V$ and $d(\omega', \omega) \leqslant M^{-(k+1)} < V^{-1}$, hence $d_{\mathcal{H}}(\Omega_V, \Omega) < V^{-1}$. The remaining assertions are clear. $\quad\square$

5.3. An IFS acting on V-tuples of tree codes

Definition 23. The metric space (Ω^V, d) is the set of V-tuples from Ω with the metric

$$d\big((\omega_1 \ldots \omega_V), (\omega_1' \ldots \omega_V')\big) = \max_{1 \leqslant v \leqslant V} d\big(\omega_v, \omega_v'\big),$$

where d on the right side is as in Definition 13.

This is a complete bounded metric and is compact if Λ is finite since the same its true for $V = 1$. See Definition 13 and the comment which follows it. The induced Hausdorff metric on $\mathcal{BC}(\Omega^V)$ is complete and bounded, and is compact if Λ is finite. See the comments following Definition 1.

The notion of V-variability extends to V-tuples of tree codes, V-tuples of sets and V-tuples of measures.

Definition 24 (*V-variable V-tuples*). The V-tuple of tree codes of the form $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_V) \in \Omega^V$ is V*-variable* if for each positive integer k there are at most V distinct subtrees of the form $\omega_v \rfloor \sigma$ with $v \in \{1, \ldots, V\}$ and $|\sigma| = k$. The *set of V-variable V-tuples of tree codes* is denoted by Ω_V^*.

Let $\{F^\lambda\}_{\lambda \in \Lambda}$ be a uniformly contractive family of IFSs. The corresponding sets $\mathcal{K}_V^*$ of V-variable V-tuples of fractal sets, and $\mathcal{M}_V^*$ of V-variable V-tuples of fractal measures, are

$$\mathcal{K}_V^* = \big\{\big(K^{\omega_1}, \ldots, K^{\omega_V}\big): (\omega_1, \ldots, \omega_V) \in \Omega_V^*\big\},$$
$$\mathcal{M}_V^* = \big\{\big(\mu^{\omega_1}, \ldots, \mu^{\omega_V}\big): (\omega_1, \ldots, \omega_V) \in \Omega_V^*\big\},$$

where K^{ω_v} and μ^{ω_v} are as in Definition 14.

Proposition 25. *The projection of Ω_V^* in any coordinate direction equals Ω_V, however $\Omega_V^* \subsetneq \ngeq (\Omega_V)^V$.*

Proof. To see the projection map is onto consider $(\omega, \ldots, \omega)$ for $\omega \in \Omega_V$. To see $\Omega_V^* \subsetneq \ngeq (\Omega_V)^V$ note that a V-tuple of V-variable tree codes need not itself be V-variable. $\quad\square$

Notation 26. Given $\lambda \in \Lambda$ and $\omega_1, \ldots, \omega_M \in \Omega$ define $\omega = \lambda * (\omega_1, \ldots, \omega_M) \in \Omega$ by $\omega(\emptyset) = \lambda$ and $\omega(m\sigma) = \omega_m(\sigma)$. Thus $\lambda * (\omega_1, \ldots, \omega_M)$ is the tree code with λ at the base node $\emptyset$ and the tree ω_m attached to the node m for $m = 1, \ldots, M$.

Similar notation applies if the $\omega_1, \ldots, \omega_M$ are finite tree codes all of the same height.

We define maps on V-tuples of tree codes and a corresponding IFS on Ω^V as follows.

Definition 27 (*The IFS acting on the set of V-tuples of tree codes*). Let V be a positive integer. Let $\mathcal{A}_V$ be the set of all pairs of maps $a = (I, J) = (I^a, J^a)$, where

$$I : \{1, \ldots, V\} \to \Lambda, \quad J : \{1, \ldots, V\} \times \{1, \ldots, M\} \to \{1, \ldots, V\}.$$

For $a \in \mathcal{A}_V$ the map $\Phi^a : \Omega^V \to \Omega^V$ is defined for $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_V)$ by

$$\Phi^a(\boldsymbol{\omega}) = \big(\Phi_1^a(\boldsymbol{\omega}), \ldots, \Phi_V^a(\boldsymbol{\omega})\big), \qquad \Phi_v^a(\boldsymbol{\omega}) = I^a(v) * \big(\omega_{J^a(v,1)}, \ldots, \omega_{J^a(v,M)}\big). \quad (38)$$

Thus $\Phi_v^a(\boldsymbol{\omega})$ is the tree code with base node $I^a(v)$, and at the end of each of its M base branches are attached copies of $\omega_{J^a(v,1)}, \ldots, \omega_{J^a(v,M)}$ respectively.

The *IFS $\boldsymbol{\Phi}_V$ acting on V-tuples of tree codes* and without a probability distribution at this stage is defined by

$$\boldsymbol{\Phi}_V := \big(\Omega^V, \Phi^a, a \in \mathcal{A}_V\big). \quad (39)$$

Note that $\Phi^a : \Omega_V^* \to \Omega_V^*$ for each $a \in \mathcal{A}_V$.

Notation 28. It is often convenient to write $a = (I^a, J^a) \in \mathcal{A}_V$ in the form

$$a = \begin{bmatrix} I^a(1) & J^a(1,1) & \ldots & J^a(1,M) \\ \vdots & \vdots & \ddots & \vdots \\ I^a(V) & J^a(V,1) & \ldots & J^a(V,M) \end{bmatrix}. \quad (40)$$

Thus $\mathcal{A}_V$ is then the set of all $V \times (1 + M)$ matrices with entries in the first column belonging to Λ and all other entries belonging to $\{1, \ldots, V\}$.

Theorem 29. *Suppose $\boldsymbol{\Phi}_V = (\Omega^V, \Phi^a, a \in \mathcal{A}_V)$ is an IFS as in Definition 27, with Λ possibly infinite. Then each Φ^a is a contraction map with Lipschitz constant $1/M$. Moreover, with $\boldsymbol{\Phi}_V$ acting on subsets of Ω^V as in (7) and using the notation of Definition 1, we have $\boldsymbol{\Phi}_V : \mathcal{BC}(\Omega^V) \to \mathcal{BC}(\Omega^V)$ and $\boldsymbol{\Phi}_V$ is a contractive map with Lipschitz constant $1/M$. The unique fixed point of $\boldsymbol{\Phi}_V$ is Ω_V^* and in particular its projection in any coordinate direction equals Ω_V.*

Proof. It is readily checked that each Φ^a is a contraction map with Lipschitz constant $1/M$.

We can establish directly that $\boldsymbol{\Phi}_V(E) := \bigcup_{a \in \mathcal{A}_V} \Phi^a(E)$ is closed if E is closed, since any Cauchy sequence from $\boldsymbol{\Phi}_V(E)$ eventually belongs to $\Phi^a(E)$ for some fixed a. It follows that $\boldsymbol{\Phi}_V$ is a contraction map on the complete space $(\mathcal{BC}(\Omega^V), d_{\mathcal{H}})$ with Lipschitz constant $1/M$ and so has a unique bounded closed fixed point (i.e. attractor).

In order to show this attractor is the set Ω_V^* from Definition 24, note that Ω_V^* is bounded and closed in Ω^V. It is closed under Φ^a for any $a \in \mathcal{A}_V$ as noted before. Moreover, each $\boldsymbol{\omega} \in \Omega_V^*$ is of the form $\Phi^a(\boldsymbol{\omega}')$ for some $\boldsymbol{\omega}' \in \Omega_V^*$ and some (in fact many) Φ^a. To see this, consider the VM tree codes of the form $\omega_v \rfloor m$ for $1 \leqslant v \leqslant V$ and $1 \leqslant m \leqslant M$, where each m is the corresponding node of T of height one. There are at most V distinct such tree codes, which we denote by $\omega_1', \ldots, \omega_V'$, possibly with repetitions. Then from (38)

$$(\omega_1, \ldots, \omega_V) = \Phi^a\big(\omega_1', \ldots, \omega_V'\big),$$

provided

$$I^a(v) = \omega_v(\emptyset), \qquad \omega_{J^a(v,m)}' = \omega_v \lfloor m.$$

So Ω_V^* is invariant under $\boldsymbol{\Phi}_V$ and hence is the unique attractor of the IFS $\boldsymbol{\Phi}_V$. $\quad \square$

In the previous theorem, although $\boldsymbol{\Phi}_V$ is an IFS, neither Theorem 8 nor the extensions in Remark 12 apply directly. If Λ is not finite then Ω^V is neither separable nor compact. Moreover, the map $\boldsymbol{\Phi}_V$ acts on sets by taking infinite unions and so we cannot apply Theorem 8(d) to find a set attractor for $\boldsymbol{\Phi}_V$, since in general the union of an infinite number of closed sets need not be closed.

As a consequence of the theorem, approximations to V-variable V-tuples of tree codes, and in particular to individual V-variable tree codes, can be built up from a V-tuple $\boldsymbol{\tau}$ of finite tree codes of height 0 such as $\tau = (\lambda^*, \dots, \lambda^*)$ for some $\lambda^* \in \Lambda$, and a finite sequence $a_0, a_2, \dots, a_k \in \mathcal{A}_V$, by computing the height k finite tree code $\Phi^{a_0} \circ \cdots \circ \Phi^{a_k}(\boldsymbol{\tau})$. Here we use the natural analogue of (38) for finite tree codes. See also the diagrams in [6, Figs. 19, 20].

5.4. Constructing tree codes from addresses and conversely

Definition 30 *(Addresses for V-variable V-tuples of tree codes).* For each sequence $\boldsymbol{a} = a_0 a_1 \dots a_k \dots$ with $a_k \in \mathcal{A}_V$ and Φ^{a_k} as in (38), define the corresponding V-tuple $\boldsymbol{\omega}^{\boldsymbol{a}}$ of tree codes by

$$\boldsymbol{\omega}^{\boldsymbol{a}} = \left(\omega_1^{\boldsymbol{a}}, \dots, \omega_V^{\boldsymbol{a}}\right) := \lim_{k \to \infty} \Phi^{a_0} \circ \Phi^{a_1} \circ \cdots \circ \Phi^{a_k}\left(\omega_1^0, \dots, \omega_V^0\right), \qquad (41)$$

for any initial $(\omega_1^0, \dots, \omega_V^0) \in \Omega^V$. The sequence $\boldsymbol{a}$ is called an *address* for the V-variable V-tuple of tree codes $\boldsymbol{\omega}^{\boldsymbol{a}}$. The set of all such addresses $\boldsymbol{a}$ is denoted by $\mathcal{A}_V^\infty$.

The tree code $\Phi^{a_0} \circ \Phi^{a_1} \circ \cdots \circ \Phi^{a_k}(\omega_1^0, \dots, \omega_V^0)$ is independent of $(\omega_1^0, \dots, \omega_V^0) \in \Omega^V$ up to and including level k, and hence agrees with $\boldsymbol{\omega}^{\boldsymbol{a}}$ for these levels. The sequence in (41) converges exponentially fast since $\mathrm{Lip}\,\Phi^{a_k}$ is $\leqslant 1/M$.

The map $\boldsymbol{a} \mapsto \boldsymbol{\omega}^{\boldsymbol{a}} : \mathcal{A}_V^\infty \to \Omega_V^*$ is many-to-one, since the composition of different $\Phi^{\boldsymbol{a}}$s may give the same map even in simple situations as the following example shows.

