Fractal Sets as Final Coalgebras Obtained by Completing an Initial Algebra

Prasit Bhattacharya, Lawrence S. Moss*,
Jayampathy Ratnayake, and Robert Rose

Department of Mathematics, Indiana University,
Bloomington, IN 47405 USA

Abstract. This paper is concerned with final coalgebra representations of fractal sets. The background to our work includes Freyd's Theorem: the unit interval is a final coalgebra of a certain endofunctor on the category of bipointed sets. Leinster's far-ranging generalization of Freyd's Theorem is also a central part of the discussion, but we do not directly build on his results. Our contributions are in two different directions. First, we demonstrate the connection of final coalgebras and initial algebras; this is an alternative development to one of his central contributions, working with resolutions. Second, we are interested in the metric space aspects of fractal sets. We work mainly with two examples: the unit interval $[0, 1]$ and the Sierpiński gasket $\mathbb{S}$ as a subset of $\mathbb{R}^2$.

1 Introduction

This paper is a contribution to the presentation of fractal sets in terms of final coalgebras. The first result on this topic was Freyd's Theorem in [6]: the unit interval $[0, 1]$ is the final coalgebra of a functor $X \mapsto X \oplus X$ on the category of bipointed sets. Leinster [10] offers a sweeping generalization of this result. He is able to represent many of what would be intuitively called *self-similar* spaces using (a) bimodules (also called profunctors or distributors), (b) an examination of non-degeneracy conditions on functors of various sorts; (c) a construction of final coalgebras for the types of functors of interest using a notion of resolution. In addition to the characterization of fractal sets as sets, his seminal paper also characterizes them as topological spaces.

Our major contribution is to suggest that in many cases of interest, point (c) above on resolutions is not needed in the construction of final coalgebras. Instead, one may obtain a number of spaces of interest as the Cauchy completion of an initial algebra, and this initial algebra is the set of points in a colimit of an ω-sequence of finite metric spaces. This generalizes Hutchinson's characterization of fractal attractors in [8] as closures of the orbits of the critical points. In addition to simplifying the overall machinery, it also presents a metric space which is "computationally related" to the overall fractal. For example, when

* This work was partially supported by a grant from the Simons Foundation (#245591 to Lawrence Moss).

applied to Freyd's construction, our method yields the metric space of dyadic rational numbers in $[0,1]$. When applied to the Sierpiński gasket $\mathbb{S}$, it yields a countable metric space whose completion is $\mathbb{S}$ and which might be taken as the set of finitely described points in $\mathbb{S}$.

Our second contribution is not completed at this time, but it is a set of results on *metric space* characterizations of final coalgebras. This point was raised as an open issue in Hasuo, Jacobs, and Niqui [7]; this paper was important to us because it emphasized algebras in addition to coalgebras, and because it highlighted endofunctors defined using quotient metrics. Indeed, we use quotient metrics in the main results of this paper. We say that our work is not completed, and by this we mean that what we have is a fully worked out example rather than a general theory. Still, since $[0,1]$ and $\mathbb{S}$ are important in their own right, we feel working them out in detail will inform a more general theory.

Related work We know of a few papers which discuss final coalgebras obtained as Cauchy completions of initial algebras. Perhaps the most relevant is Adámek [2], since it summarizes and extends work that came before it. All of these papers work under assumptions that are not satisfied in our setting. Specifically, [2] has results on locally finitely presentable categories; these are complete, in contrast to the main categories here. To make this point differently, our main categories, Bi (bipointed sets) and Tri (tripointed sets) do not have final objects. Most of the constructions of final coalgebras in the literature *start* from a final object. So for this reason, what we are doing is not a special case of the results in [2].

2 Freyd's Theorem on the Unit Interval

We begin our work by reviewing Peter Freyd's characterization in [6] of the unit interval $[0,1]$ in terms of bipointed sets. Our proof is a little simpler than the original. In our presentation, we include a detour, connecting the final coalgebra characterization of the unit interval to an initial algebra characterization of the dyadic rationals in $[0,1]$.

A *bipointed set* is a triple $(X, \top, \bot)$, where X is an arbitrary set, and $\top$ and $\bot$ belong to X. A bipointed set is *proper* (or *non-degenerate*) if $\bot \neq \top$. We restrict attention to the proper bipointed sets from now on. Let Bi be the category of *proper bipointed sets*; a morphism in Bi is required to preserve the constants $\top$ and $\bot$.

We are mainly interested in a certain endofunctor on bipointed sets which we'll write as $X \mapsto M \otimes X$. (The original notation for $M \otimes X$ was $X \oplus X$, but we have changed this to match Leinster [10].) Here M is a two-element set $\{\ell, r\}$. For a bipointed set X, $M \otimes X$ is the product $M \times X$, modulo the identification $(\ell, \top) = (r, \bot)$. Further, we take $\top_{M\otimes X} = (r, \top)$, and $\bot_{M\otimes X} = (\ell, \bot)$. We complete the definition of $M \otimes X$ as an endofunctor in the following way: if $f : X \to Y$ is a morphism of bipointed sets, then $M \otimes f : M \otimes X \to M \otimes Y$ is $(m, x) \mapsto (m, f(x))$.

Background on initial algebras and final coalgebras. We now recall some general definitions from category theory. Given an endofunctor, that is a functor from some category to itself, say $F : C \to C$, an *algebra* for F is a pair $(A, a : FA \to A)$. Dually, a *coalgebra* for F is a pair $(A, a : A \to FA)$. The difference is the direction of the arrow. In either case, the object A is called the *carrier*, and the morphism a is the *structure*. Often we suppress the structure when referring to algebras or coalgebras. Algebras for F are often called F-algebras, and similarly for coalgebras.

A morphism of algebras $f : A \to B$ is a morphism in the underlying category such that $a \circ f = b \circ Ff$, as on the left below:

$$\begin{array}{ccc} FA & \xrightarrow{a} & A \\ {\scriptstyle Ff}\downarrow & & \downarrow{\scriptstyle f} \\ FB & \xrightarrow[b]{} & B \end{array} \qquad\qquad \begin{array}{ccc} A & \xrightarrow{a} & FA \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle Ff} \\ B & \xrightarrow[b]{} & FB \end{array}$$

A morphism of coalgebras is a morphism $f : A \to B$ such that the diagram on the right commutes. An *initial algebra* is an F-algebra with the property that there is a unique algebra morphism from it to any algebra (of the same functor). A *final coalgebra* is a coalgebra with the property that for every coalgebra, there is a unique coalgebra morphism into it. Initial algebras and final coalgebras occur in many places in mathematics and theoretical computer science. Many examples of these, along with their theory, may be found in the survey volume [3]. We do need two of the central results in the theory.

Lemma 1 (Lambek's Lemma [9]). *The structure map of an initial algebra is a categorical isomorphism. That is, if (A, a) is an initial algebra, then a has an inverse $a^{-1} : FA \to A$ such that $a^{-1} \circ a = \mathsf{id}_{FA}$, and $a \circ a^{-1} = \mathsf{id}_A$. The same hold for final coalgebra structures, mutatis mutandis.*

Theorem 2 (Adámek [1]). *Let C be a category with initial object 0. Let $F : C \to C$ be an endofunctor. Consider the* initial chain

$$0 \xrightarrow{!} F0 \xrightarrow{F!} F^2 0 \xrightarrow{F^2!} \cdots \qquad F^n 0 \xrightarrow{F^n!} F^{n+1}0 \qquad \cdots$$

Suppose that the colimit

$$A = colim_{n<\omega} F^n 0$$

exists, and write $i_n : F^n 0 \to A$ for the cocone morphism. Suppose that F preserves this colimit. Let $a : FA \to A$ be the unique morphism so that $a \circ Fi_n = i_{n+1}$ for all n. Then (A, a) is an initial algebra of F.

