General self-similarity: an overview

Tom Leinster

Department of Mathematics, University of Glasgow
www.maths.gla.ac.uk/∼tl
tl@maths.gla.ac.uk

Abstract

Consider a self-similar space X. A typical situation is that X looks like several copies of itself glued to several copies of another space Y, and that Y looks like several copies of itself glued to several copies of X— or the same kind of thing with more than two spaces. Thus, we have a system of simultaneous equations in which the right-hand sides (the gluing instructions) are 'higher-dimensional formulas'.

This idea is developed in detail in [Lei1] and [Lei2]. The present informal seminar notes explain the theory in outline.

I want to tell you about a very general theory of self-similar objects that I've been developing recently. In principle this theory can handle self-similar objects of any kind whatsoever—algebraic, analytic, geometric, probabilistic, and so on. At present it's the topological case that I understand best, so that's what I'll concentrate on today. This concerns the self-similar or fractal spaces that you've all seen pictures of.

Some of the most important self-similar spaces in mathematics are Julia sets. For the purposes of this talk you don't need to know the definition of Julia set, but the bare facts are these: to every complex rational function f there is associated a closed subset $J(f)$ of the Riemann sphere $\mathbb{C} \cup \{\infty\}$, its Julia set, which almost always has a highly intricate, fractal appearance. If you look in a textbook on complex dynamics, you'll find theorems about 'local self-similarity' of Julia sets. For example, given almost any point z in a Julia set, points locally isomorphic to z occur densely throughout the set [Mil, Ch. 4]. On the other hand, the kind of self-similarity I'm going to talk about today is the dual idea, 'global self-similarity', where you say that the *whole* space looks like several copies of itself stuck together—or some statement of the kind. So it's a top-down, rather than bottom-up, point of view.

A long-term goal is to develop the algebraic topology of self-similar spaces. The usual invariants coming from homotopy and homology are pretty much useless (e.g. for a fractal subset of the plane all you've got is π_1, which is usually either trivial or infinite-dimensional), but a description by global self-similarity is discrete and so might provide useful invariants at some point in the future.
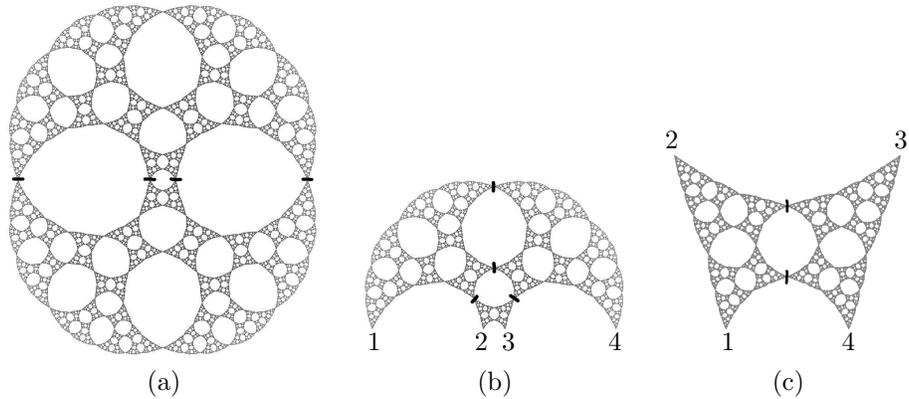
Figure 1: (a) The Julia set of $z \longmapsto (2z/(1+z^2))^2$; (b), (c) certain subsets

This theory is about self-similarity as an *intrinsic* structure on an object: there is no reference to an ambient space, and in fact no ambient space at all. This is like doing group theory rather than representation theory, or the theory of abstract manifolds rather than the theory of manifolds embedded in $\mathbb{R}^n$. We can worry about representations later. For instance, the Koch snowflake is just a circle for us: its self-similar aspect is the way it's embedded in the plane.

Later I'll show you the general language of self-similarity, but first here are some concrete examples to indicate the kind of situation that I want to describe.