Example 31 *(Non-uniqueness of addresses).* The map $\boldsymbol{a} \mapsto \boldsymbol{\omega}^{\boldsymbol{a}} : \mathcal{A}_V^\infty \to \Omega_V^*$ is many-to-one, since the composition of different $\Phi^{\boldsymbol{a}}$s may give the same map even in simple situations. For example, suppose

$$M = 1, \quad V = 2, \quad F \in \Lambda, \quad \Phi^a = \begin{bmatrix} F & 1 \\ F & 1 \end{bmatrix}, \quad \Phi^b = \begin{bmatrix} F & 1 \\ F & 2 \end{bmatrix}.$$

Since $M = 1$ tree codes here are infinite sequences, i.e. 1-branching tree codes. One readily checks from (38) that

$$\Phi^a(\omega, \omega') = (F * \omega, F * \omega), \qquad \Phi^b(\omega, \omega') = (F * \omega, F * \omega'),$$

and so

$$\Phi^a \circ \Phi^b(\omega, \omega') = \Phi^a \circ \Phi^a(\omega, \omega') = \Phi^b \circ \Phi^a(\omega, \omega') = (F * F * \omega, F * F * \omega),$$

$$\Phi^b \circ \Phi^b(\omega, \omega') = (F * F * \omega, F * F * \omega').$$

The following definition is best understood from Example 34.

Definition 32 *(Tree skeletons)*. Given an address $\boldsymbol{a} = a_0 a_1 \ldots a_k \ldots \in \mathcal{A}_V^\infty$ and $v \in \{1, \ldots, V\}$ the corresponding *tree skeleton* $\hat{J}_v^{\boldsymbol{a}} : T \to \{1, \ldots, V\}$ is defined by

$$\hat{J}_v^{\boldsymbol{a}}(\emptyset) = v, \quad \hat{J}_v^{\boldsymbol{a}}(m_1) = J^{a_0}(v, m_1), \quad \hat{J}_v^{\boldsymbol{a}}(m_1 m_2) = J^{a_1}\big(\hat{J}_v^{\boldsymbol{a}}(m_1), m_2\big), \quad \ldots,$$

$$\hat{J}_v^{\boldsymbol{a}}(m_1 \ldots m_k) = J^{a_{k-1}}\big(\hat{J}_v^{\boldsymbol{a}}(m_1 \ldots m_{k-1}), m_k\big), \quad \ldots, \tag{42}$$

where the maps $J^{a_k}(v, m)$ are as in (40).

The tree skeleton depends on the maps $J^{a_k}(w, m)$, but not on the maps $I^{a_k}(w, m)$ and hence not on the set $\{F^\lambda\}_{\lambda \in \Lambda}$ of IFSs and its indexing set Λ.

The V-tuple of tree codes $(\omega_1^{\boldsymbol{a}}, \ldots, \omega_V^{\boldsymbol{a}})$ can be recovered from the address $\boldsymbol{a} = a_0 a_1 \ldots a_k \ldots$ as follows.

Proposition 33 *(Tree codes from addresses)*. If $\boldsymbol{a} = a_0 a_1 \ldots a_k \ldots \in \mathcal{A}_V^\infty$ is an address, I^{a_k} and J^{a_k} are as in (40), and $\hat{J}_v^{\boldsymbol{a}}$ is the tree skeleton corresponding to the J^{a_k}, then for each $\sigma \in T$ and $1 \leqslant v \leqslant V$,

$$\omega_v^{\boldsymbol{a}}(\sigma) = I^{a_k}\big(\hat{J}_v^{\boldsymbol{a}}(\sigma)\big) \quad \text{where } k = |\sigma|. \tag{43}$$

Proof. The proof is implicit in Example 34. A formal proof can be given by induction. $\square$

Example 34 *(The espalier technique)*. [1] We use this to find tree codes from addresses and addresses from tree codes.

We first show how to represent an address $\boldsymbol{a} = a_0 a_1 \ldots a_k \ldots \in \mathcal{A}_V^\infty$ by means of a diagram as in Fig. 3. From this we construct the V-variable V-tuple of tree codes $(\omega_1, \ldots, \omega_V) \in \Omega_V^*$ with address $\boldsymbol{a}$.

Conversely, given a V-variable V-tuple of tree codes $(\omega_1, \ldots, \omega_V) \in \Omega_V^*$ we show how to find the set of all its possible addresses. Moreover, given a single tree code $\omega \in \Omega_V$ we find all possible $(\omega_1, \ldots, \omega_V) \in \Omega_V^*$ with $\omega_1 = \omega$ and all possible addresses in this case.

For the example here let $\Lambda = \{F, G\}$ where F and G are symbols. Suppose $M = 3$ and $V = 5$.

Suppose $\boldsymbol{a} = a_0 a_1 \ldots a_k \ldots \in \mathcal{A}_V^\infty$ is an address of $(\omega_1, \ldots, \omega_V) \in \Omega_V^*$, where

$$a_0 = \begin{bmatrix} F & 3 & 1 & 2 \\ F & 3 & 1 & 3 \\ G & 4 & 3 & 1 \\ F & 4 & 3 & 4 \\ G & 3 & 4 & 4 \end{bmatrix}, \quad a_1 = \begin{bmatrix} F & 1 & 2 & 2 \\ G & 2 & 2 & 5 \\ G & 4 & 5 & 3 \\ F & 5 & 1 & 5 \\ F & 3 & 5 & 3 \end{bmatrix},$$

$$a_2 = \begin{bmatrix} F & 1 & 2 & 3 \\ G & 2 & 5 & 2 \\ F & 3 & 2 & 5 \\ F & 1 & 4 & 4 \\ G & 5 & 3 & 4 \end{bmatrix}, \quad a_3 = \begin{bmatrix} G & * & * & * \\ G & * & * & * \\ F & * & * & * \\ G & * & * & * \\ G & * & * & * \end{bmatrix}. \tag{44}$$

---

[1] Espalier [verb]: to train a fruit tree or ornamental shrub to grow flat against a wall, supported on a lattice.

$I^{a_3}(1) = G$	$I^{a_3}(2) = G$	$I^{a_3}(3) = F$	$I^{a_3}(4) = G$	$I^{a_3}(5) = G$

$J^{a_2}(1,1)=1$

$J^{a_2}(1,2)=2$ $J^{a_2}(2,2)=5$ $J^{a_2}(5,1)=5$

$J^{a_2}(4,3)=4$

$I^{a_2}(1) = F$	$I^{a_2}(2) = G$	$I^{a_2}(3) = F$	$I^{a_2}(4) = F$	$I^{a_2}(5) = G$

$I^{a_1}(1) = F$	$I^{a_1}(2) = G$	$I^{a_1}(3) = G$	$I^{a_1}(4) = F$	$I^{a_1}(5) = F$

$J^{a_0}(1,1)=3$

$J^{a_0}(1,2)=1$

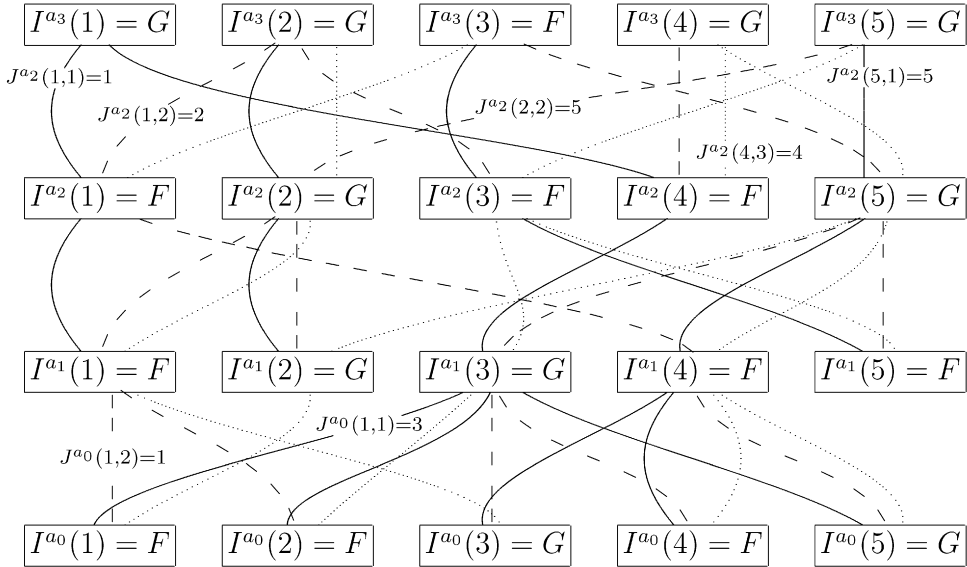| $I^{a_0}(1) = F$	$I^{a_0}(2) = F$	$I^{a_0}(3) = G$	$I^{a_0}(4) = F$	$I^{a_0}(5) = G$

Fig. 3. Constructing tree codes from an address.

We will see that up to level 2 the tree codes ω_1 and ω_2 are those shown in Fig. 2. Although ω_1 and ω_2 are 3-variable and 2-variable respectively up to level 2, it will follow from Fig. 3 that they are 5-variable up to level 3 and are not 4-variable.

The diagram in Fig. 3 is obtained from $\boldsymbol{a} = a_0 a_1 a_2 a_3 \ldots$ by espaliering V copies of the tree T in Definition 13 up through an infinite lattice of V boxes at each level $0, 1, 2, 3, \ldots$. One tree grows out of each box at level 0, and one element from Λ is assigned to each box at each level. When two or more branches pass through the same box from below they inosculate, i.e. their sub branches merge and are indistinguishable from that point upwards. More precisely, $\boldsymbol{a}$ determines the diagram in the following manner. For each level k and starting from each box v at that level, a branch ——— terminates in box number $J^{a_k}(v, 1)$ at level $k + 1$, a branch $---$ terminates in box $J^{a_k}(v, 2)$ at level $k + 1$ and a branch $\cdots\cdots$ terminates in box $J^{a_k}(v, 3)$ at level $k + 1$. The element $I^{a_k}(v) \in \Lambda$ is assigned to box v at level k.

Conversely, any such diagram determines a unique address $\boldsymbol{a} = a_0 a_1 a_2 a_3 \ldots$. More precisely, consider an infinite lattice of V boxes at each level $0, 1, 2, 3, \ldots$. Suppose at each level k there is either F or G in each of the V boxes, and from each box there are 3 branches ———, $---$ and $\cdots\cdots$, each branch terminating in a box at level $k + 1$. From this information one can read off $I^{a_k}(v)$ and $J^{a_k}(v, m)$ for each $k \geqslant 0$, $1 \leqslant v \leqslant V$ and $1 \leqslant m \leqslant M$, and hence determine $\boldsymbol{a}$.

The diagram, and hence the address $\boldsymbol{a}$, determines the tree skeleton $\hat{J}^{\boldsymbol{a}}_v$ by assigning to each node of the copy of T growing out of box v at level 0, the number of the particular box in which that node sits. If $\boldsymbol{\omega}^{\boldsymbol{a}} = (\omega_1, \ldots, \omega_V)$ is the V-tuple of tree codes with address $\boldsymbol{a}$ then the tree code ω_v is obtained by assigning to each node of the copy of T growing out of box v at level 0 the element (fruit?) from Λ in the particular box in which that node sits.