We apply this with C being Bi. The category does have an initial algebra, the two-point bipointed set $\{\bot, \top\}$. For the functor F, we take $X \mapsto M \otimes X$. So $F0$ is a three-point bipointed set $\{\bot, \top, \frac{1}{2}\}$, and the map $! : 0 \to F0$ is the inclusion. The $\frac{1}{2}$ here is merely suggestive; we could use any object. But using $\frac{1}{2}$ foreshadows what we shall soon see in the metric setting. Continuing, $F^2 0$ would

be $\{\bot, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \top\}$; again, the map $F! : F0 \to F^2 0$ is the inclusion. This explains how the initial chain is constructed. Its colimit in the category is the set $\mathbb{D}$ of dyadic rationals in $[0,1]$ with $\bot = 0$ and $\top = 1$. (The dyadic rationals are the rationals whose denominator is a power of two.) The structure $d : F\mathbb{D} \to \mathbb{D}$ is

$$\begin{aligned} d(\ell, x) &= x/2 \\ d(r, x) &= (x+1)/2 \end{aligned}$$

It is easy to check that this is well-defined. To apply Theorem 2, we must verify that F preserves the colimit of the initial chain. We omit this detail, and merely state the result which we are after; see [3] for more on this result.

Theorem 3 ([3]). *$(\mathbb{D}, d)$ is an initial algebra of $F : \mathsf{Bi} \to \mathsf{Bi}$.*

We now turn to the final coalgebra of F. There is a dual of Theorem 2, first stated by Barr [5]. However, Barr's result requires that the underlying category C have a final object, and the category Bi of bipointed sets does not have a final object. [To see this, suppose towards a contradiction that A were final. Let B be the three-element Bi-object $\{\top, \bot, *\}$. There are at least two Bi-morphisms $i : B \to A$. One is $\top \mapsto \top$, $\bot \mapsto \bot$, and $* \mapsto \top$. The other is $\top \mapsto \top$, $\bot \mapsto \bot$, and $* \mapsto \bot$. This shows that A is not final.] Instead, we use a different approach. One way is to basically "guess" the final coalgebra and verify the guess. This is what we shall do in Theorem 4 below.

Let I be the unit interval $[0,1]$, with $\bot = 0$ and $\top = 1$. So I is an object in Bi. We have a coalgebra structure $i : I \to FI$ quite closely related to the inverse of d above:

$$i(x) = \begin{cases} (\ell, 2x) & \text{if } x \leq 1/2 \\ (r, 1-2x) & \text{if } x \geq 1/2 \end{cases}$$

(If $x = 1/2$, then both cases agree because $(\ell, 1) = (r, 0)$.) Note that i is a bijection, and indeed an isomorphism in Bi. In particular, i^{-1} exists in Bi.

Theorem 4 (Freyd [6]). *(I, i) is a final coalgebra of $F : \mathsf{Bi} \to \mathsf{Bi}$.*

Proof. Let $(X, e : X \to FX)$ be a coalgebra, where X is any biointed set. The homset $\mathsf{Set}(X, I)$ is a complete metric space with the metric given by the supremum of pointwise distances. And the homset $\mathsf{Bi}(X, I)$ is a closed subset of this space. Thus $\mathsf{Bi}(X, I)$ is a complete metric space. It is also non-empty. Consider the function

$$\phi : \mathsf{Bi}(X, I) \to \mathsf{Bi}(X, I)$$

given by

$$\phi(f) = i^{-1} \circ (Ff) \circ e.$$

In other words, $\phi(f)$ makes the diagram below commute:

$$\begin{array}{ccc} X & \xrightarrow{e} & FX \\ {\scriptstyle \phi(f)}\downarrow & & \downarrow{\scriptstyle Ff} \\ I & \xrightarrow[i]{} & FI \end{array}$$

Then ϕ is a contraction of a non-empty complete metric space. (The contractive property means that in the natural metric on $\mathsf{Bi}(X, I)$,

$$d(\phi(f), \phi(g)) < \tfrac{1}{2} d(f, g).$$

This property is easy to check; it uses the fact that in our definition of FX, we scaled the metric of X by $\frac{1}{2}$.) Thus ϕ has a unique fixed point. A fixed point ϕ^* of ϕ is exactly a coalgebra morphism from X to I, and so we are done.

We mentioned before that this proof of Theorem 4 is simpler than the original argument. Note also that the proof does not really use Theorem 3. So it is not clear that there is a connection between these results at all. The central point in this paper is to draw such a connection, based on the observation that the unit interval is the Cauchy completion of the dyadic rationals in $[0, 1]$. Our way of making the connection is via a metric version of Bi.

MS is the category of metric spaces, with all distance bounded by 1. The morphisms in MS are the *short maps* $f : X \to Y$ of metric spaces: for all $x, y \in X$,

$$d_Y(f(x), f(y)) \leq d_X(x, y)$$

These are also called *non-expanding maps*. (Incidentally, we choose the short morphisms as the morphisms in this category because it is the most standard choice in the literature on coalgebras. However, at the very end of this paper we shall see a reason to perhaps re-consider this choice.)

Let the category BiMS of *bipointed metric spaces* be the category whose objects are bipointed sets which are also metric spaces, and with the property that $d(\bot, \top) = 1$. The morphisms are short maps which preserve $\bot$ and $\top$. The endofunctor $X \to M \otimes X$ works in the metric setting.

Theorem 5. *The initial algebra of $X \to M \otimes X$ on BiMS is the bipointed metric space of dyadic rationals $(\mathbb{D}, d)$. The final coalgebra is the unit interval (I, i).*

The initiality results is similar to what we saw in Theorem 3. The finality result is proved in nearly the same way as Theorem 4, except that we must show one extra point: if X is a bipointed set, then the set of short maps from X to I is non-empty. (We need this because we want to apply the contraction mapping theorem to get a fixed point, but to do this we must know that the space is actually non-empty.) Fortunately, in this case, the result is easy: $x \mapsto d(x, \bot)$ is a BiMS-morphism of any space into $[0, 1]$. The analogous result for the Sierpiński gasket will be much more difficult to obtain.

3 The Sierpiński Gasket

Freyd's Theorem suggests that many interesting topological or metric spaces might have final coalgebra characterizations. This point has been substantiated by Leinster [10], especially on topics related to the topology. We are after something similar for the metric aspects. This paper begins this development by an

extended discussion of the *Sierpiński gasket* (or *Sierpiński triangle*) $\mathbb{S} \subseteq \mathbb{R}^2$. This is sometimes taken to be the prototypical fractal set, and so it is a good place to begin. As we shall see, not all of the results on $\mathbb{S}$ are straightforward generalizations of what we have seen for the unit interval.

We first recall the characterization of $\mathbb{S}$ using *iterated function systems*. Consider the maps $\sigma_a, \sigma_b, \sigma_c : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\begin{aligned} \sigma_a(x,y) &= (x/2, y/2) + (1/4, \sqrt{3}/4) \\ \sigma_b(x,y) &= (x/2, y/2) \\ \sigma_c(x,y) &= (x/2, y/2) + (1/2, 0) \end{aligned} \tag{1}$$

The idea is that σ_a takes a point to the midpoint of the segment determined by it and $(1/2, \sqrt{3}/2)$; σ_b acts similarly using $(0,0)$, and σ_c uses $(1,0)$. These maps σ_i extend to subsets of $\mathbb{R}^2$ by taking images.

Definition 1. *The Sierpiński gasket is the unique non-empty compact subset* $\mathbb{S} \subseteq \mathbb{R}^2$ *such that*

$$\mathbb{S} = \sigma_a(\mathbb{S}) \cup \sigma_b(\mathbb{S}) \cup \sigma_c(\mathbb{S}). \tag{2}$$

$\mathbb{S}$ is shown in Figure 1.

Later on, we shall need some notation for certain subsets of the plane. Let

$$M = \{a, b, c\}$$

We use the letter m as a variable over M in the sequel.