1   First example: a Julia set

Let's look at one particular Julia set in detail: Figure 1(a). I'll write I_1 for this Julia set. It clearly has reflectional symmetry in a horizontal axis, so if we cut at the four points shown then we get a decomposition

$$I_1 = \qquad\qquad \tag{1}$$

where I_2 is a certain space with four marked points (or 'basepoints'). Now consider I_2 (Figure 1(b)). Cutting at the points shown gives a decomposition

$$I_2 = \qquad\qquad \tag{2}$$

2

where I_3 is another space with four marked points. Next, consider I_3 (Figure 1(c)); it decomposes as

$$I_3 = \qquad . \tag{3}$$

Here we can stop, since no new spaces are involved. Or nearly: there's a hidden role being played by the one-point space I_0, since that's what we've been gluing along, and I'll record the trivial equation

$$I_0 = I_0. \tag{4}$$

What we have here is a system of four simultaneous equations, with the unusual feature that the right-hand sides are not algebraic formulas of the usual type, but rather '2-dimensional formulas' expressing how the spaces are glued together.

(There's a conceptual link between this and the world of n-categories, where there are 2-dimensional and higher-dimensional morphisms which you're allowed to compose or 'glue' in various ways. Both can be regarded as a kind of higher-dimensional algebra. The *cognoscenti* will see a technological link too: in both contexts the gluing can be described by pullback-preserving functors on categories of presheaves.)

The simultaneous equations (1)–(4) can be expressed as follows. First, we have our spaces I_1, I_2, I_3 with their marked points, which together form a functor from the category

$$\mathbb{A} = \left(\begin{array}{c} 1 \\ 0 \Longrightarrow 2 \\ 3 \end{array} \right)$$

to the category **Set** of sets (or a category of spaces, but let's be conservative for the moment). Second, the gluing formulas define a functor

$$G : [\mathbb{A}, \mathbf{Set}] \longrightarrow [\mathbb{A}, \mathbf{Set}],$$

where $[\mathbb{A}, \mathbf{Set}]$ is the category of functors $\mathbb{A} \longrightarrow \mathbf{Set}$: given $X \in [\mathbb{A}, \mathbf{Set}]$, put

$$(G(X))_1 \quad = \quad = (X_2 \amalg X_2)/\sim,$$

3

$$(G(X))_2 \quad = \quad \qquad\qquad\qquad = (X_2 \amalg X_2 \amalg X_3)/\sim,$$

and so on. (I've drawn the pictures as if X_0 were a single point.) Then the simultaneous equations assert precisely that

$$I \cong G(I)$$

—I is a fixed point of G.

Although these simultaneous equations have many solutions (G has many fixed points), I is in some sense the universal one. This means that the simple diagrams (1)–(4) contain just as much information as the apparently very complex spaces in Figure 1: for given the system of equations, we recover these spaces as the universal solution. Caveats: we're only interested in the intrinsic, topological aspects of self-similar spaces, not how they're embedded into an ambient space (in this case, the plane) or the metrics on them.

Next we have to find some general way of making rigorous the idea of 'gluing formula'; so far I've just drawn pictures. We have a small category $\mathbb{A}$ whose objects index the spaces involved, and I claim that the self-similarity equations are described by a functor $M : \mathbb{A}^{\mathrm{op}} \times \mathbb{A} \longrightarrow \mathbf{Set}$ (a '2-sided $\mathbb{A}$-module'). The idea is that for $b, a \in \mathbb{A}$,

$$M(b,a) \quad = \quad \{\text{copies of the } b\text{th space used in the gluing formula}$$
$$\text{for the } a\text{th space}\}.$$

Take, for instance, our Julia set. In the gluing formula for I_2, the one-point space I_0 appears 8 times, I_1 doesn't appear at all, I_2 appears twice, and I_3 appears once, so, writing n for an n-element set,

$$M(0,2) = 8, \qquad M(1,2) = \emptyset, \qquad M(2,2) = 2, \qquad M(3,2) = 1.$$

(It's easy to get confused between the arrows $b \longrightarrow a$ in $\mathbb{A}$ and the elements of $M(b,a)$. The arrows of $\mathbb{A}$ say nothing whatsoever about the gluing formulas, although they determine where gluing may *potentially* take place. The elements of M embody the gluing formulas themselves.)

This is an example of what I'll call a 'self-similarity system':

Definition *A **self-similarity system** is a small category $\mathbb{A}$ together with a functor $M : \mathbb{A}^{\mathrm{op}} \times \mathbb{A} \longrightarrow \mathbf{Set}$ such that*

   a. for each $a \in \mathbb{A}$, the set $\coprod_{c,b \in \mathbb{A}} \mathbb{A}(c,b) \times M(b,a)$ is finite

   b. (a condition to be described later).