Conversely, suppose $\omega \in \Omega_V$ is a single V-variable tree code. Then the set of all possible $\boldsymbol{\omega} \in \Omega_V^* \subset \Omega^V$ of the form $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_V)$ with $\omega_1 = \omega$, and the set of all diagrams and corresponding addresses $\boldsymbol{a} \in \mathcal{A}_V^\infty$ for such $\boldsymbol{\omega}$, is found as follows. Espalier a copy of T up through the infinite lattice with V boxes at each level in such a way that if $\sigma_0 \ldots \sigma_k$ and $\sigma'_0 \ldots \sigma'_k$ sit in

the same box at level k then the sub tree codes $\omega\lfloor\sigma_0\ldots\sigma_k$ and $\omega\lfloor\sigma_0'\ldots\sigma_k'$ are equal. Since ω is V-variable, this is always possible. From level k onwards the two sub trees are fused together.

The possible diagrams corresponding to this espaliered T are constructed as follows. For each $\sigma \in T$ the element $\omega(\sigma) \in \Lambda$ is assigned to the box containing σ. By construction, this is the same element for any two σ's in the same box. The three branches of the diagram from this box up to the next level are given by the three sub branches of the espaliered T growing out of that box. If T does not pass through some box, then the F or G in that box, and the three branches of the diagram from that box to the next level up, can be assigned arbitrarily.

In this manner one obtains all possible diagrams for which the tree growing out of box 1 at level 0 is ω. Each diagram gives an address $\boldsymbol{a} \in \mathcal{A}_V^\infty$ as before, and the corresponding $\boldsymbol{\omega}^{\boldsymbol{a}} = (\omega_1, \ldots, \omega_V) \in \Omega_V^*$ with address $\boldsymbol{a}$ satisfies $\omega_1 = \omega$. In a similar manner, the set of possible diagrams and corresponding addresses can be obtained for any $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_V) \in \Omega_V^* \subset \Omega^V$.

5.5. The probability distribution on V-variable tree codes

Corresponding to the probability distribution P on Λ in (2) there are natural probability distributions ρ_V on Ω_V, $\mathcal{K}_V$ on $\mathcal{K}_V$ and $\mathfrak{M}_V$ on $\mathcal{M}_V$. See Definition 18 for notation and for the following also note Definition 24.

Definition 35 (*Probability distributions on addresses and tree codes*). The probability distribution P_V on $\mathcal{A}_V$, with notation as in (40), is defined by selecting $a = (I, J) \in \mathcal{A}_V$ so that $I(1), \ldots, I(V) \in \Lambda$ are iid with distribution P, so that $J(1, 1), \ldots, J(V, M) \in \{1, \ldots, V\}$ are iid with the uniform distribution $\{V^{-1}, \ldots, V^{-1}\}$, and so the $I(v)$ and $J(w, m)$ are independent of one another.

The probability distribution P_V^∞ on $\mathcal{A}_V^\infty$, the set of addresses $\boldsymbol{a} = a_0 a_1 \ldots$, is defined by choosing the a_k to be iid with distribution P_V. The probability distribution ρ_V^* on Ω_V^* is the image of P_V^∞ under the map $\boldsymbol{a} \mapsto \boldsymbol{\omega}^{\boldsymbol{a}}$ in (41). The probability distribution ρ_V on Ω_V is the projection of ρ_V^* in any of the V coordinate directions. (By symmetry of the construction this is independent of choice of direction.)

One obtains natural probability distributions on fractals sets and measures, and on V-tuples of fractal sets and measures as follows.

Definition 36 (*Probability distributions on V-variable fractals*). Let $(F^\lambda)_{\lambda \in \Lambda}$ be a uniformly contractive family of IFSs. The probability distributions $\mathcal{K}_V^*$ and $\mathcal{K}_V$ on $\mathcal{K}_V^*$ and $\mathcal{K}_V$ respectively are those induced from ρ_V^* and ρ_V by the maps $(\omega_1, \ldots, \omega_V) \mapsto (K^{\omega_1}, \ldots, K^{\omega_V})$ and $\omega \mapsto K^\omega$ in Definitions 24 and 18.

Similarly, the probability distributions $\mathfrak{M}_V^*$ and $\mathfrak{M}_V$ on $\mathcal{M}_V^*$ and $\mathcal{M}_V$ respectively are those induced from ρ_V^* and ρ_V by the maps $(\omega_1, \ldots, \omega_V) \mapsto (\mu^{\omega_1}, \ldots, \mu^{\omega_V})$ and $\omega \mapsto \mu^\omega$.

That is, $\mathcal{K}_V^*$, $\mathcal{K}_V$, $\mathfrak{M}_V^*$ and $\mathfrak{M}_V$ are the probability distributions of the random objects $(K^{\omega_1}, \ldots, K^{\omega_V})$, K^ω, $(\mu^{\omega_1}, \ldots, \mu^{\omega_V})$ and μ^ω respectively, under the probability distributions ρ_V^* and ρ_V on $(\omega_1, \ldots, \omega_V)$ and ω. Since the projection of ρ_V^* in each coordinate direction is ρ_V it follows that the projection of $\mathcal{K}_V^*$ in each coordinate direction is $\mathcal{K}_V$ and the projection of $\mathfrak{M}_V^*$ in each coordinate direction is $\mathfrak{M}_V$. However, there is a high degree of dependence between the components and in general $\rho_V^* \neq \rho_V{}^V$, $\mathcal{K}_V^* \neq \mathcal{K}_V{}^V$ and $\mathfrak{M}_V^* \neq \mathfrak{M}_V{}^V$.

Definition 37 *(The IFS acting on the set of V-tuples of tree codes).* The IFS $\boldsymbol{\Phi}_V$ in (39) is extended to an IFS with probabilities by

$$\boldsymbol{\Phi}_V := \left(\Omega^V, \Phi^a, a \in \mathcal{A}_V, P_V \right). \tag{45}$$

Theorem 38. *A unique measure attractor exists for $\boldsymbol{\Phi}_V$ and equals ρ_V^*. In particular, the projection of ρ_V^* in any coordinate direction is ρ_V.*

Proof. For $a \in \mathcal{A}_V$ let $R^a : \mathcal{A}_V^\infty \to \mathcal{A}_V^\infty$ denote the operator $\boldsymbol{a} \mapsto a * \boldsymbol{a}$. Then $(\mathcal{A}_V^\infty, R^a,$ $a \in \mathcal{A}_V, P_V)$ is an IFS and the R^a are contractive with Lipschitz constant $1/2$ under the metric $d(\boldsymbol{a}, \boldsymbol{b}) = 2^{-k}$, where k is the least integer such that $a_k \neq b_k$. Thus this IFS has a unique attractor which from Definition 35 is P_V^∞.

Since each $\Phi^a : \Omega^V \to \Omega^V$ has Lipschitz constant $1/M$ from Theorem 29, it follows that Φ^a has Lipschitz constant $1/M$ in the strong Prokhorov (and Monge–Kantorovich) metric as a map on measures. It also follows that $\boldsymbol{\Phi}_V$ has a unique attractor from Theorem 29 and Remark 12.

Finally, if Π is the projection $\boldsymbol{a} \mapsto \omega^a : \mathcal{A}_V^\infty \to \Omega^V$ in (41), it is immediate that $\Pi \circ R^a = \Phi^a \circ \Pi$ and hence the attractor of $\boldsymbol{\Phi}_V$ is $\Pi(P_V^\infty) = \rho_V^*$ by Definition 35.

The projection of ρ_V^* in any coordinate direction is ρ_V from Definition 35. $\quad\square$

Remark 39 *(Connection with other types of random fractals).* The probability distribution ρ_V on Ω_V is obtained by projection from the probability distribution P_V^∞ on $\mathcal{A}_V^\infty$, which is constructed in a simple iid manner. However, because of the combinatorial nature of the many-to-one map

$$\boldsymbol{a} \mapsto \boldsymbol{\omega^a} \mapsto \omega_1^{\boldsymbol{a}} : \mathcal{A}_V^\infty \to \Omega_V^* \to \Omega_V \quad \text{in (41), inducing } P_V^\infty \to \rho_V^* \to \rho_V,$$

the distribution ρ_V is very difficult to analyse in terms of tree codes. In particular, under the distribution ρ_V on Ω_V and hence on Ω, the set of random IFSs $\omega \mapsto F^{\omega(\sigma)}$ for $\sigma \in T$ has a complicated long range dependence structure. (See the comments following Definition 13 for the notation $F^{\omega(\sigma)}$.)

For each V-variable tree code there are at most V isomorphism classes of subtree codes at each level, but the isomorphism classes are level dependent. Moreover, each set of realisations of a V-variable random fractal, as well as its associated probability distribution, is the projection of the fractal attractor of a single *deterministic* IFS operating on V-tuples of sets or measures. See Theorems 43 and 45.

For these reasons ρ_V, $\mathfrak{K}_V$ and $\mathfrak{M}_V$ are very different from other notions of a random fractal distribution in the literature. See also Remark 48.

6. Convergence and existence results for superIFSs

We continue with the assumption that $\boldsymbol{F} = \{X, F^\lambda, \lambda \in \Lambda, P\}$ is a family of IFSs as in (1) and (2).

The set $\mathcal{K}_V$ of V-variable fractal sets and the set $\mathcal{M}_V$ of V-variable fractal measures from Definition 18, together with their natural probability distribution $\mathfrak{K}_V$ and $\mathfrak{M}_V$ from Definition 36, are obtained as the attractors of IFSs $\mathfrak{F}_V^\mathcal{C}$, $\mathfrak{F}_V^{\mathcal{M}_c}$ or $\mathfrak{F}_V^{\mathcal{M}_1}$ under suitable conditions, see Theorems 43 and 45. The Markov chains corresponding to these IFSs provide MCMC algorithms,

such as the "chaos game," for generating samples of V-variable fractal sets and V-variable fractal measures whose empirical distributions converge to the stationary distributions $\mathfrak{K}_V$ and $\mathfrak{M}_V$ respectively.

Definition 40. The metrics $d_{\mathcal{H}}$, d_P and d_{MK} are defined on $\mathcal{C}(X)^V$, $\mathcal{M}_c(X)^V$ and $\mathcal{M}_1(X)^V$ by

$$d_{\mathcal{H}}\big((K_1,\ldots,K_V),(K_1',\ldots,K_V')\big) = \max_v d_{\mathcal{H}}\big(K_v,K_v'\big),$$

$$d_P\big((\mu_1,\ldots,\mu_V),(\mu_1',\ldots,\mu_V')\big) = \max_v d_P\big(\mu_v,\mu_v'\big),$$

$$d_{MK}\big((\mu_1,\ldots,\mu_V),(\mu_1',\ldots,\mu_V')\big) = V^{-1}\sum_v d_{MK}\big(\mu_v,\mu_v'\big), \tag{46}$$

where the metrics on the right are as in Section 2.

The metrics $d_{\mathcal{H}}$ and d_{MK} are complete and separable, while the metric d_P is complete but usually not separable. See Definitions 1, 3 and 4, the comments which follow them, Proposition 5 and Remark 6. The metric d_{MK} is usually not locally compact, see Remark 11.