Let M^* be the set of finite words from M. We define triangles tr_w for $w \in M^*$ by recursion:

$$\begin{aligned} tr_\varepsilon &= \text{the triangle with vertices} \\ &\quad \mathsf{top} = (1/2, \sqrt{3}/2),\ \mathsf{left} = (0,0), \text{ and } \mathsf{right} = (1,0) \\ tr_{mw} &= \sigma_m(tr_w) \end{aligned} \tag{3}$$

Let

$$R_n = \bigcup \{tr_w : w \text{ is a word of length } n\}.$$

Also, for an infinite sequence $\alpha = \alpha_0\alpha_1 \cdots \alpha_n \cdots \in M^\omega$, let

$$p_\alpha = \text{ the unique point in } \bigcap_n tr_{\alpha_0\alpha_1\cdots\alpha_n}.$$

That is, $\bigcap_n tr_{\alpha_0\alpha_1\cdots\alpha_n}$ is a nested intersection of a family of non-empty compact sets, and the diameter of the sets tends to 0. By the Cantor intersection theorem, this intersection is a singleton.

The following is a special case of the classical result of Hutchinson [8] on fractal attractors.

Proposition 6. $\mathbb{S}$ *has the following characterizations:*

$$\begin{aligned} \mathbb{S} &= \bigcap_n R_n \\ &= \{p_\alpha : \alpha \in M^\omega\} \\ &= \textit{the closure of } \bigcup_{w \in M^*} \{x : x \textit{ is a vertex of } t_w\}. \end{aligned}$$

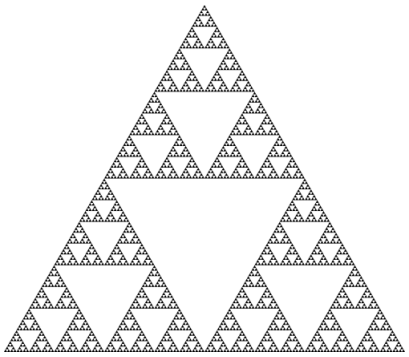


Fig. 1. The Sierpiński gasket $\mathbb{S} \subseteq \mathbb{R}^2$.

Different sequences in M^ω might well be associated with the same point via the operation p. For example, let $\alpha = baccc\cdots$, and let $\beta = bcaaa\cdots$. Then $p_\alpha = p_\beta = (3/8, \sqrt{3}/8)$. However, the sequence α is a *resolution* of p_α. And so what we have is the familiar phenomenon that a given point might have more than one resolution.

In addition, Proposition 6 gives three different ways to think about $\mathbb{S}$. We just saw that writing $\mathbb{S} = \{p_\alpha : \alpha \in M^\omega\}$ connects $\mathbb{S}$ with resolutions. But the third way is closest to what we do in this paper.

4 Tripointed Sets

The main categories involved in our rendering the Sierpiński gasket as a final coalgebra are *tripointed sets* and *tripointed metric spaces*.

Definition 2. *A* tripointed set *is a set X together with distinguished elements* top, left, *and* right. *X is* proper *(or* non-degenerate*) if the distinguished elements are distinct. As with bipointed sets, we are only really interested in proper tripointed sets.*

We let Tri *be the category of proper tripointed sets, taking as morphisms the functions respecting the distinguished points.*

Example 1. The initial object I of Tri is $\{\mathsf{top}, \mathsf{left}, \mathsf{right}\}$. But Tri has no final object. To see this, suppose that X were final. Let Y contain top, left, and right, and also a fourth point $*$. There are at least three short maps from Y to X, since $*$ can be sent to any of the distinguished points.

Example 2. The Sierpiński gasket is an object of Tri. For the structure, we take $\mathsf{top} = (1/2, \sqrt{3}/2)$, $\mathsf{left} = (0, 0)$, and $\mathsf{right} = (1, 0)$.

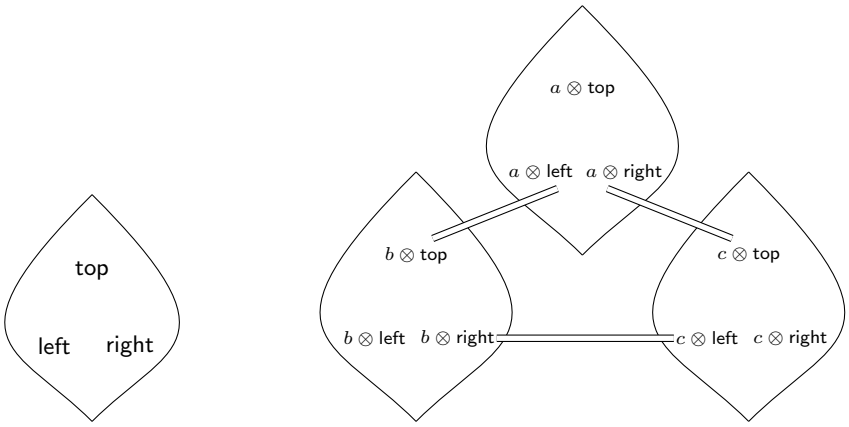


Fig. 2. On the left we have a Tri-object X. On the right we show $M \otimes X$. It is the cartesian product of $M = \{a, b, c\}$ with X, with three pairs of points identified. The identifications are shown with the doubled lines. The top of $M \otimes X$ is $a \otimes \mathsf{top}$, and similarly for left and right. In Section 5, we shall study a version of this functor for TriMS-objects. In that setting the metric multiplies the distances in the copies of M by $\frac{1}{2}$, and then takes the quotient metric associated with the identifications shown above. Pairs of points identified by the doubled lines count as distance 0 apart.

The functor $X \mapsto M \otimes X$ on Tri. Recall the Set endofunctor $X \mapsto M \times X$, where M is $\{a, b, c\}$. For a function $f : X \to Y$, $M \times f : M \times X \to M \times Y$ is $(m, x) \mapsto (m, fx)$. We are are interested in a version of this functor for Tri.

We define $\sim$ on $M \times X$ to be the relation

$$(a, \mathsf{left}) \sim (b, \mathsf{top}) \qquad (a, \mathsf{right}) \sim (c, \mathsf{top}) \qquad (b, \mathsf{right}) \sim (c, \mathsf{left}) \tag{4}$$

Let $\approx$ be the reflexive and symmetric closure of $\sim$. This relation $\approx$ is an equivalence relation. We take $M \otimes X$ to be the quotient of $M \times X$ by the relation $\approx$. See Figure 2 for a picture. We write $m \otimes x$ instead of the equivalence class $[(m, x)]$. So we have equalities

$$a \otimes \mathsf{left} = b \otimes \mathsf{top} \qquad a \otimes \mathsf{right} = c \otimes \mathsf{top} \qquad b \otimes \mathsf{right} = c \otimes \mathsf{left}$$

We call these points the *connection points of $M \otimes X$*.

Returning to our tripointed set $M \otimes X$, we now see that this set is a tripointed space with

$$\begin{aligned} \mathsf{top}_{M\otimes X} &= a \otimes \mathsf{top} \\ \mathsf{left}_{M\otimes X} &= b \otimes \mathsf{left} \\ \mathsf{right}_{M\otimes X} &= c \otimes \mathsf{right} \end{aligned}$$

If X is a proper tripointed set, so is $M \otimes X$. Moreover, the operation $X \mapsto M \otimes X$ extends to a functor $\mathsf{Tri} \to \mathsf{Tri}$.

The Sierpiński gasket $\mathbb{S}$ carries a coalgebra structure σ for this functor. It is easier to define the algebra structure which is the inverse of this coalgebra structure,

$$\sigma^{-1} : M \otimes \mathbb{S} \to \mathbb{S}$$

σ^{-1} is given in terms of the maps σ_a, σ_b, and σ_c in (1):

$$\sigma^{-1}(m \otimes x) = \sigma_m(x). \tag{5}$$

We need to check that this map is well-defined. For example, $a \otimes \mathsf{left} = b \otimes \mathsf{top}$ in $M \otimes \mathbb{S}$. And indeed $\sigma_a(\mathsf{left}) = (\frac{1}{4}, \frac{\sqrt{3}}{2}) = \sigma_b(\mathsf{top})$. We have to check the same thing for the other pairs in (4); the verifications are similar. It is also easy to check that σ^{-1} preserves top, left, and right. Finally, it is one-to-one. All of this goes to show that $\sigma : \mathbb{S} \to M \otimes \mathbb{S}$ is a coalgebra on Tri, and also an isomorphism.