4

Part (a) says that in the system of simultaneous equations, each right-hand side is a gluing of only a *finite* family of spaces. So we might have infinitely many spaces (in which case $\mathbb{A}$ would be infinite), but each one is described as a finite gluing. The condition is more gracefully expressed in categorical language: 'for each a, the category of elements of $M(-, a)$ is finite'.

As in our example, any self-similarity system $(\mathbb{A}, M)$ induces an endofunctor G of $[\mathbb{A}, \mathbf{Set}]$. This works as follows. First note that if A is a ring (not necessarily commutative), Y a right A-module, and X a left A-module, there is a tensor product $Y \otimes_A X$ (a mere abelian group). Similarly, if $\mathbb{A}$ is a small category, $Y : \mathbb{A}^{\mathrm{op}} \longrightarrow \mathbf{Set}$ a contravariant functor, and $X : \mathbb{A} \longrightarrow \mathbf{Set}$ a covariant functor, there is a tensor product

$$Y \otimes X = \left(\coprod_{a \in \mathbb{A}} Ya \times Xa \right) / \sim$$

(a mere set): see [Mac, IX.6]. So if $(\mathbb{A}, M)$ is a self-similarity system then there is an endofunctor $G = M \otimes -$ of $[\mathbb{A}, \mathbf{Set}]$ defined by

$$(M \otimes X)(a) = M(-, a) \otimes X = \left(\coprod_{b \in \mathbb{A}} M(b, a) \times Xb \right) / \sim .$$

($X \in [\mathbb{A}, \mathbf{Set}]$, $a \in \mathbb{A}$). We are interested in finding a fixed point of G that is in some sense 'universal'.

2   Second example: Freyd's Theorem

The second example I'll show you comes from a very different direction. In December 1999, Peter Freyd posted a message [Fre] on the categories mailing list that caused a lot of excitement, especially among the theoretical computer scientists.

We'll need some terminology. Given a category $\mathcal{C}$ and an endofunctor G of $\mathcal{C}$, a G-**coalgebra** is an object X of $\mathcal{C}$ together with a map $\xi : X \longrightarrow GX$. (For instance, if $\mathcal{C}$ is a category of modules and $GX = X \otimes X$ then a G-coalgebra is a coalgebra—not necessarily coassociative—in the usual sense.) A **map** $(X, \xi) \longrightarrow (X', \xi')$ of coalgebras is a map $X \longrightarrow X'$ in $\mathcal{C}$ making the evident square commute. Depending on what G is, the category of G-coalgebras may or may not have a terminal object, but if it does then it's a fixed point:

Lemma 1 (Lambek [Lam]) *Let $\mathcal{C}$ be a category and G an endofunctor of $\mathcal{C}$. If (I, ι) is terminal in the category of G-coalgebras then $\iota : I \longrightarrow GI$ is an isomorphism.*

Proof Short and elementary.   □

Here's what Freyd said, modified slightly. Let $\mathcal{C}$ be the category whose objects are diagrams $X_0 \underset{v}{\overset{u}{\rightrightarrows}} X_1$ where X_0 and X_1 are sets and u and v are
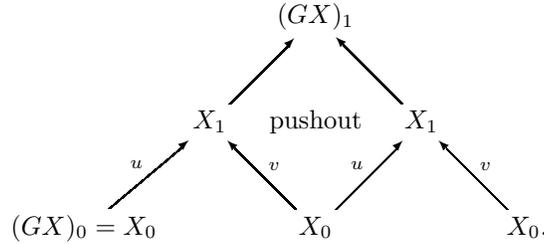
injections with disjoint images; then an object of $\mathcal{C}$ can be drawn as
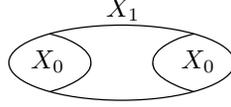
where the copies of X_0 on the left and the right are the images of u and v respectively. A map $X \longrightarrow X'$ in $\mathcal{C}$ consists of functions $X_0 \longrightarrow X_0'$ and $X_1 \longrightarrow X_1'$ making the evident two squares commute. Now, given $X \in \mathcal{C}$ we can form a new object GX of $\mathcal{C}$ by gluing two copies of X end to end:
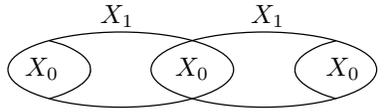
Formally, GX is defined by pushout:

For example, the unit interval with its endpoints distinguished forms an object

$$I = \left(\{\star\} \underset{1}{\overset{0}{\rightrightarrows}} [0,1] \right)$$

of $\mathcal{C}$, and GI is naturally described as an interval of length 2, again with its endpoints distinguished:

$$GI = \left(\{\star\} \underset{2}{\overset{0}{\rightrightarrows}} [0,2] \right).$$

So there is a coalgebra structure $\iota : I \longrightarrow GI$ on I given by multiplication by two. Freyd's Theorem says that this is, in fact, the *universal* example of a G-coalgebra:

Theorem 2 (Freyd$+\varepsilon$) (I, ι) *is terminal in the category of G-coalgebras.*

I won't go into the proof, but clearly it's going to have to involve the completeness of the real numbers in an essential way. Once you've worked out what 'terminal coalgebra' is saying, it's easy to see that the proof is going to be something to do with binary expansions. Note that although ι is an isomorphism (as predicted by Lambek's Lemma), this by no means determines (I, ι): consider, for instance, the unique coalgebra satisfying $X_0 = X_1 = \emptyset$, or the evident coalgebra in which $X_0 = \{\star\}$ and $X_1 = [0,1] \cap \{\text{dyadic rationals}\}$.

The striking thing about Freyd's result is that we started with just some extremely primitive notions of set, function, and gluing—and suddenly, out popped the real numbers. What excited the computer scientists was that it suggested new ways of representing the reals. But its relevance for us here is that it describes the self-similarity of the unit interval—in other words, the fact that it's isomorphic to two copies of itself stuck end to end. We'll take this idea of Freyd, describing a very simple self-similar space as a terminal coalgebra, and generalize it dramatically.

Freyd's Theorem concerns the self-similarity system $(\mathbb{A}, M)$ in which

$$\mathbb{A} = (0 \rightrightarrows 1)$$

and $M : \mathbb{A}^{\mathrm{op}} \times \mathbb{A} \longrightarrow \mathbf{Set}$ is given by

$$M(-, 0): \qquad \{\star\} \leftleftarrows \emptyset$$
$$0 \Big\downarrow \Big\downarrow 1 \qquad \qquad \Big\downarrow\Big\downarrow$$
$$M(-, 1): \qquad \{0, \tfrac{1}{2}, 1\} \overset{\mathrm{inf}}{\underset{\mathrm{sup}}{\leftleftarrows}} \{[0, \tfrac{1}{2}], [\tfrac{1}{2}, 1]\}.$$

(Here $M(0, 1)$ is just a 3-element set and $M(1, 1)$ a 2-element set, but their elements have been named suggestively.) The category $\mathcal{C}$ is a full subcategory of $[\mathbb{A}, \mathbf{Set}]$, and the endofunctor G of $\mathcal{C}$ is the restriction of the endofunctor $M \otimes -$ of $[\mathbb{A}, \mathbf{Set}]$.

The only thing that looks like a barrier to generalization is the condition that $u, v : X_0 \longrightarrow X_1$ are injective with disjoint images (which is the difference between $\mathcal{C}$ and $[\mathbb{A}, \mathbf{Set}]$). If this were dropped then $(\{\star\} \rightrightarrows \{\star\})$ would be the terminal coalgebra, so the theorem would degenerate entirely. It turns out that the condition is really a kind of flatness.

A module X over a ring is called 'flat' if the functor $- \otimes X$ preserves finite limits. There is an analogous definition when X is a functor, but actually we want something weaker:

Definition *Let $\mathbb{A}$ be a small category. A functor $X : \mathbb{A} \longrightarrow \mathbf{Set}$ is* **nondegenerate** *if the functor*

$$- \otimes X : [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}] \longrightarrow \mathbf{Set}$$

preserves finite connected *limits. Write $[\mathbb{A}, \mathbf{Set}]_{\mathrm{nondegen}}$ for the category of non-degenerate functors $\mathbb{A} \longrightarrow \mathbf{Set}$ and natural transformations between them.*

It looks as if this is very abstract, that in order to show that X was degenerate you'd need to test it against all possible finite connected limits in $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$, but in fact there's an equivalent explicit condition. This can be used to show that in the case at hand, a functor $X : \mathbb{A} \longrightarrow \mathbf{Set}$ is degenerate precisely when the two functions $u, v : X_0 \longrightarrow X_1$ are injective with disjoint images.