The following IFSs (48) are analogues of the tree IFS $\boldsymbol{\Phi}_V := (\Omega^V, \Phi^a, a \in \mathcal{A}_V)$ in (39).

Definition 41 *(SuperIFS).* For $a \in \mathcal{A}_V$ as in Definition 27 let

$$\mathcal{F}^a : \mathcal{C}(X)^V \to \mathcal{C}(X)^V, \qquad \mathcal{F}^a : \mathcal{M}_c(X)^V \to \mathcal{M}_c(X)^V, \qquad \mathcal{F}^a : \mathcal{M}_1(X)^V \to \mathcal{M}_1(X)^V$$

be given by

$$\mathcal{F}^a(K_1,\ldots,K_V)$$
$$= \big(F^{I^a(1)}(K_{J^a(1,1)},\ldots,K_{J^a(1,M)}),\ldots,F^{I^a(V)}(K_{J^a(V,1)},\ldots,K_{J^a(V,M)})\big),$$
$$\mathcal{F}^a(\mu_1,\ldots,\mu_V)$$
$$= \big(F^{I^a(1)}(\mu_{J^a(1,1)},\ldots,\mu_{J^a(1,M)}),\ldots,F^{I^a(V)}(\mu_{J^a(V,1)},\ldots,\mu_{J^a(V,M)})\big), \tag{47}$$

where the action of $F^{I^a(v)}$ is defined in (35).

Let

$$\mathfrak{F}_V^{\mathcal{C}} = \big(\mathcal{C}(X)^V, \mathcal{F}^a, a \in \mathcal{A}_V, P_V\big),$$

$$\mathfrak{F}_V^{\mathcal{M}_c} = \big(\mathcal{M}_c(X)^V, \mathcal{F}^a, a \in \mathcal{A}_V, P_V\big),$$

$$\mathfrak{F}_V^{\mathcal{M}_1} = \big(\mathcal{M}_1(X)^V, \mathcal{F}^a, a \in \mathcal{A}_V, P_V\big) \tag{48}$$

be the corresponding IFSs, with P_V from Definition 35. These IFSs are called *superIFSs*.

Two types of conditions will be used on families of IFSs. The first was introduced in (28).

Definition 42 (*Contractivity conditions for a family of IFSs*). The family $\boldsymbol{F} = \{X, F^\lambda, \lambda \in \Lambda, P\}$ of IFSs is *uniformly contractive* and *uniformly bounded* if for some $0 \leqslant r < 1$,

$$\sup_\lambda \max_m d\big(f_m^\lambda(x), f_m^\lambda(y)\big) \leqslant r\, d(x, y) \quad \text{and} \quad \sup_\lambda \max_m d\big(f_m^\lambda(a), a\big) < \infty \qquad (49)$$

for all $x, y \in X$ and some $a \in X$. (The probability distribution P is not used in (49). The second condition is immediate if Λ is finite.)

The family $\boldsymbol{F}$ is *pointwise average contractive* and *average bounded* if for some $0 \leqslant r < 1$

$$\mathbb{E}_\lambda \mathbb{E}_m d\big(f_m^\lambda(x), f_m^\lambda(y)\big) \leqslant r d(x, y) \quad \text{and} \quad \mathbb{E}_\lambda \mathbb{E}_m d\big(f_m^\lambda(a), a\big) < \infty \qquad (50)$$

for all $x, y \in X$ and some $a \in X$.

The following theorem includes and strengthens Theorems 15–24 from [6]. The space X may be noncompact and the strong Prokhorov metric d_P is used rather than the Monge–Kantorovitch metric d_{MK}.

Theorem 43 (*Uniformly contractive conditions*). *Let $\boldsymbol{F} = \{X, F^\lambda, \lambda \in \Lambda, P\}$ be a* finite *family of IFSs on a complete separable metric space (X, d) satisfying (49).*

Then the superIFSs $\mathfrak{F}_V^{\mathcal{C}}$ and $\mathfrak{F}_V^{\mathcal{M}_c}$ satisfy the uniform contractive condition $\operatorname{Lip} \mathcal{F}^a \leqslant r$. Since $(\mathcal{C}(X)^V, d_\mathcal{H})$ is complete and separable, and $(\mathcal{M}_c(X)^V, d_P)$ is complete but not necessarily separable, the corresponding conclusions of Theorem 8 and Remark 12 are valid.

In particular

(1) *$\mathfrak{F}_V^{\mathcal{C}}$ and $\mathfrak{F}_V^{\mathcal{M}_c}$ each have unique compact set attractors and compactly supported separable measure attractors. The attractors are $\mathcal{K}_V^*$ and $\mathfrak{K}_V^*$, and $\mathcal{M}_V^*$ and $\mathfrak{M}_V^*$, respectively. Their projections in any coordinate direction are $\mathcal{K}_V$, $\mathfrak{K}_V$, $\mathcal{M}_V$ and $\mathfrak{M}_V$, respectively.*
(2) *The Markov chains generated by the superIFSs converge at an exponential rate.*
(3) *Suppose $(K_1^0, \ldots, K_V^0) \in \mathcal{C}(X)^V$. If $\boldsymbol{a} = a_0 a_1 \cdots \in \mathcal{A}_V^\infty$ then for some $(K_1, \ldots, K_V) \in \mathcal{K}_V^*$ which is independent of $(K_1^0, \ldots, K_V^0)$,*

$$\mathcal{F}^{a_0} \circ \cdots \circ \mathcal{F}^{a_k}\big(K_1^0, \ldots, K_V^0\big) \to (K_1, \ldots, K_V)$$

in $(\mathcal{C}(X)^V, d_\mathcal{H})$ as $k \to \infty$. Moreover

$$\big\{\mathcal{F}^{a_0} \circ \cdots \circ \mathcal{F}^{a_k}\big(K_1^0, \ldots, K_V^0\big): a_0, \ldots, a_k \in \mathcal{A}_V\big\} \to \mathcal{K}_V^*$$

in $(\mathcal{C}(\mathcal{C}(X)^V, d_\mathcal{H}), d_\mathcal{H})$. Convergence is exponential in both cases.
Analogous results apply for $\mathcal{M}_V^*$, $\mathfrak{K}_V^*$ and $\mathfrak{M}_V^*$, with d_P or $d_\mathcal{H}$ as appropriate.
(4) *Suppose $\boldsymbol{B}^0 = (B_1^0, \ldots, B_V^0) \in \mathcal{C}(X)^V$ and $\boldsymbol{a} = a_0 a_1 \cdots \in \mathcal{A}_V^\infty$ and let $\boldsymbol{B}^k(\boldsymbol{a}) = \boldsymbol{B}^k = \mathcal{F}^{a_k}(\boldsymbol{B}^{k-1})$ if $k \geqslant 1$. Let $B^k(\boldsymbol{a})$ be the first component of $\boldsymbol{B}^k(\boldsymbol{a})$. Then for a.e. $\boldsymbol{a}$ and every $\boldsymbol{B}^0$,*

$$\frac{1}{k} \sum_{n=0}^{k-1} \delta_{B^n(\boldsymbol{a})} \to \mathfrak{K}_V \qquad (51)$$

weakly in the space of probability distributions on $(\mathcal{C}(X), d_{\mathcal{H}})$.

For starting measures $(\mu_1^0, \ldots, \mu_V^0) \in \mathcal{M}_c(X)^V$, there are analogous results modified as in Remark 12 to account for the fact that $(\mathcal{M}_c(X), d_P)$ is not separable.

There are similar results for V-tuples of sets or measures.

Proof. The assertion $\mathrm{Lip}\,\mathcal{F}^a \leqslant r$ follows from (46) by a straightforward argument using Definitions 1 and 4, Eq. (6) and the comment following it, and Eqs. (34), (46) and (47). So the analogue of the uniform contractive condition (19) in Theorem 8 is satisfied, while the uniform boundedness condition is immediate since $\boldsymbol{F}$ is finite. From Theorem 8(d) and Remark 12, $\mathfrak{F}_V^{\mathcal{C}}$ and $\mathfrak{F}_V^{\mathcal{M}_c}$ each has a unique set attractor which is a subset of $\mathcal{C}(X)^V$ and $\mathcal{M}_c(X)^V$ respectively, and a measure attractor which is a probability distribution on the respective set attractor.

It remains to identify the attractors with the sets and distributions in Definitions 18, 24, 35 and 36. Let $\widehat{\Pi}$ denote one of the maps

$$\omega \mapsto K^\omega, \quad \omega \mapsto \mu^\omega,$$
$$(\omega_1, \ldots, \omega_V) \mapsto \left(K^{\omega_1}, \ldots, K^{\omega_V}\right), \quad (\omega_1, \ldots, \omega_V) \mapsto \left(\mu^{\omega_1}, \ldots, \mu^{\omega_V}\right), \tag{52}$$

depending on the context. In the last two cases it follows from (38) and (47) that $\widehat{\Pi} \circ \Phi^a = \mathcal{F}^a \circ \widehat{\Pi}$. Also denote by $\widehat{\Pi}$ the extension of $\widehat{\Pi}$ to a map on sets, on V-tuples of sets, on measures or on V-tuples of measures, respectively. It follows from Theorem 8(d) together with Definitions 18, 24 and 35 that

$$\mathcal{K}_V^* = \widehat{\Pi}\left(\Omega_V^*\right), \qquad \mathcal{K}_V = \widehat{\Pi}(\Omega_V), \qquad \mathfrak{K}_V^* = \widehat{\Pi}\left(\rho_V^*\right), \qquad \mathfrak{K}_V = \widehat{\Pi}(\rho_V),$$
$$\mathcal{M}_V^* = \widehat{\Pi}\left(\Omega_V^*\right), \qquad \mathcal{M}_V = \widehat{\Pi}(\Omega_V), \qquad \mathfrak{M}_V^* = \widehat{\Pi}\left(\rho_V^*\right), \qquad \mathfrak{M}_V = \widehat{\Pi}(\rho_V). \tag{53}$$

The rest of (i) follows from Theorems 29 and 38.

The remaining parts of the theorem follow from Theorem 8 and Remark 12. $\quad\square$

Remark 44 *(Why use the d_P metric?).* For computing approximations to the set of V-variable fractals and its associated probability distribution which correspond to $\boldsymbol{F}$, the main part of the theorem is (iv) with either sets or measures. The advantage of $(\mathcal{M}_c(X), d_P)$ over $(\mathcal{M}_c(X), d_{MK})$ is that for use in the analogue of (27) the space $\mathcal{BC}(\mathcal{M}_c(X), d_P)$ is much larger than $\mathcal{BC}(\mathcal{M}_c(X), d_{MK})$. For example, if $\phi(\mu) = \psi(d_P(\mu, \mu_1))$, where $\mu_1 \in \mathcal{M}_c(X)$ and ψ is a continuous cut-off approximation to the characteristic function of $[0, \epsilon] \subset \mathbb{R}$, then $\phi \in \mathcal{BC}(\mathcal{M}_c(X), d_P)$ but is not continuous or even Borel over $(\mathcal{M}_c(X), d_{MK})$.