5 Tripointed Metric Spaces

The functor $X \mapsto M \times X$ on MS. The operation $X \mapsto M \times X$ extends to an endofunctor on MS in the following way. We take

$$d((m,x),(n,y)) = \begin{cases} \frac{1}{2}d(x,y) & \text{if } m = n \\ 1 & \text{if } m \neq n \end{cases}$$

If $f : X \to Y$ is a morphism in MS, then so is $M \times f : M \times X \to M \times Y$.

It is easy to show that if X is complete, then $M \times X$ is also complete.

Tripointed metric spaces. An object of TriMS is an object of MS (a metric space with distances bounded by 1) with three distinguished points top, left, and right which are required to be of pairwise distance 1. Morphisms are short maps which preserve the distinguished points.

Example 3. The initial object I of TriMS is $I = \{\mathsf{top}, \mathsf{left}, \mathsf{right}\}$ from Example 1 with pairwise distances 1.

Example 4. Another example is the unit triangle $\mathbb{T}$, where

$$\mathbb{T} = \{(x,y) \in \mathbb{R}^2 : y \geq 0 \text{ and } y \leq x\sqrt{3} \text{ and } y \leq \sqrt{3}(1-x)\}.$$

We take $\mathsf{top} = (\sqrt{3}/2, 1)$, $\mathsf{left} = (0,0)$, and $\mathsf{right} = (1,0)$.

The functor $X \mapsto M \otimes X$ on TriMS. We turn $X \mapsto M \otimes X$ into an endofunctor on TriMS by taking the set $M \otimes X$ and then using the quotient metric. This means that the distance from $m \otimes x$ to $n \otimes y$ is the infimum over all finite paths in $M \times X$ of the *score*, where the score is the sum of the distances (in $M \times X$) along the path, but where we count 0 for pairs in the relation $\sim$.

Example 5. Let X be the initial object of TriMS (see Example 3) together with another point x such that

$$d(\mathsf{top}, x) = 1 = d(\mathsf{left}, x) \qquad d(\mathsf{right}, x) = 1/3.$$

There are essentially two paths from $a \otimes x$ to $b \otimes x$: via the path

$$(a, x) \overset{\frac{1}{2}}{\text{———}} (a, \mathsf{left}) \approx (b, \mathsf{top}) \overset{\frac{1}{2}}{\text{———————}} (b, x) \quad .$$

$$(a, x) \overset{\frac{1}{6}}{\text{———}} (a, \mathsf{right}) \approx (c, \mathsf{top}) \overset{\frac{1}{2}}{\text{———}} (c, \mathsf{left}) \approx (b, \mathsf{right}) \overset{\frac{1}{6}}{\text{———}} (b, x)$$

In terms of the score, the second path is shorter, and its score is $\frac{5}{6}$. Indeed, $d(a \otimes x, b \otimes x) = \frac{5}{6}$.

Generalizing what we saw in Example 5 just above, we have the following result.

Lemma 7. *For all $x, y \in X$, distances in $M \otimes X$ are calculated as follows:*

$$\begin{aligned} d(a \otimes x, a \otimes y) &= \tfrac{1}{2} d(x, y) \\ d(a \otimes x, b \otimes y) &= \min\left(\tfrac{1}{2}(d(x, \mathsf{left}) + d(\mathsf{top}, y)), \tfrac{1}{2}(d(x, \mathsf{right}) + 1 + d(\mathsf{right}, y)) \right) \end{aligned}$$

and similarly for the other distance calculations.

Proof. We check the first assertion, concerning two pairs in the same copy of X, say (a, x) and (a, y).

Consider a finite sequence of pairs in $M \times X$, and the sum of the distances between pairs which are not related by $\sim$. Clearly we have

$$d(a \otimes x, a \otimes y) \leq \tfrac{1}{2} d(x, y)$$

The main work is to show the reverse inequalities by a detailed examination of the quotient metric.

If our sequence consists only of pairs in the a-copy, then by the triangle inequality of X, the score has to be at most half the distance from x to y in the original space X. If the sequence has two points in the b-copy, we may assume that those points are different connection points; if they are the same point, then we could get a smaller score by omitting everything between them. But if the sequence has two connection points, then its score is at least $1/2$. The last case is when the sequence has two connection points of different types other than a, say (c, top) and (b, right). In this case, again we have a score of at least $\frac{1}{2}$.

The second assertion is proved similarly. The point is that no sequence of minimal score may use all three connection points. The shortest path between $a \otimes x$ and $b \otimes y$ could either go through $(a, \mathsf{left}) = (b, \mathsf{top})$; or it could go through $(a, \mathsf{right}) = (c, \mathsf{top})$, and then to $(c, \mathsf{left}) = (b, \mathsf{right})$. (This last option occurs in Example 5.)

Proposition 8. *$X \mapsto M \otimes X$ is an endofunctor on* TriMS.

Proof. Lemma 7 implies that the distances in $M \otimes X$ between distinct corner points is 1. For example, consider top and left.

$$\begin{aligned} d(\mathsf{top}_{M\otimes X}, \mathsf{left}_{M\otimes X}) &= d(a \otimes \mathsf{top}, b \otimes \mathsf{left}) \\ &= \min(\tfrac{1}{2}(1+1) + \tfrac{1}{2}(1+1+1)) \\ &= 1 \end{aligned}$$

This checks that $M \otimes X$ is an object in $M \otimes X$. We next check the parallel assertion for morphisms.

Let $f : X \to Y$ be a short map. We show that $M \otimes f : M \otimes X \to M \otimes Y$. One example case concerns $x_1, x_2 \in X$ and asks us that show that

$$d_{M\otimes Y}(a \otimes f(x_1), b \otimes f(x_2)) \leq d_{M\otimes X}(a \otimes x_1, b \otimes x_2).$$

We sketch the proof. Both of the distances above are minima, and so it is sufficient to show that the following two inequalities hold:

$$\begin{aligned} d_Y(f(x_1), \mathsf{left}) + d_Y(\mathsf{top}, f(x_2)) &\leq d_X(x_1, \mathsf{left}) + d_X(\mathsf{top}, x_2) \\ d_Y(f(x_1), \mathsf{right}) + d_Y(\mathsf{top}, f(x_2)) &\leq d_X(x_1, \mathsf{right}) + d_X(\mathsf{top}, x_2) \end{aligned}$$

Now these inequalities hold because f preserves top, left, and right and is itself a short map. For example,

$$d_Y(f(x_1), \mathsf{left}) = d_Y(f(x_1), f(\mathsf{left})) \leq d_X(x_1, \mathsf{left}).$$

The full details of the proof are based on these observations and similar ones.

Example 6. With $\mathbb{S}$ the Sierpiński gasket in the usual metric, we have an algebra $\sigma^{-1} : M \otimes \mathbb{S} \to \mathbb{S}$ (see (5)). This map σ^{-1} is a short map, as is easily verified. On the other hand, its inverse $\sigma : \mathbb{S} \to M \otimes \mathbb{S}$ is not a short map. For example, $\sigma((\frac{1}{2}, \frac{\sqrt{3}}{2})) = (a, \mathsf{top})$, $\sigma((\frac{1}{2}, 0)) = (b, \mathsf{right})$, and

$$\begin{aligned} d_{\mathbb{S}}((\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}), (\tfrac{1}{2}, 0)) &= \tfrac{\sqrt{3}}{2} \\ d_{M\otimes\mathbb{S}}((a, \mathsf{top}), (b, \mathsf{left})) &= 1 \end{aligned}$$

The theorem below is about tripointed *sets*, but the proof makes a detour into tripointed *metric spaces*.

Theorem 9. *$\mathbb{S} \to M \otimes \mathbb{S}$ is a final coalgebra of $M \otimes X$ on* Tri.

Proof. The proof is almost the same as that of Theorem 4, so we merely sketch this one and point out where it differs from the earlier proof. Let $(X, e : X \to M \otimes X)$ be a coalgebra, where X is any tripointed set. The main point is to consider the homset $\mathsf{Tri}(X, \mathbb{S})$. This is easily seen to be a complete metric space.