Theorem 45 *(Average contractive conditions).* *Let $\boldsymbol{F} = \{X, F^\lambda, \lambda \in \Lambda, P\}$ be a possibly infinite family of IFSs on a complete separable metric space (X, d) satisfying (50).*

Then the superIFS $\mathfrak{F}_V^{\mathcal{M}_1}$ satisfies the pointwise average contractive and average boundedness conditions

$$\mathbb{E}_a\,d_{MK}\left(\mathcal{F}^a(\boldsymbol{\mu}), \mathcal{F}^a(\boldsymbol{\mu}')\right) \leqslant r d_{MK}(\boldsymbol{\mu}, \boldsymbol{\mu}'), \qquad \mathbb{E}_a\,d_{MK}\left(\mathcal{F}^a\left(\boldsymbol{\mu}^0\right), \boldsymbol{\mu}^0\right) < \infty \tag{54}$$

for all $\boldsymbol{\mu}, \boldsymbol{\mu}' \in \mathcal{M}_1(X)^V$ and some $\boldsymbol{\mu}^0 \in \mathcal{M}_1(X)^V$. Since $(\mathcal{M}_1(X)^V, d_{MK})$ is complete and separable, the corresponding conclusions of Theorem 8 are valid.

In particular

(1) $\mathfrak{F}_V^{\mathcal{M}_1}$ *has a unique measure attractor and its projection in any coordinate direction is the same. The attractor and the projection are denoted by $\mathfrak{M}_V^*$ and $\mathfrak{M}_V$, and extend the corresponding distributions in Theorem 43.*

(2) *For a.e. $\boldsymbol{a} = a_0 a_1 \cdots \in \mathcal{A}_V^\infty$, if $(\mu_1^0, \ldots, \mu_V^0) \in \mathcal{M}_1(X)^V$ then $\mathcal{F}^{a_0} \circ \cdots \circ \mathcal{F}^{a_k}(\mu_1^0, \ldots, \mu_V^0)$ converges at an exponential rate. The limit random V-tuple of measures has probability distribution $\mathfrak{M}_V^*$.*

(3) *If $\boldsymbol{a} = a_0 a_1 \cdots \in \mathcal{A}_V^\infty$ and $\boldsymbol{\mu}^0 = (\mu_1^0, \ldots, \mu_V^0) \in \mathcal{M}_1(X)^V$ let $\boldsymbol{\mu}^k(\boldsymbol{a}) = \mathcal{F}^{a_k}(\boldsymbol{\mu}^{k-1})$ for $k \geqslant 1$. Let $\mu^k(\boldsymbol{a})$ be the first component of $\boldsymbol{\mu}^k(\boldsymbol{a})$. Then for a.e. $\boldsymbol{a}$ and every $\boldsymbol{\mu}^0$,*

$$\frac{1}{k} \sum_{n=0}^{k-1} \delta_{\mu^n(\boldsymbol{a})} \to \mathfrak{M}_V \tag{55}$$

weakly in the space of probability distributions on $(\mathcal{M}_1(X), d_{MK})$.

Proof. To establish average boundedness in (54) let $\boldsymbol{\mu}^0 = (\delta_b, \ldots, \delta_b) \in \mathcal{M}_1(X)^V$ for some $b \in X$. Then

$$\mathbb{E}_a \, d_{MK}\big(\mathcal{F}^a(\delta_b, \ldots, \delta_b), (\delta_b, \ldots, \delta_b)\big)$$

$$= \mathbb{E}_a \frac{1}{V} \sum_v d_{MK}\bigg(\sum_m w_m^{I^a(v)} \delta_{f_m^{I^a(v)}(b)}, \delta_b\bigg) \quad \text{from (46), (47), (1) and (35)}$$

$$\leqslant \mathbb{E}_a \frac{1}{V} \sum_v \sum_m w_m^{I^a(v)} d_{MK}\big(\delta_{f_m^{I^a(v)}(b)}, \delta_b\big) \quad \text{by basic properties of } d_{MK}$$

$$= \mathbb{E}_a \frac{1}{V} \sum_v \sum_m w_m^{I^a(v)} d\big(f_m^{I^a(v)}(b), b\big) \quad \text{basic properties of } d_{MK}$$

$$= \mathbb{E}_\lambda \frac{1}{V} \sum_v \sum_m w_m^\lambda d\big(f_m^\lambda(b), b\big) \text{ since dist } I^a(v) = P = \text{dist } \lambda \text{ by Definition 35}$$

$$= \mathbb{E}_\lambda \mathbb{E}_m \, d\big(f_m^\lambda(b), b\big) < \infty.$$

To establish average contractivity in (54) let $(\mu_1, \ldots, \mu_V), (\mu_1', \ldots, \mu_V') \in \mathcal{M}_1(X)^V$. Then

$$\mathbb{E}_a \, d_{MK}\big(\mathcal{F}^a(\mu_1, \ldots, \mu_V), \mathcal{F}^a(\mu_1', \ldots, \mu_V')\big)$$

$$\leqslant \mathbb{E}_a \frac{1}{V} \sum_v d_{MK}\bigg(\sum_m w_m^{I^a(v)} f_m^{I^a(v)}(\mu_{J^a(v,m)}), \sum_m w_m^{I^a(v)} f_m^{I^a(v)}\big(\mu_{J^a(v,m)}'\big)\bigg)$$

$$\leqslant \mathbb{E}_a \frac{1}{V} \sum_v \sum_m w_m^{I^a(v)} d_{MK}\big(f_m^{I^a(v)}(\mu_{J^a(v,m)}), f_m^{I^a(v)}\big(\mu_{J^a(v,m)}'\big)\big)$$

$$= \mathbb{E}_\lambda \frac{1}{V} \sum_v \sum_m w_m^\lambda \, \mathbb{E}_a \, d_{MK}\big(f_m^\lambda(\mu_{J^a(v,m)}), f_m^\lambda\big(\mu_{J^a(v,m)}'\big)\big)$$

$$= \mathbb{E}_\lambda \frac{1}{V} \sum_v \sum_m w_m^\lambda \, \mathbb{E}_t \, d_{MK}\big(f_m^\lambda(\mu_t), f_m^\lambda\big(\mu_t'\big)\big)$$

$$= \mathbb{E}_t \, \mathbb{E}_\lambda \, \mathbb{E}_m \, d_{MK}\big(f_m^\lambda(\mu_t), f_m^\lambda\big(\mu_t'\big)\big).$$

The first inequality is from (47), (1) and (35), the second by properties of d_{MK}, the first equality by the independence of $I^a(v)$ and $J^a(v, n)$ in Definition 35 and since dist $I^a(v) = P = \text{dist } \lambda$, and the second equality by the uniform distribution of $J^a(v, m)$ for fixed (v, m) where t is distributed uniformly over $\{1, \ldots, V\}$.

Next let W_t, W_t' be random variables on X such that dist $W_t = \mu_t$, dist $W_t' = \mu_t'$ and $\mathbb{E} d(W_t, W_t') = d_{MK}(\mu_t, \mu_t')$, where $\mathbb{E}$ without a subscript here and later refers to expectations from the sample space over which the W_t and W_t' are jointly defined. This is possible by [12, Theorem 11.8.2]. Then

$$\mathbb{E}_t \, \mathbb{E}_\lambda \, \mathbb{E}_m \, d_{MK}\big(f_m^\lambda(\mu_t), f_m^\lambda(\mu_t')\big)$$
$$\leqslant \mathbb{E}_t \, \mathbb{E}_\lambda \, \mathbb{E}_m \, \mathbb{E} \, d\big(f_m^\lambda(W_t), f_m^\lambda(W_t')\big) \quad \text{by the third version of (4)}$$
$$\leqslant r \, \mathbb{E}_t \, \mathbb{E} \, d\big(W_t, W_t'\big) \quad \text{from (50)}$$
$$= r \, \mathbb{E}_t \, d_{MK}\big(\mu_t, \mu_t'\big) \quad \text{by choice of } W_t \text{ and } W_t'$$
$$= r d_{MK}\big((\mu_1, \ldots, \mu_V), (\mu_1', \ldots, \mu_V')\big).$$

This completes the proof of (54). The remaining conclusions now follow by Theorem 8. $\quad\square$

Remark 46 *(Global average contractivity).* One might expect that the global average contractive condition $\mathbb{E}_\lambda \, \mathbb{E}_m \, \text{Lip} \, f_m^\lambda < 1$ on the family $\boldsymbol{F}$ would imply the global average contractive condition $\mathbb{E}_a \, \text{Lip} \, \mathcal{F}^a < 1$, i.e. would imply that the superIFS $\mathfrak{F}_V^{\mathcal{M}_1}$ is global average contractive. However, this is not the case.

For example, let $X = \mathbb{R}$, $V = 2$ and $M = 2$. Let $\boldsymbol{F}$ contain a single IFS $F = (f_1, f_2; 1/2, 1/2)$ where

$$f_1(x) = -\frac{3}{2}x, \qquad f_2(x) = \frac{1+\epsilon}{2} + \frac{1-\epsilon}{2}x.$$

Then $\text{Lip} \, f_1 = 3/2$, $\text{Lip} \, f_2 = (1 - \epsilon)/2$ and $\mathbb{E}_m \, \text{Lip} \, f_m = 1 - \epsilon/4$. So $\boldsymbol{F}$ is global average contractive.

Note $f_1(0) = 0$ and $f_2(1) = 1$. Let $\boldsymbol{\mu} = (\delta_0, \delta_0)$, $\boldsymbol{\mu}' = (\delta_1, \delta_1)$ and note $d_{MK}(\boldsymbol{\mu}, \boldsymbol{\mu}') = 1$. Then for any $a \in \mathcal{A}_V$ as in (40),

$$\mathcal{F}^a(\boldsymbol{\mu}) = \left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\frac{1+\epsilon}{2}}, \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\frac{1+\epsilon}{2}}\right), \qquad \mathcal{F}^a(\boldsymbol{\mu}') = \left(\frac{1}{2}\delta_{-\frac{3}{2}} + \frac{1}{2}\delta_1, \frac{1}{2}\delta_{-\frac{3}{2}} + \frac{1}{2}\delta_1\right).$$

From the first form of (4) with $f(x) = |x|$ and from (46), $d_{MK}(\mathcal{F}^a(\boldsymbol{\mu}), \mathcal{F}^a(\boldsymbol{\mu}')) \geqslant 1 - \epsilon/4$ and so $\text{Lip} \, \mathcal{F}^a \geqslant 1 - \epsilon/4$ for every a.

Next let $a^* = \begin{bmatrix} F & 1 & 2 \\ F & 1 & 2 \end{bmatrix}$ and choose $\boldsymbol{\mu} = (\delta_0, \delta_0)$, $\boldsymbol{\mu}' = (\delta_1, \delta_0)$, so $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ differ in the box on which f_1 always acts and agree in the box on which f_2 always acts. Note $d_{MK}(\boldsymbol{\mu}, \boldsymbol{\mu}') = 1/2$. Then

$$\mathcal{F}^{a^*}(\boldsymbol{\mu}) = \left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\frac{1+\epsilon}{2}}, \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\frac{1+\epsilon}{2}}\right), \qquad \mathcal{F}^{a^*}(\boldsymbol{\mu}') = \left(\frac{1}{2}\delta_{-\frac{3}{2}} + \frac{1}{2}\delta_{\frac{1+\epsilon}{2}}, \frac{1}{2}\delta_{-\frac{3}{2}} + \frac{1}{2}\delta_{\frac{1+\epsilon}{2}}\right).$$

Again using the first form of (4) with $f(x) = |x|$, $d_{MK}(\mathcal{F}^{a^*}(\boldsymbol{\mu}), \mathcal{F}^{a^*}(\boldsymbol{\mu}')) \geqslant 3/4$, so $\text{Lip} \, \mathcal{F}^{a^*} \geqslant 3/2$.

Since there are 16 possible maps $a \in \mathcal{A}_V$, each selected with probability $1/16$, it follows that

$$\mathbb{E}_a \operatorname{Lip} \mathcal{F}^a \geqslant \frac{15}{16}\left(1 - \frac{\epsilon}{4}\right) + \frac{1}{16} \cdot \frac{3}{2} > 1 \quad \text{if } \epsilon < \frac{2}{15}.$$

So for such $0 < \epsilon < 2/15$ the IFS $\mathfrak{F}_V^{\mathcal{M}_1}$ is not global average contractive. But since $\mathbb{E}_m \operatorname{Lip} f_m = 1 - \epsilon/4$ it follows from Theorem 45 that $\mathfrak{F}_V^{\mathcal{M}_1}$ is pointwise average contractive, and so Theorem 8 can be applied.

Example 47 *(Random curves in the plane).* The following shows why it is natural to consider families of IFSs which are both infinite and not uniformly contractive. Such examples can be modified to model Brownian motion and other stochastic processes, see [18, Section 5.2] and [23, pp. 120–122].

Let $\boldsymbol{F} = \{\mathbb{R}^2, F^\lambda, \lambda \in \mathbb{R}^2, N(0, \sigma^2 I)\}$ where $F^\lambda = \{f_1^\lambda, f_2^\lambda; 1/2, 1/2\}$ and $N(0, \sigma^2 I)$ is the symmetric normal distribution in $\mathbb{R}^2$ with variance σ^2. The functions f_1^λ and f_2^λ are uniquely specified by the requirements that they be similitudes with positive determinant and

$$f_1^\lambda(-1, 0) = (-1, 0), \quad f_1^\lambda(1, 0) = \lambda, \quad f_2^\lambda(-1, 0) = \lambda, \quad f_2^\lambda(1, 0) = (1, 0).$$

A calculation shows $\sigma = 1.42$ implies $\mathbb{E}_\lambda \mathbb{E}_m \operatorname{Lip} f_m^\lambda \approx 0.9969$ and so average contractivity holds if $\sigma \leqslant 1.42$. If $|\lambda|$ is sufficiently large then neither f_1^λ nor f_2^λ are contractive.

The IFS F^λ can also be interpreted as a map from the space $\mathcal{C}([0, 1], \mathbb{R}^2)$, of continuous paths from $[0, 1]$ to $\mathbb{R}^2$, into itself as follows:

$$\big(F^\lambda(\phi)\big)(t) = \begin{cases} f_1^\lambda(\phi(2t)), & 0 \leqslant t \leqslant \frac{1}{2}, \\ f_2^\lambda(\phi(2t - 1)), & \frac{1}{2} \leqslant t \leqslant 1. \end{cases}$$

Then one can define a superIFS acting on such functions in a manner analogous to that for the superIFS acting on sets or measures. Under the average contractive condition one obtains L^1 convergence to a class of V-variable fractal paths, and in particular V-variable fractal curves, from $(-1, 0)$ to $(1, 0)$. We omit the details.

Remark 48 *(Graph directed fractals).* Fractals generated by a graph directed system [GDS] or more generally by a graph directed Markov system [GDMS], or by random versions of these, have been considered by many authors. See [29,32] for the definitions and references. We comment here on the connection between these fractals and V-variable fractals.

In particular, for each experimental run of its generation process, a random GDS or GDMS generates a single realisation of the associated random fractal. On the other hand, for each run, a superIFS generates a *family* of realisations whose empirical distributions converge a.s. to the probability distribution given by the associated V-variable random fractal.

To make a more careful comparison, allow the number of functions M in each IFS F^λ as in (1) to depend on λ. A GDS together with its associated contraction maps can be interpreted as a map from V-tuples of sets to V-tuples of sets. The map can be coded up by a matrix a as in (40), where M there is now the number of edges in the GDS. If V is the number of vertices in a GDS, the V-tuple $(K^1, \ldots, K^V)$ of fractal sets generated by the GDS is a very particular V-variable V-tuple. If the address $\boldsymbol{a} = a_0 a_1 \ldots a_k \ldots$ for $(K^1, \ldots, K^V)$ is as in (41), then $a_k = a$ for all k. Unlike the situation for V-variable fractals as discussed in Remark 19, there are at most

V distinct subtrees which can be obtained from the tree codes ω^v for K^v *regardless* of the level of the initial node of the subtree.

More generally, if $(K^1, \ldots, K^V)$ is generated by a GDMS then for each k, a_{k+1} is determined just by a_k and by the incidence matrix for the GDMS. Each subtree $\omega^v \rfloor \sigma$ is completely determined by the value $\omega^v(\sigma) \in \Lambda$ at its base node σ and by the "branch" σ_k in $\sigma = \sigma_1 \ldots \sigma_k$.

Realisations of random fractals generated by a random GDS are almost surely not V-variable, and are more akin to standard random fractals as in Definition 17. One comes closer to V-variable fractal sets by introducing a notion of a homogeneous random GDS fractal set analogous to that of a homogeneous random fractal as in Remark 55. But then one does not obtain a class of V-variable fractals together with its associated probability distribution unless one makes the same definitions as in Section 5. This would be quite unnatural in the setting of GDS fractals, for example it would require one edge from any vertex to any vertex.

7. Approximation results as $V \to \infty$

Theorems 49 and 51 enable one to obtain empirical samples of standard random fractals up to any prescribed degree of approximation by using sufficiently large V in Theorem 43(4). This is useful since even single realisations of random fractals are computationally expensive to generate by standard methods. Note that although the matrices used to compute samples of V-variable fractals are typically of order $V \times V$, they are sparse with bandwidth M.

The next theorem improves the exponent in [6, Theorem 12] and removes the dependence on M. The difference comes from using the third rather than the first version of (4) in the proof.

Theorem 49. *If d_{MK} is the Monge–Kantorovitch metric then $d_{MK}(\rho_V, \rho_\infty) \leqslant 1.4 V^{-1/3}$.*

Proof. We construct random tree codes W_V and W_∞ with $\mathrm{dist}\, W_V = \rho_V$ and $\mathrm{dist}\, W_\infty = \rho_\infty$. In order to apply the last equality in (4) we want the expected distance between W_V and W_∞, determined by their joint distribution, to be as small as possible.

Suppose $A = A_0 A_1 A_2 \ldots$ is a random address with $\mathrm{dist}\, A = P_V^\infty$. Let $W_V = \omega_1^A(\sigma)$ be the corresponding random tree code, using the notation of (43) and (42). It follows from Definition 35 that $\mathrm{dist}\, W_V = \rho_V$.

Let the random integer $K = K(A)$ be the greatest integer such that, for $0 \leqslant j \leqslant K$, if $|\sigma| = |\sigma'| = j$ and $\sigma \neq \sigma'$ then $\hat{J}_1^A(\sigma) \neq \hat{J}_1^A(\sigma')$ in (42). Thus with $v = 1$ as in Example 34 the nodes of T are placed in distinct buffers up to and including level K.

Let W_∞ be any random tree code such that if $|\sigma| \leqslant K$ then $W_\infty(\sigma) = W_V(\sigma)$, while if $|\sigma| > K$ then $\mathrm{dist}\, W_\infty(\sigma) = P$ and $W_\infty(\sigma)$ is independent of $W_\infty(\sigma')$ for all $\sigma' \neq \sigma$. It follows from the definition of K that $W_\infty(\sigma)$ are iid with distribution P *for all* σ and so $\mathrm{dist}\, W_\infty = \rho_\infty$.

For any k and for $V \geqslant M^k$,

$$
\mathbb{E}\, d(W_V, W_\infty)
$$
$$
= \mathbb{E}\big(d(W_V, W_\infty) \mid K \geqslant k\big) \cdot \mathrm{Prob}(K \geqslant k) + \mathbb{E}\big(d(W_V, W_\infty) \mid K < k\big) \cdot \mathrm{Prob}(K < k)
$$
$$
\leqslant \frac{1}{M^{k+1}} + \frac{1}{M}\, \mathrm{Prob}(K < k) \quad \text{by (29) and since } W_V(\emptyset) = W_\infty(\emptyset)
$$
$$
= \frac{1}{M^{k+1}} + \frac{1}{M}\left(1 - \prod_{i=1}^{M-1}\left(1 - \frac{i}{V}\right) \prod_{i=1}^{M^2-1}\left(1 - \frac{i}{V}\right) \cdot \ldots \cdot \prod_{i=1}^{M^k-1}\left(1 - \frac{i}{V}\right)\right)
$$

$$\leqslant \frac{1}{M^{k+1}} + \frac{1}{MV}\left(\sum_{i=1}^{M-1} i + \sum_{i=1}^{M^2-1} i + \cdots + \sum_{i=1}^{M^k-1} i\right)$$

$$\leqslant \frac{1}{M^{k+1}} + \frac{1}{2MV}\left(M^2 + M^4 + \cdots + M^{2k}\right)$$

$$\leqslant \frac{1}{M^{k+1}} + \frac{M^{2(k+1)}}{2MV(M^2-1)} \leqslant \frac{1}{M^{k+1}} + \frac{2M^{2k-1}}{3V},$$

noting $\Pi_{i=1}^n (1 - a_i) \geqslant 1 - \sum_{i=1}^n a_i$ for $a_i \geqslant 0$ to obtain the second inequality and assuming $M \geqslant 2$ for the last inequality. The estimate is trivially true if $M = 1$ or $V < M^k$.

Choose x so $M^x = (\frac{3V}{4})^{1/3}$, this being the value of x which minimises $\frac{1}{M^{x+1}} + \frac{2M^{2x-1}}{3V}$. Choose k so $k \leqslant x < k + 1$. Hence from (4)

$$d_{MK}(\rho_V, \rho_\infty) \leqslant \mathbb{E}\, d(W_V, W_\infty) \leqslant \left(\frac{3V}{4}\right)^{-1/3} + \frac{2}{3MV}\left(\frac{3V}{4}\right)^{2/3}$$

$$\leqslant V^{-1/3}\left(\left(\frac{3}{4}\right)^{-1/3} + \frac{1}{3}\left(\frac{3}{4}\right)^{2/3}\right) \leqslant 1.37 V^{-1/3}. \qquad \Box$$

Remark 50 (*No similar estimate for d_P is possible in Theorem 49*). The support of ρ_V converges to the support of ρ_∞ in the Hausdorff metric by Theorem 22. However, $\rho_V \nrightarrow \rho_\infty$ in the d_P metric as $V \to \infty$.