It is also non-empty: we can give X the discrete metric, and then take any function preserving top, left, and right. Consider the function

$$\phi : \mathsf{Tri}(X, \mathbb{S}) \to \mathsf{Tri}(X, \mathbb{S})$$

given by

$$\phi(f) = \sigma^{-1} \circ (M \otimes f) \circ e.$$

(Here σ^{-1} is from (5), and we also saw it in Example 6.) In other words, $\phi(f)$ makes the diagram below commute:

$$\begin{array}{ccc} X & \xrightarrow{e} & M \otimes X \\ {\scriptstyle \phi(f)}\downarrow & & \downarrow{\scriptstyle M \otimes f} \\ \mathbb{S} & \xleftarrow[\sigma^{-1}]{} & M \otimes \mathbb{S} \end{array}$$

(We use σ^{-1} rather than σ because σ^{-1} is a short map.) Then ϕ is a contraction of a non-empty complete metric space. Thus ϕ has a unique fixed point ϕ^*, and the fixed points of ϕ are exactly the coalgebra morphisms from X to $\mathbb{S}$. Note also that there is no reason to expect that ϕ^* would be a short map, even if X carried a metric and e were itself short, since our proof starts be throwing away the metric on X.

6 Isometric Embeddings and the Initial Chain

As with any morphism of metric spaces, we say that a TriMS morphism $\eta : X \to Y$ is *an isometric embedding* if it preserves all distances.

Proposition 10. *Concerning isometric embeddings:*

1. *The unique map $\eta : I \to M \otimes I$ is an isometric embedding, where I is the initial object of TriMS from Example 3.*
2. *If $\eta : X \to Y$ is an isometric embedding, so is $M \otimes \eta : M \otimes X \to M \otimes Y$.*

Proof. The first part follows from the fact that $M \otimes I$ is an object in TriMS. The second follows easily from Lemma 7; see the proof of Proposition 8 for a similar argument.

As a result, we have a chain of isometric embeddings

$$I \xrightarrow{\eta} M \otimes I \xrightarrow{M \otimes \eta} M \otimes M \otimes I \xrightarrow{M \otimes M \otimes \eta} \cdots \tag{6}$$

In this situation, the colimit exists and is preserved by the functor $M \otimes -$.

Theorem 11. *The colimit of the chain in (6) exists. Call this space G. G carries the structure of an initial algebra $\eta : M \otimes G \to G$ for $M \otimes X$ on TriMS.*

Proof. We are going to use Adámek's Theorem 2. The category TriMS has an initial object. Given a chain of isometric embeddings as in (6), the colimit exists in the category of metric spaces, and is the set-theoretic union of the spaces. This colimit is an object of Tri, since the initial object is isometrically embedded in it. This TriMS object is G in our theorem. The map $i_n : M^n \otimes I \to G$ is the inclusion. This space G is a fixed point: $M \otimes G \to G$. Moreover, the mediating morphism $g : M \otimes G \to G$ given by the colimit is

$$m \otimes (n_1 \otimes n_2 \otimes \cdots \otimes n_k \otimes d) \quad \mapsto \quad m \otimes n_1 \otimes n_2 \otimes \cdots \otimes n_k \otimes d.$$

By Adámek's Theorem 2, (G, g) is an initial algebra.

6.1 Concrete presentation of G.

It might be useful to see a presentation of G in very concrete terms. We have in mind readers not familiar with Adámek's Theorem, but who might be familiar with the basics of universal algebra or abstract data types.

Consider a signature Σ of terms which takes symbols top, left, and right as constants, and symbols a, b, and c as unary function symbols. Then we would get terms such as $a(b(c(\mathsf{right})))$ and $b(b(\mathsf{left}))$. In order to match our earlier notation, we are going to write these as $a \otimes b \otimes c \otimes \mathsf{right}$ and $b \otimes b \otimes \mathsf{left}$, respectively. Consider the free Σ-algebra modulo the equations E

$$\begin{array}{ll} \mathsf{top} = a \otimes \mathsf{top} & a \otimes \mathsf{left} = b \otimes \mathsf{top} \\ \mathsf{left} = b \otimes \mathsf{left} & a \otimes \mathsf{right} = c \otimes \mathsf{top} \\ \mathsf{right} = c \otimes \mathsf{right} & b \otimes \mathsf{right} = c \otimes \mathsf{left} \end{array} \tag{7}$$

This is the set G of all terms on the signature Σ modulo the smallest equivalence relation containing the pairs in E above and closed under the three function symbols. Technically, the elements of G are not terms but rather equivalence classes of terms. (This is often written as $T_{\Sigma,E}$.) In fact, an $(M \otimes X)$-algebra on Tri is exactly a Σ-algebra satisfying E with the property that top, left, and right are distinct. This correspondence extends to Σ-algebra morphisms and $M \otimes X$-algebra morphisms. Then it is a standard result that $G = T_{\Sigma,E}$ is the initial Σ-algebra satisfying E.

We take G and impose the unique metric such that $d(\mathsf{top}, \mathsf{left}) = d(\mathsf{top}, \mathsf{right}) = d(\mathsf{left}, \mathsf{right}) = 1$, and also such that the equations in Lemma 7 hold. This concludes our digression on the concrete presentation of G as a metric space.

Proposition 12. *As a metric space, G is totally bounded.*

Proof. Let $M^n \otimes I$ be the nth term in the chain in (6), starting with $M^0 \otimes I = I$, and $M^1 \otimes I = M \otimes I$. One checks easily that for all n, each point in $M^{n+1} \otimes I$ is of distance $\leq 2^{-(n+1)}$ from some point in M^n. Then using the triangle inequality, it follows that the distance of each point in $M^{n+k} \otimes I$ to some point in $M^n \otimes I$ is at most $\sum_{i=1}^{k} 2^{-(n+1+i)} < 2^{-n}$. Thus, $M^n \otimes I$ is a finite set of points in G with the property that every point in G is within 2^{-n} of some point in it.

6.2 The final coalgebra of $X \to M \otimes X$

This section constructs the final coalgebra of $X \to M \otimes X$ on TriMS. We need a few preliminaries.

Complete tripointed spaces. We let TriCMS be the tripointed metric spaces which are *complete*: every Cauchy sequence has a limit. Every metric space is isometrically embedded in its Cauchy completion, and so we have a Cauchy completion functor

$$C : \mathsf{TriMS} \to \mathsf{TriCMS}.$$

For a tripointed metric space X, consider

$$M \otimes C(X) \xrightarrow{\beta_X} C(M \otimes X) \xrightarrow{\gamma_X} M \otimes C(X)$$

given by

$$\begin{aligned} \beta_X(m \otimes (x_0, x_1, \ldots)) &= (m \otimes x_0, m \otimes x_1, \ldots) \\ \gamma_X((m_i \otimes x_i)_i) &= m^* \otimes (x_{j_1}, x_{j_2}, \ldots,) \end{aligned}$$

In the definition of γ_X, for each sequence of points in $M \otimes X$, some $m \in M$ must occur infinitely many times as the first coordinate m_i, since M is finite. Let m^* be "first" in some pre-chosen order on M. Let $(x_{j_i})_i$ be the subsequence of $(x_i)_i$ such that the corresponding first coordinates are all m^*.

Lemma 13. *$\beta : C(M \otimes -) \to M \otimes C(-)$ and $\gamma : M \otimes C(-) \to C(M \otimes -)$ are natural isomorphisms.*

Proof. One checks that for all X, β_X and γ_X well-defined, that they are inverses (modulo equivalence of Cauchy sequences), that they are short maps (hence they are isometries), that they preserve the distinguished points, and that they are the components of natural transformations. All of these verifications are elementary.

The final coalgebra of $X \to M \otimes X$. We have seen the initial algebra

$$(G, \eta : M \otimes G \to G)$$

of $X \to M \otimes X$ on TriMS. By Lambek's Lemma 1, η is an isomorphism. Applying the Cauchy completion functor C, we have another algebra

$$(CG, C\eta \circ \beta_G : M \otimes CG \to C(M \otimes G) \to CG).$$

Moreover, $C\eta \circ \beta_G$ is an isomorphism. Let us shorten our notation by writing S for CG and s for $(C\eta \circ \beta_G)^{-1}$. So we have a coalgebra $(S, s : S \to M \otimes S)$ on TriMS. S is the completion of a totally bounded space (Proposition 12) and is therefore compact. Forgetting the metric structure, S a coalgebra for $M \otimes X$ on Tri.