To see this suppose $M \geqslant 2$, fix $j, k \in \{1, \ldots, M\}$ with $j \neq k$ and let $E = \{\omega \in \Omega : \omega(j) = \omega(k)\}$, where j and k are interpreted as sequences of length one in T. According to the probability distribution ρ_∞, $\omega(j)$ and $\omega(k)$ are independent if $j \neq k$. For the probability distribution ρ_V^* there is a positive probability $1/V$ that $J(1, j) = J(1, k)$, in which case $\omega_1(j) = \omega_1(k)$ must be equal from Proposition 33, while if $J(1, j) \neq J(1, k)$ then $\omega_1(j)$ and $\omega_1(k)$ are independent. Identifying the probability distribution ρ_V on Ω with the projection of ρ_V^* on Ω^V in the first coordinate direction it follows $\rho_\infty(E) < \rho_V(E)$.

However, $d(\omega', E) \geqslant 1/M$ if $\omega' \notin E$ since in this case for $\omega \in E$ either $\omega'(j) \neq \omega(j)$ or $\omega'(k) \neq \omega(k)$. Hence for $\epsilon < 1/M$, $E^\epsilon = E$ and so $\rho_\infty(E^\epsilon) = \rho_\infty(E) < \rho_V(E)$. It follows that $d_P(\rho_V, \rho_\infty) \geqslant 1/M$ for all V if $M \geqslant 2$.

Theorem 51. *Under the assumptions of Theorem* 43,

$$d_{\mathcal{H}}(\mathcal{K}_V, \mathcal{K}_\infty), d_{\mathcal{H}}(\mathcal{M}_V, \mathcal{M}_\infty) < \frac{2L}{1-r}V^{-\alpha},$$

$$d_{MK}(\mathfrak{K}_V, \mathfrak{K}_\infty), d_{MK}(\mathfrak{M}_V, \mathfrak{M}_\infty) < \frac{2.8L}{1-r}V^{-\widehat{\alpha}}, \tag{56}$$

where

$$L = \sup_\lambda \max_m d\big(f_m^\lambda(a), a\big), \quad \alpha = \frac{\log(1/r)}{\log M}, \quad \widehat{\alpha} = \begin{cases} \alpha/3 & \text{if } \alpha \leqslant 1, \\ 1 & \text{if } \alpha \geqslant 1. \end{cases}$$

Proof. The first two estimates follow from Theorem 22 and Proposition 15. For the third, let W_V and W_∞ be the random codes from the proof of Theorem 49. In particular,

$$\rho_V = \text{dist } W_V, \qquad \rho_\infty = \text{dist } W_\infty, \qquad \mathbb{E}\, d(W_V, W_\infty) \leqslant 1.4 V^{-1/3}. \tag{57}$$

Let $\widehat{\Pi}$ be the projection map $\Omega \mapsto K^\omega$ given by (30) and (53). Then $\mathfrak{K}_V = \text{dist } \widehat{\Pi} \circ W_V$, $\mathfrak{K}_\infty = \text{dist } \widehat{\Pi} \circ W_\infty$, and from the last condition in (4)

$$d_{MK}(\mathfrak{K}_V, \mathfrak{K}_\infty) \leqslant \mathbb{E}\, d_{\mathcal{H}}(\widehat{\Pi} \circ W_V, \widehat{\Pi} \circ W_\infty). \tag{58}$$

If $\alpha \leqslant 1$ then from (33) on taking expectations of both sides, using Hölder's inequality and applying (57),

$$\mathbb{E}\, d_{\mathcal{H}}(\widehat{\Pi} \circ W_V, \widehat{\Pi} \circ W_\infty) \leqslant \frac{2L}{1-r} \mathbb{E}\, d^\alpha(W_V, W_\infty) \leqslant \frac{2.8L}{1-r} V^{-\alpha/3}.$$

If $\alpha \geqslant 1$ then from the last two lines in the proof of Proposition (15),

$$d_{\mathcal{H}}\big(K^\omega, K^{\omega'}\big) \leqslant \frac{2L}{1-r}\, d^\alpha(\omega, \omega'),$$

and so, using this and arguing as before,

$$\mathbb{E}\, d_{\mathcal{H}}(\widehat{\Pi} \circ W_V, \widehat{\Pi} \circ W_\infty) \leqslant \frac{2.8L}{1-r} V^{-1/3}.$$

This gives the third estimate. The fourth estimate is proved in an analogous manner. $\quad\square$

Sharper estimates can be obtained arguing directly as in the proof of Theorem 49. In particular, the exponent $\widehat{\alpha}$ can be replaced by $\frac{\log(1/r)}{\log(M^2/r)}$.

8. Example of 2-variable fractals

Consider the family $F = \{\mathbb{R}^2, U, D, \frac{1}{2}, \frac{1}{2}\}$ consisting of two IFSs $U = (f_1, f_2)$ (up with a reflection) and $D = (g_1, g_2)$ (down) acting on $\mathbb{R}^2$, where

$$f_1(x, y) = \left(\frac{x}{2} + \frac{3y}{8} - \frac{1}{16}, \frac{x}{2} - \frac{3y}{8} + \frac{9}{16}\right),$$

$$f_2(x, y) = \left(\frac{x}{2} - \frac{3y}{8} + \frac{9}{16}, -\frac{x}{2} - \frac{3y}{8} + \frac{17}{16}\right),$$

$$g_1(x, y) = \left(\frac{x}{2} + \frac{3y}{8} - \frac{1}{16}, -\frac{x}{2} + \frac{3y}{8} + \frac{7}{16}\right),$$

$$g_2(x, y) = \left(\frac{x}{2} - \frac{3y}{8} + \frac{9}{16}, \frac{x}{2} + \frac{3y}{8} - \frac{1}{16}\right).$$

The corresponding fractal attractors of U and F are shown at the beginning of Fig. 4. The probability of choice of U and D is $\frac{1}{2}$ in each case.

U (above) and D (below) attractors　　　　　　　Initial Sets

1: $D(1,2)$　　$U(2,2)$　　　　　2: $U(2,2)$　　$U(1,1)$

3: $D(1,1)$　　$U(1,2)$　　　　　4: $U(2,1)$　　$U(1,2)$

5: $D(1,2)$　　$U(1,2)$　　　　　6: $D(1,2)$　　$U(2,2)$

23: $D(1,2)$　　$D(2,2)$　　　　　24: $U(2,1)$　　$D(2,1)$
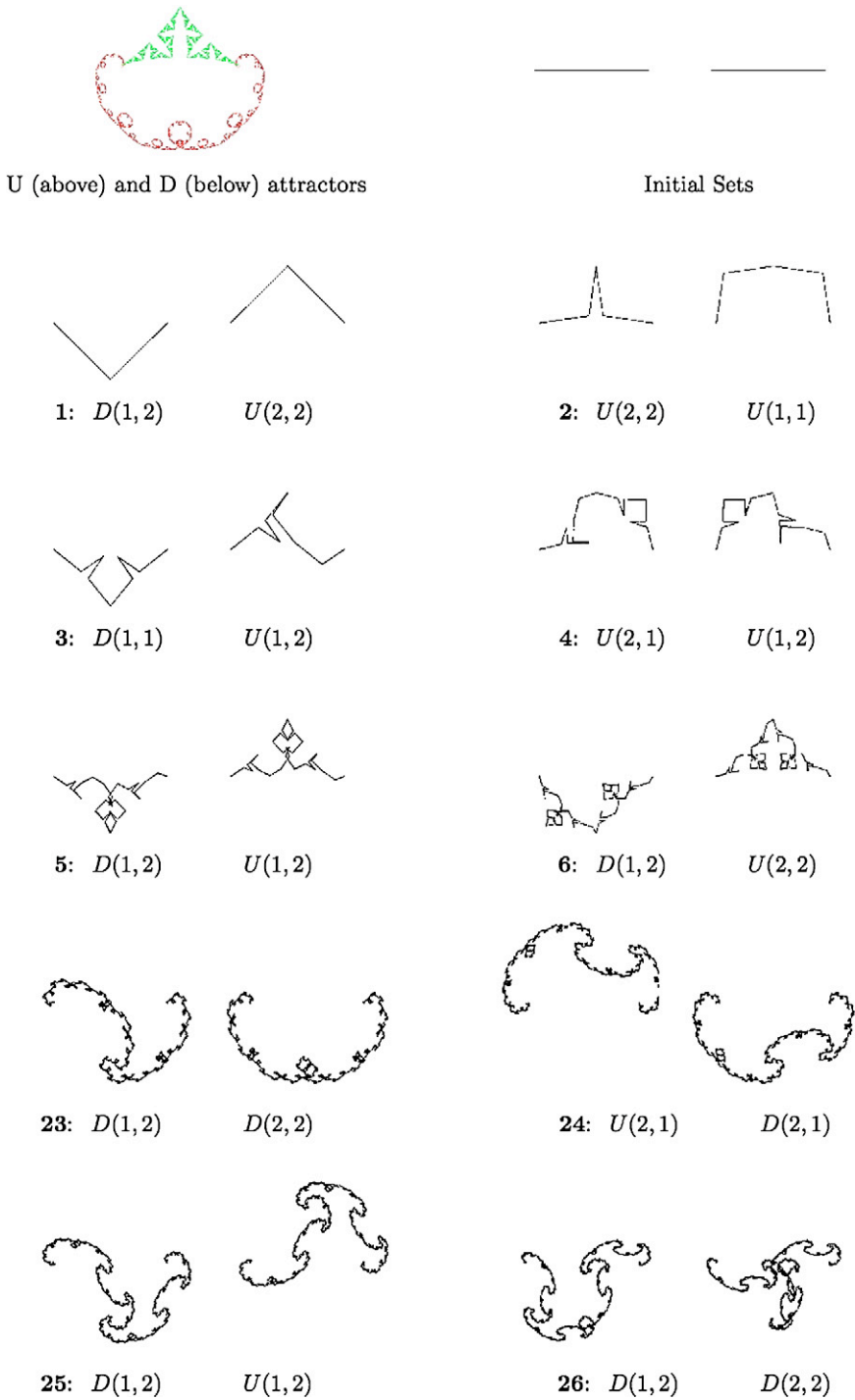

25: $D(1,2)$　　$U(1,2)$　　　　　26: $D(1,2)$　　$D(2,2)$

Fig. 4. Sampling 2-variable fractals.

The 2-variable superIFS acting on pairs of compact sets is $\mathfrak{F}_2^{\mathcal{C}} = (\mathcal{C}(\mathbb{R}^2)^2, \mathcal{F}^a, a \in \mathcal{A}_2, P_2)$. There are 64 maps $a \in \mathcal{A}_2$, each a 2×3 matrix. The probability distribution P_2 assigns probability $\frac{1}{64}$ to each $a \in \mathcal{A}_2$. For example, in iteration 4 the matrix is $a = \begin{bmatrix} U & 2 & 1 \\ U & 1 & 2 \end{bmatrix}$. Applying $\mathcal{F}^a$ to the pair of sets (E_1, E_2) from iteration 3 gives

$$\mathcal{F}^a(E_1, E_2) = \big(U(E_2, E_1), U(E_1, E_2)\big) = \big(f_1(E_2) \cup f_2(E_1), f_1(E_1) \cup f_2(E_2)\big).$$

The process in Fig. 4 begins with a pair of line segments. The first 6 iterations and iterations 23–26 are shown. After about 12 iterations the sets are independent of the initial sets up to screen resolution. After this the pairs of sets can be considered as examples of 2-variable 2-tuples of fractal sets corresponding to F.