Theorem 14 (Leinster [10]). *$(S, s : S \to M \otimes S)$ is a final coalgebra for $M \otimes X$ on Tri.*

Proof. We have seen the argument twice before, in Theorems 4 and 9. The main points are that S is complete, so therefore $\mathsf{Tri}(X, S)$ is complete; also $\mathsf{Tri}(X, S)$ is non-empty, and i is an isometry.

Theorem 14 is due to Leinster, but as with Freyd's Theorem, our proof is different. We feel that going through the initial algebra G offers an advantage of sorts: G is the countable metric space of approximations to points in S. It is a little like what is called the set of finite addresses of points in $\mathbb{S}$, but it is obtained as the colimit of the initial sequence of an endofunctor on TriMS.

7 Finality of S for $M \otimes X$ on TriMS

Theorem 14 shows that S is the final coalgebra of $M \otimes X$ on Tri. In this section, we present a related result. To begin, recall from Proposition 8 that $M \otimes X$ is an endofunctor on TriMS. As S is an object of TriMS and its structure s is a short map, we thus can ask whether S is a final coalgebra of $M \otimes X$ on TriMS. It is important to see that this is not an immediate consequence of Theorem 14. Given a short map of tripointed metric spaces $e : X \to M \otimes X$, we do have a unique coalgebra morphism $f : X \to S$. We must show that f is a short map.

Notation Let $\boldsymbol{m} = m_1, m_2, \ldots, m_k \in M^k$. For $x \in X$, write $\boldsymbol{m} \otimes x$ for $m_1 \otimes \cdots \otimes m_k \otimes x$. Further, $M^i \otimes X$ denotes the space

$$\overbrace{M \otimes M \otimes \cdots \otimes M}^{i} \otimes X.$$

Finally, we use a notation like $M^j \otimes e$ as a shorthand for the obvious map.

Lemma 15. *Let X be an object in TriMS. For all natural numbers i, all $\boldsymbol{m} \in M^i$ and all $x, y \in X$ the distance in $M^i \otimes X$ between*

$$d(\boldsymbol{m} \otimes x, \boldsymbol{m} \otimes y) \leq 2^{-i}.$$

Thus for all $x, x', y, y' \in X$

$$|d(\boldsymbol{m} \otimes x, \boldsymbol{n} \otimes y) - d(\boldsymbol{m} \otimes x', \boldsymbol{n} \otimes y')| \leq 2^{1-i}. \tag{8}$$

Proof. The first assertion is proved by an easy induction on i. For the second, we use the triangle inequality and two applications of the first assertion:

$$\begin{aligned} d(\boldsymbol{m} \otimes x, \boldsymbol{n} \otimes y) - d(\boldsymbol{m} \otimes x', \boldsymbol{n} \otimes y')| &\leq d(\boldsymbol{m} \otimes x, \boldsymbol{m} \otimes x') + d(\boldsymbol{n} \otimes y, \boldsymbol{n} \otimes y') \\ &\leq 2^{-i} + 2^{-i} \end{aligned}$$

Theorem 16. *(S, s) is a final coalgebra of $M \otimes X$ on TriMS.*

Proof. Let $(X, e : X \to M \otimes X)$ be a coalgebra of $M \otimes -$ on TriMS. Forgetting the metric structure and the shortness of e, we know from Theorem 9 that there is a unique coalgebra morphism $f : X \to S$. This map preserves top, left, and right, and the main task is to show that f is a short map. We do this by showing that for $\varepsilon > 0$, and all $x, y \in X$,

$$d(f(x), f(y)) \leq d(x, y) + 2^{2-i}.$$

An easy induction on i shows that the composite below is a short map:

$$X \xrightarrow{\quad e \quad} M \otimes X \xrightarrow{M \otimes e} M \otimes M \otimes X \xrightarrow{M \otimes M \otimes e} \cdots \xrightarrow{M^{i-1} \otimes e} M^i \otimes X \ . \tag{9}$$

Let $x, y \in X$. Denote the image of x and y under the composite above by $\boldsymbol{m} \otimes x'$ and $\boldsymbol{m} \otimes y'$ respectively, where $\boldsymbol{m} = m_0, \ldots, m_{k-1} \in M^i$, $\boldsymbol{n} = n_0, \ldots, n_{k-1} \in M^i$ and $x', y' \in X$ Using the shortness of the map in (9), we see that

$$d(\boldsymbol{m} \otimes x', \boldsymbol{n} \otimes y') \leq d(x, y) \tag{10}$$

Since f is a coalgebra morphism to S we get:

$$\begin{aligned}(M^{i-1} \otimes s \circ \ldots \circ s)(f(x)) &= \boldsymbol{m} \otimes f(x') \\ (M^{i-1} \otimes s \circ \ldots \circ s)(f(y)) &= \boldsymbol{n} \otimes f(y')\end{aligned}$$

Now s is an isometry, as are $M \otimes s$, …, $M^{k-1} \otimes s$. Using the version of (9) that starts with S and s, we see that

$$d(f(x), f(y)) = d(\boldsymbol{m} \otimes f(x'), \boldsymbol{n} \otimes f(y')). \tag{11}$$

(Note the contrast to (10), where we have an inequality instead.)

Recall that the category TriMS has an initial object $I = \{\mathsf{top}, \mathsf{left}, \mathsf{right}\}$, and for any object in Z of TriMS, the initial map

$$i_Z : I \to Z$$

is an isometric embedding and so are the maps

$$M^i \otimes i_Z : M^i \otimes Z \to M^i \otimes Z$$

for all i (see Proposition 10). In our case we use the following diagram with both arrows being isometric embeddings

$$M^i \otimes S \hookleftarrow M^i \otimes I \hookrightarrow M^i \otimes X$$

to conclude that

$$d_{M^i \otimes S}(\boldsymbol{m} \otimes \mathsf{top}_S, n \otimes \mathsf{top}_S) = d_{M^i \otimes X}(\boldsymbol{m} \otimes \mathsf{top}_X, \boldsymbol{n} \otimes \mathsf{top}_X). \tag{12}$$

where top_X and top_S denote the images of $\mathsf{top} \in I$ under the two evident initial maps. Therefore

$$\begin{aligned}
d_{M^i \otimes S}(f(x), f(y)) &= d_{M^i \otimes S}(\boldsymbol{m} \otimes f(x'), \boldsymbol{n} \otimes f(y')) && \text{by (11)} \\
&\leq d_{M^i \otimes S}(\boldsymbol{m} \otimes \mathsf{top}_S, \boldsymbol{n} \otimes \mathsf{top}_S) + 2^{1-i} && \text{by Lemma 15} \\
&= d_{M^i \otimes X}(\boldsymbol{m} \otimes \mathsf{top}_X, \boldsymbol{n} \otimes \mathsf{top}_X) + 2^{1-i} && \text{by (12)} \\
&\leq d_{M^i \otimes X}(\boldsymbol{m} \otimes x', \boldsymbol{m} \otimes y') + 2^{1-i} + 2^{1-i} && \text{by Lemma 15} \\
&\leq d_{M^i \otimes X}(x, y) + 2^{2-i} && \text{by (10)}
\end{aligned}$$

This for all i completes the proof.

8 The Relation Between $\mathbb{S}$ and S

Let us summarize what we know about the spaces $\mathbb{S}$, G, and S:

1. $\mathbb{S}$ is the Sierpiński gasket, the unique non-empty compact subset of $\mathbb{R}^2$ satisfying
$$\mathbb{S} = \sigma_a(\mathbb{S}) \cup \sigma_b(\mathbb{S}) \cup \sigma_c(\mathbb{S}).$$
We know that as a tripointed set, it is a final coalgebra of $X \mapsto M \otimes X$.
2. As explained in Section 6.1, G is the set of expressions $m_1 \otimes m_2 \otimes \cdots \otimes m_k \otimes x$, where $m_i \in M$ and $x \in \{\mathsf{top}, \mathsf{left}, \mathsf{right}\}$; we also quotient by a certain set E of equations. G carries a metric structure, and as such it is an initial algebra of $X \mapsto M \otimes X$, considered as an endofunctor on TriMS. It also is the case that forgetting the metric structure, G is an initial algebra for this endofunctor on Tri.
3. S is the Cauchy completion of G. S is a final coalgebra of $X \mapsto M \otimes X$ on Tri, and also of this endofunctor on TriMS.

As tripointed sets, S and $\mathbb{S}$ are final coalgebras for the same endofunctor. Thus there are inverse coalgebra morphisms i and j:

$$\begin{array}{ccc} S & \xrightarrow{s} & M \otimes S \\ {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle M \otimes i} \\ \mathbb{S} & \xrightarrow[\sigma]{} & M \otimes \mathbb{S} \end{array} \qquad \begin{array}{ccc} \mathbb{S} & \xrightarrow{\sigma} & M \otimes \mathbb{S} \\ {\scriptstyle j}\downarrow & & \downarrow{\scriptstyle M \otimes j} \\ S & \xrightarrow[s]{} & M \otimes S \end{array}$$

Both i and j are Tri-morphisms, and they are the topic of this section.

Recall from Example 6 that σ^{-1} is short. By initiality in the metric setting, there is a unique short map $h : G \to \mathbb{S}$ such that the diagram below commutes:

$$\begin{array}{ccc} M \otimes G & \xrightarrow{\eta} & G \\ {\scriptstyle M \otimes h}\downarrow & & \downarrow{\scriptstyle h} \\ M \otimes \mathbb{S} & \xrightarrow[\sigma^{-1}]{} & \mathbb{S} \end{array} \tag{13}$$

As an illustration of how h works, $h(a \otimes b \otimes \mathsf{top}) = \sigma_a(\sigma_b(\mathsf{top}))$.

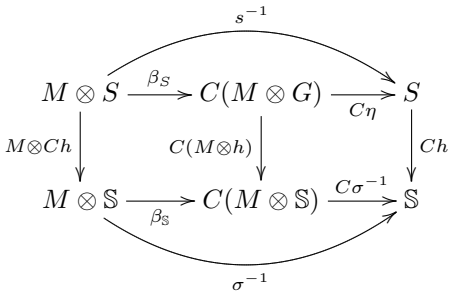
Example 7. The map h is not an isometry. For example, $d_G(b \otimes \mathsf{left}, a \otimes \mathsf{right}) = 1$. But $h(b \otimes \mathsf{left}) = h(\mathsf{left}) = (0,0)$, $h(a \otimes \mathsf{right}) = (3/4, \sqrt{3}/4)$, and

$$d_{\mathbb{S}}((0,0),(3/4,\sqrt{3}/4)) = \sqrt{3}/2.$$

We need an observation concerning the map i.

Lemma 17. *$i = Ch$.*

Proof. Consider the diagram below:



(14)

The square on the right comes from the diagram in (13), applying C. The square on the left is the naturality noted in Lemma 13. The top region commutes by the definition of s. It is easy to check that the bottom part of the diagram also commutes. It follows that $h : S \to \mathbb{S}$ is an algebra morphism. We take the inverses of s^{-1} and σ^{-1}, and we forget the metric structure. (We need to forget this metric structure, since σ is not a short map.) As a map of tripointed sets, h is a coalgebra morphism. By finality, $Ch = i$.

One last preliminary is needed before our main result in this section. It is an addendum to the triangle inequality.

Lemma 18. *Let $\triangle ABC$ be a triangle in the plane. Then*

$$\overline{AB} + \overline{AC} \leq \overline{BC}\sqrt{\frac{2}{1-\cos A}}$$

Proof. Let us write K for $\sqrt{2/(1-\cos A)}$. Since $\angle A$ is part of a triangle, $-1 < \cos A < 1$. Thus $K > 1$. Also, $2K^2 \cos A + 2 = 2K^2 - 2$. For all x and y,

$$\begin{aligned} &(K^2-1)x^2 - 2(K^2\cos A + 2)xy + (K^2-1)y^2 \\ &= (K^2-1)x^2 - 2(K^2-1)xy + (K^2-1)y^2 \\ &= (K^2-1)(x-y)^2 \\ &\geq 0 \end{aligned}$$

Thus the first of the four lines above is ≥ 0. Rearranging things,

$$x^2 + 2xy + y^2 \leq K^2(x^2 + y^2 - 2xy\cos A). \tag{15}$$

We apply (15) to the sides of our triangle, taking x to be $\overline{AB}$, and y to be $\overline{AC}$:

$$(\overline{AB} + \overline{AC})^2 \leq K^2(\overline{AB}^2 + \overline{AC}^2 - 2\overline{AB} \cdot \overline{AC} \cos A) = K^2(\overline{BC})^2.$$

In the line just above we used the law of cosines from trigonometry. Taking square roots gives the assertion in our lemma.

Two cases of Lemma 18 will be of interest. When $\angle A = 60^o$, we get $\overline{AB} + \overline{AC} \leq 2\overline{BC}$. When $\angle A = 120^o$, we get $\overline{AB} + \overline{AC} \leq \frac{2\sqrt{3}}{3}\overline{BC}$.

Theorem 19. *$i : S \to \mathbb{S}$ is a bilipschitz isomorphism. Specifically, for all $x, y \in S$,*

$$\frac{\sqrt{3}}{4} d(x, y) \leq d(i(x), i(y)) \leq d(x, y) \ . \tag{16}$$

Proof. We showed in Lemma 17 above that $i = Ch$. Since G is dense in S, we need only verify (16) for $h : G \to \mathbb{S}$. It is important to recall that distances in S are calculated using the quotient metric; in other words, we use the equations in Lemma 7. But for $\mathbb{S}$, we are using the Euclidean metric.

It is convenient to picture the points in $M^n \otimes I$ as arranged in a graph (see some pictures below). For each n, we take the distance of a segment in $M^n \otimes I$ to be 2^{-n}. Then the distance between two points in $M^n \otimes I$ (as we have been considering it) is exactly the length of the shortest path in the graph. Thus, we need to show that the graph-theoretic distance between two points is at most the Euclidean distance multiplied by $4\sqrt{3}/3$.

Lemma 20. *For all $x \in M^n \otimes I$, there is a path from the top of $M^n \otimes I$ to x whose graph-theoretic distance is at most $2\sqrt{3}/3$ times the Euclidean distance.*

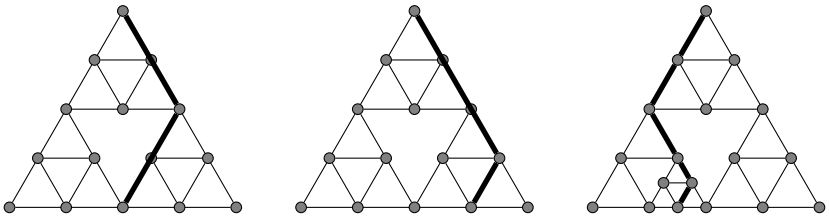
Proof. For each x, we first construct a *principal path* $\mathsf{top} = p_0, p_1, \ldots, p_n = x$ from the top to x. Each p_i will belong to $M^i \otimes I$. We take $p_0 = \mathsf{top}$.

If $x = p_0$, then the path is simply p_0.

If x belongs to the top main triangle tr_a, then it is of the form $a \otimes y$. And by induction hypothesis, we already have principal path for y, say $\mathsf{top} = p_0, p_1, \ldots, p_n = y$. We take the principal path for x to be $\mathsf{top} = a \otimes p_0, a \otimes p_1, \ldots, a \otimes p_n = x$. (In other words, we take the principal path for y and then apply σ_a to each point.)