The generation process gives an MCMC algorithm or "chaos game" and acts on the infinite state space $(\mathcal{C}(\mathbb{R}^2)^2, d_{\mathcal{H}})$ of pairs of compact sets with the $d_{\mathcal{H}}$ metric. The empirical distribution along a.e. trajectory from any starting pair of sets converges weakly to the 2-variable superfractal probability distribution on 2-variable 2-tuples of fractal sets corresponding to F. The empirical distribution of first (and second) components converges weakly to the corresponding natural probability distribution on 2-variable fractal sets corresponding to F.

9. Concluding comments

Remark 52 *(Extensions)*. The restriction that each IFS F^λ in (1) has the same number M of functions is for notational simplicity only.

In Definition 35 the independence conditions on the I and J may be relaxed. In some modelling situations it would be natural to have a degree of local dependance between $(I(v), J(v))$ and $(I(w), J(w))$ for v "near" w.

The probability distribution ρ_V is in some sense the most natural probability distribution on the set of V-variable code trees since it is inherited from the probability distribution P_V^∞ with the simplest possible probabilistic structure. We may construct more general distributions on the set of V-variable code trees by letting P_V^∞ be non-Bernoulli, e.g. stationary.

Instead of beginning with a family of IFSs in (2) one could begin with a family of graph directed IFSs and obtain in this manner the corresponding class of V-variable graph directed fractals.

Remark 53 *(Dimensions)*. Suppose $F = \{\mathbb{R}^n, F^\lambda; \lambda \in \Lambda, P\}$ is a family of IFSs satisfying the strong uniform open set condition and whose maps are similitudes. In a forthcoming paper we compute the a.s. dimension of the associated family of V-variable random fractals. The idea is to associate to each $a \in \mathcal{A}_V$ a certain $V \times V$ matrix and then use the Furstenberg Kesten theory for products of random matrices to compute a type of pressure function.

Remark 54 *(Motivation for the construction of V-variable fractals)*. The original motivation was to find a chaos game type algorithm for generating collections of fractal sets whose empirical distributions approximated the probability distribution of standard random fractals.

More precisely, suppose $F = \{(X, d), F^\lambda, \lambda \in \Lambda, P\}$ is a family of IFSs as in (2). Let V be a large positive integer and $\mathcal{S}$ be a collection of V compact subsets of X, such that the empirical distribution of $\mathcal{S}$ approximates the distribution $\mathfrak{K}_\infty$ of the standard random fractal associated to F by Definition 17. Suppose $\mathcal{S}^*$ is a second collection of V compact subsets of X obtained from F and $\mathcal{S}$ as follows. For each $v \in \{1, \ldots, V\}$ and independently of other $w \in \{1, \ldots, V\}$,

select $E_1, \ldots, E_M$ from $\mathcal{S}$ according to the uniform distribution independently with replacement, and independently of this select F^λ from $\boldsymbol{F}$ according to P. Let the vth set in $\mathcal{S}^*$ be $F^\lambda(E_1, \ldots, E_M) = \bigcup_{1 \leqslant m \leqslant M} f_m^\lambda(E_m)$. Then one expects the empirical distribution of $\mathcal{S}^*$ to also approximate $\mathfrak{K}_\infty$.

The random operator constructed in this manner for passing from $\mathcal{S}$ to $\mathcal{S}^*$ is essentially the random operator $\mathcal{F}^a$ in Definition 41 with $a \in \mathcal{A}_V$ chosen according to P_V.

Remark 55 (*A hierarchy of fractals*). See Fig. 1.

If $M = 1$ in (1) then each F^λ is a trivial IFS (f^λ) containing just one map, and the family $\boldsymbol{F}$ in (2) can be interpreted as a standard IFS. If moreover $V = 1$ then the corresponding superIFS in Definition 41 can be interpreted as a standard IFS operating on (X, d) with set and measure attractors K and μ, essentially by identifying singleton subsets of X with elements in X. For $M = 1$ and $V > 1$ the superIFS can be identified with an IFS operating on X^V with set and measure attractors K_V^* and μ_V^*. Conversely, any standard IFS can be extended to a superIFS in this manner. The projection of K_V^* in any coordinate direction is K, but $K_V^* \neq K^V$. The attractors K_V^* and μ_V^* are called correlated fractals. The measure μ^* provides information on a certain "correlation" between subsets of K. This provides a new tool for studying the structure of standard IFS fractals as we show in a forthcoming paper.

The case $V = 1$ corresponds to homogeneous random fractals and has been studied in [19,25,36]. The case $V \to \infty$ corresponds to standard random fractals as defined in Definition 17, see also Section 7. See also [1] for some graphical examples.

For a given class $\boldsymbol{F}$ of IFSs and positive integer $V > 1$, one obtains a new class of fractals each with the prescribed degree V of self similarity at every scale. The associated superIFS provides a rapid way of generating a sample from this class of V-variable fractals whose empirical distribution approximates the natural probability distribution on the class.

Large V provides a method for generating a class of correctly distributed approximations to standard random fractals. Small V provides a class of fractals with useful modelling properties.

References

[1] Takahiro Asai, Fractal image generation with iterated function set, Technical Report 24, Ricoh, November 1998.

[2] M.F. Barnsley, Superfractals, Cambridge University Press, Cambridge, 2006.

[3] M.F. Barnsley, S. Demko, Iterated function systems and the global construction of fractals, Proc. Roy. Soc. London Ser. A 399 (1985) 243–275.

[4] Michael F. Barnsley, John H. Elton, A new class of Markov processes for image encoding, Adv. in Appl. Probab. 20 (1988) 14–32.

[5] Michael F. Barnsley, John H. Elton, Douglas P. Hardin, Recurrent iterated function systems, in: Fractal Approximation, Constr. Approx. 5 (1989) 3–31.

[6] Michael Barnsley, John Hutchinson, Örjan Stenflo, A fractal valued random iteration algorithm and fractal hierarchy, Fractals 13 (2005) 111–146.

[7] Patrick Billingsley, Convergence of Probability Measures, Wiley Ser. Probab. Stat., John Wiley & Sons Inc., New York, 1999.

[8] Leo Breiman, The strong law of large numbers for a class of Markov chains, Ann. Math. Statist. 31 (1960) 801–803.

[9] Persi Diaconis, David Freedman, Iterated random functions, SIAM Rev. 41 (1999) 45–76.

[10] Persi Diaconis, Mehrdad Shahshahani, Products of random matrices and computer image generation, in: Random Matrices and Their Applications, Brunswick, Maine, 1984, in: Contemp. Math., vol. 50, Amer. Math. Soc., Providence, RI, 1986, pp. 173–182.

[11] Wolfgang Doeblin, Robert Fortet, Sur des chaînes à liaisons complètes, Bull. Soc. Math. France 65 (1937) 132–148 (in French).

[12] R.M. Dudley, Real Analysis and Probability, Cambridge Stud. Adv. Math., vol. 74, Cambridge University Press, Cambridge, 2002, revised reprint of the 1989 original.

[13] John H. Elton, An ergodic theorem for iterated maps, Ergodic Theory Dynam. Systems 7 (1987) 481–488.
[14] John H. Elton, A multiplicative ergodic theorem for Lipschitz maps, Stochastic Process. Appl. 34 (1990) 39–47.
[15] Kenneth Falconer, Random fractals, Math. Proc. Cambridge Philos. Soc. 100 (1986) 559–582.
[16] Herbert Federer, Geometric Measure Theory, Grundlehren Math. Wiss., Band 153, Springer-Verlag New York Inc., New York, 1969.
[17] Siegfried Graf, Statistically self-similar fractals, Probab. Theory Related Fields 74 (1987) 357–392.
[18] Siegfried Graf, Random fractals, in: School on Measure Theory and Real Analysis, Grado, 1991, Rend. Istit. Mat. Univ. Trieste 23 (1991) 81–144 (1993).
[19] B.M. Hambly, Brownian motion on a homogeneous random fractal, Probab. Theory Related Fields 94 (1992) 1–38.
[20] John E. Hutchinson, Fractals and self-similarity, Indiana Univ. Math. J. 30 (1981) 713–747.
[21] John E. Hutchinson, Ludger Rüschendorf, Random fractal measures via the contraction method, Indiana Univ. Math. J. 47 (1998) 471–487.
[22] John E. Hutchinson, Ludger Rüschendorf, Random fractals and probability metrics, Adv. in Appl. Probab. 32 (2000) 925–947.
[23] John E. Hutchinson, Ludger Rüschendorf, Selfsimilar fractals and selfsimilar random fractals, in: Fractal Geometry and Stochastics, II, Greifswald/Koserow, 1998, in: Progr. Probab., vol. 46, Birkhäuser, Basel, 2000, pp. 109–123.
[24] Richard Isaac, Markov processes and unique stationary probability measures, Pacific J. Math. 12 (1962) 273–286.
[25] Yuri Kifer, Fractals via random iterated function systems and random geometric constructions, in: Fractal Geometry and Stochastics, Finsterbergen, 1994, in: Progr. Probab., vol. 37, Birkhäuser, Basel, 1995, pp. 145–164.
[26] A.S. Kravchenko, Completeness of the space of separable measures in the Kantorovich–Rubinshteĭn metric, Sibirsk. Mat. Zh. 47 (1) (2006) 85–96; English transl.: Siberian Math. J. 47 (1) (2006) 68–76.
[27] Benoit B. Mandelbrot, Fractals: Form, Chance, and Dimension, revised ed., W.H. Freeman and Co., San Francisco, CA, 1977, translated from the French.
[28] Benoit B. Mandelbrot, The Fractal Geometry of Nature, W.H. Freeman and Co., San Francisco, CA, 1982.
[29] R. Daniel Mauldin, Mariusz Urbański, Graph Directed Markov Systems, Cambridge Tracts in Math., vol. 148, Cambridge University Press, Cambridge, 2003.
[30] R. Daniel Mauldin, S.C. Williams, Random recursive constructions: Asymptotic geometric and topological properties, Trans. Amer. Math. Soc. 295 (1986) 325–346.
[31] S.P. Meyn, R.L. Tweedie, Markov Chains and Stochastic Stability, Springer-Verlag London Ltd., London, 1993.
[32] Lars Olsen, Random Geometrically Graph Directed Self-similar Multifractals, Pitman Res. Notes in Math. Ser., vol. 307, Longman Scientific & Technical, Harlow, 1994.
[33] Svetlozar T. Rachev, Probability Metrics and the Stability of Stochastic Models, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Ltd., Chichester, 1991.
[34] Örjan Stenflo, Ergodic theorems for iterated function systems controlled by stochastic sequences, Umeå University, Sweden, 1998, Doctoral thesis, no. 14.
[35] Örjan Stenflo, Ergodic theorems for Markov chains represented by iterated function systems, Bull. Polish Acad. Sci. Math. 49 (2001) 27–43.
[36] Örjan Stenflo, Markov chains in random environments and random iterated function systems, Trans. Amer. Math. Soc. 353 (2001) 3547–3562.
[37] Cédric Villani, Topics in Optimal Transportation, Grad. Stud. in Math., vol. 58, American Mathematical Society, Providence, RI, 2003.
[38] Wei Biao Wu, Michael Woodroofe, A central limit theorem for iterated random functions, J. Appl. Probab. 37 (3) (2000) 748–755.