If x belongs to the left main triangle tr_b, then it is $b \otimes y$. Again by induction hypothesis, we already have principal path for y, say $\mathsf{top} = p_0, p_1, \ldots, p_n = y$. We have three subcases. If p_1 is to the left of p_0, then we take the principal path for x to be $\mathsf{top} = p_0, b \otimes p_2, \ldots, b \otimes p_n = b \otimes y = x$. Note that we omit $b \otimes p_1$ in this subcase. The second subcase is when p_1 is to the right of p_0. In this case, we take the principal path for x to be $\mathsf{top} = p_0, b \otimes p_1, b \otimes p_2, \ldots, b \otimes p_n = b \otimes y = x$. In the last subcase, x is directly under $p_0 = \mathsf{top}$. In this case, $x = b \otimes \mathsf{right} = c \otimes \mathsf{left}$, and we may take the principal path for x to be either $\mathsf{top}, a \times \mathsf{left}, x$, or else $\mathsf{top}, a \times \mathsf{right}, x$

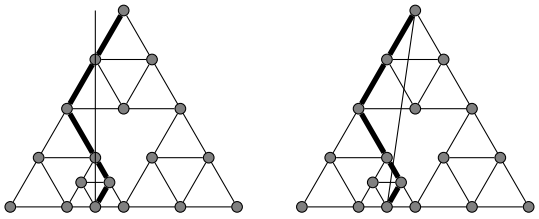
If x belongs to the right main triangle tr_c, then the definitions work similarly. Here are some illustrations of points x and the principal paths for them.



The path in the first figure is used in the second, and the path in the second is used in the third. In fact, the path in the first figure could go the other way, left instead of right. This possibility is important, as we shall see shortly.

We might have to modify the end of the path, because we would like to arrange that no point on the principal path for x is directly above x, except when x is of the form $a \otimes a \otimes \cdots \otimes a \otimes b \otimes \mathsf{left} = a \otimes a \otimes \cdots \otimes a \otimes c \otimes \mathsf{right}$. For all other x, we can indeed arrange that the principal path not have any point directly above x. To do this, we would only need to take the last two points on the path, and change the direction.

An easy induction on n now shows that the path zig-zags across the vertical line through x, as shown on the left below.



Next, it follows easily that if we incline the line so that it intersects top, the principal path zig-zags across it. This uses the assumption that no point on the principal path is directly above x. For example, if x is in the left main triangle (as in our pictures above), then the inclined line on the right still intersects all segments determined by the principal path.

To complete the proof, we use Lemma 18 in each of the triangles formed by the principal path and the line from x to top.

Now we complete the proof of Theorem 19. We are done when one of the points is top; and similarly we are done when one of the points is left or right. In both of those cases, we have a better bound than the general bound stated in Theorem 19: $\sqrt{3}/2 \cdot d(x,y) \leq d(i(x), i(y))$. But in general, we combine paths to one of the connection points, and again use Lemma 18; we use the special case of 60^o-triangles. We would get $\sqrt{3}/4 \cdot d(x,y) \leq d(i(x), i(y))$.

This completes the proof of Theorem 19.

Remark 1. The bound of $\frac{2}{3}\sqrt{3}$ in Lemma 20 is sharp. The lower bound of $\sqrt{3}/4 \approx .433$ in Theorem 19 may be improved to $1/2$. However, the argument in this case is much more involved than what we have seen in Lemma 20. In effect, one has to show that all lines in $\mathbb{R}^2$ connecting pairs of points in $M^n \otimes I$ may be "zig-zagged" by paths in $M^n \otimes I$ itself. (One would like to use results like those in Aichholzer [4], but our graphs are not triangulations; they involve colinear triples, for example.) Of course, it is sufficient in Theorem 19 to have *any* bound.

Metric spaces with lipschitz maps The reader might wonder about the significance of the bilipschitz result, Theorem 19. This points to the subtle question of the maps in the category of "metric spaces." We have been working with metric spaces and short maps. But here are settings where one makes other choices. Another choice would be the continuous maps. An even smaller choice would be the Lipschtiz continuous functions. These are the maps $f : X \to Y$ of metric spaces with the property that there is a constant K such that $d(f(x), f(y)) \leq K \cdot d(x, y)$ for all $x, y \in X$. Let MS^{Lip} be the category of 1-bounded metric spaces and Lipschitz continuous functions. An isomorphism in MS^{Lip} is exactly an bilipschtiz bijection. Frequently in the literature on metric spaces one does indeed identify spaces which are bilipschitz isomorphism. In comparison with other notions of equivalence for metric spaces, we have

$$\text{isometry} \Longrightarrow \text{bilipschitz isomorphism} \Longrightarrow \text{homeomorphism}$$

We had hoped to find $\mathbb{S}$ and S to be related by an isometry, but this was not to be. (The parallel result for the unit interval did hold, but this seems like a very special result.) We have the following corollary to Theorem 19:

Corollary 21. *$\sigma : M \otimes \mathbb{S} \to \mathbb{S}$ is a coalgebra in MS^{Lip}, and it is isomorphic to (S, s).*

Having said this, a good next question in this line of work would be: in MS^{Lip}, is $\mathbb{S}$ a final coalgebra of $M \otimes -$? We have not yet pursued this question.

9 Conclusion

This paper presented some specific results on the Sierpiński gasket $\mathbb{S}$. We considered the categories Tri of tripointed sets and TriMS of tripointed metric spaces. Both categories carry an endofunctor $X \to M \otimes X$ defined in a natural way. The initial algebra of this functor on TriMS is the set of finite addresses (tensor expressions) of points in $\mathbb{S}$. Its Cauchy completion, S, is the final coalgebra of the endofunctor. This finality result goes through in the metric setting, and this is one of our main results. As tripointed sets, S and $\mathbb{S}$ are isomorphic. But the isomorphism is not an isometry of the associated metric spaces. Indeed, the natural metric structure on $\mathbb{S}$ as a coalgebra is not an isometry. The best we can say is that as metric spaces, $\mathbb{S}$ and S are bilipschitz equivalent: there is a bijection between these spaces with the property that distances in one space

are uniformly bounded by multiples of the corresponding distances in the other space. This bilipschitz equivalence involves work on paths in the finite approximations to the Sierpiński gasket. It also extends to show that the coalgebras (S, s) and $(\mathbb{S}, \sigma)$ are isomorphic in MS^{Lip}.

Naturally, the main open question in all of this work is whether the results on the Sierpiński gasket extend to all of the other fractal sets in $\mathbb{R}^n$. That is, it is possible to consider specific self-similar sets, such as the Sierpiński carpet or the unit interval, and then render them as finial coalgebras, the same way as we have seen for the Sierpiński gasket. The initial algebras in these cases are also interesting. However, we do not yet have a general theory that offers an insightful generalization of these examples.

Acknowledgment

We thank anonymous referees for comments and suggestions which have improved our presentation.

References

1. J. Adámek, Free algebras and automata realizations in the language of categories, *Comment. Math. Univ. Carolinae* 14 (1974), 589–602.
2. J. Adámek, On final coalgebras of continuous functors. *Theoret. Comput. Sci.* 294 (2003), no. 1–2, 3–29.
3. J. Adámek, S. Milius and L. S. Moss, Initial Algebras and Terminal Coalgebras: a Survey, unpublished ms. 2011.
4. O. Aichholzer. The path of a triangulation. *Proc. 15th Ann. ACM Symp. Computational Geometry*, Miami Beach, Florida, USA, 1999, 14–23.
5. M. Barr, Terminal coalgebras in well-founded set theory, *Theoret. Comput. Sci.* 114 (1993) no. 2, 299–315. Additions and corrections in *Theoret. Comput. Sci.* 124 (1994), no. 1, 189–192.
6. P. Freyd, Real coalgebra, post on the Categories mailing list, December 22, 1999, available at www.mta.ca/∼cat-dist
7. I. Hasuo, B. Jacobs, and M. Niqui, Coalgebraic representation theory of fractals (Extended Abstract), *Proc. Mathematical Foundations of Programming Semantics (MFPS XXVI)*, Electr. Notes Comp. Sci. 265:351-368, 2010.
8. John Hutchinson, Fractals and self-similarity, *Indiana Univ. Math. Journal* 30 No. 5 (1981), 713747.
9. J. Lambek, A Fixpoint Theorem for Complete Categories, *Math. Z.* 103 (1968), 151–161.
10. T. Leinster, A general theory of self-similarity, *Advances in Mathematics* 226 (2011), 29353017.