The missing condition (b) in the definition of self-similarity system is that for each $b \in \mathbb{A}$, the functor $M(b, -) : \mathbb{A} \longrightarrow \mathbf{Set}$ is nondegenerate. This guarantees that the endofunctor $M \otimes -$ of $[\mathbb{A}, \mathbf{Set}]$ restricts to an endofunctor of $[\mathbb{A}, \mathbf{Set}]_{\text{nondegen}}$. A terminal coalgebra for this restricted endofunctor is called a **universal solution** of the self-similarity system; if one exists, it's unique up to canonical isomorphism. Lambek's Lemma implies that if (I, ι) is a universal solution then, as the terminology suggests, $M \otimes I \cong I$. Freyd's Theorem describes the universal solution of a certain self-similarity system.

Before we move on, I'll show you a version of Freyd's Theorem in which the unit interval is characterized not just as a set but as a topological space. The simplest thing would be to change 'set' to 'space' and 'function' to 'continuous map' throughout the above. Unsurprisingly, this gives a boring topology on $[0, 1]$: the indiscrete one, as it happens. But all we need to do to get the usual topology is to insist that the maps u and v are closed. That is:

Theorem 2′ *Define $\mathcal{C}'$ and G' as $\mathcal{C}$ and G were defined above, changing 'set' to 'space' and 'function' to 'continuous map', and adding the condition that u and v are closed maps. Then the terminal G'-coalgebra is (I, ι) where $[0, 1]$ is equipped with the Euclidean topology.*

Generally, a functor $X : \mathbb{A} \longrightarrow \mathbf{Top}$ is **nondegenerate** if its underlying $\mathbf{Set}$-valued functor is nondegenerate and Xf is a closed map for every map f in $\mathbb{A}$. This gives a notion of **universal topological solution**, just as in the set-theoretic scenario. So Theorem 2′ describes the universal topological solution of the Freyd self-similarity system.

3   Results

Just as some systems of equations have no solution, some self-similarity systems have no universal solution. But it's easy to tell whether there is one:

Theorem 3 *A self-similarity system has a universal solution if and only if it satisfies a certain condition $\mathbf{S}$.*

I won't say what $\mathbf{S}$ is, but it is completely explicit. So too is the construction of the universal solution when it does exist; it is similar in spirit to constructing the real numbers as infinite decimals, although smoother than that would suggest.

Let $(\mathbb{A}, M)$ be a self-similarity system with universal solution (I, ι). Then there is a canonical topology on each space Ia, with the property that all the maps If are continuous and closed and all the maps ι_a are continuous. Again, the topology can be defined in a completely explicit way.

Theorem 4 *(I, ι) with this topology is the universal topological solution.*

Call a space **self-similar** if it is homeomorphic to Ia for some self-similarity system $(\mathbb{A}, M)$ and some $a \in \mathbb{A}$, where (I, ι) is the universal solution of $(\mathbb{A}, M)$. There is a 'recognition theorem' giving a practical way to recognize universal solutions, and this generates some examples of self-similar spaces:

- $[0, 1]$, as in the Freyd example

- $[0, 1]^n$ for any $n \in \mathbb{N}$; more generally, the product of two self-similar spaces is self-similar

- the n-simplex Δ^n for any $n \in \mathbb{N}$, by barycentric subdivision

- the Cantor set (isomorphic to two disjoint copies of itself)

- Sierpiński's gasket, and many other spaces defined by iterated function systems.

The proof of Theorem 4 involves showing that each of the spaces Ia is compact and metrizable (or equivalently, compact Hausdorff with a countable basis of open sets). So every self-similar space is compact and metrizable. The shock is that the converse holds:

Theorem 5 *For topological spaces,*

$$\text{self-similar} \iff \text{compact metrizable}.$$

This looks like madness, so let me explain.

First, the result is non-trivial: the classical result that every nonempty compact metrizable space is a continuous image of the Cantor set can be derived as a corollary.

Second, the word 'self-similar' is problematic (even putting aside the obvious objection: what could be more similar to a thing than itself?) When we formalized the idea of a system of self-similarity equations, we allowed ourselves to have infinitely many equations, even though each individual equation could involve only finitely many spaces. So there might be infinite regress: for instance, X_1 could be described as a copy of itself glued to a copy of X_2, X_2 as a copy of itself glued to a copy of X_3, and so on. Perhaps 'recursively realizable' would be better than 'self-similar'.

Finally, this theorem does not exhaust the subject: it characterizes the spaces admitting *at least one* self-similarity structure, but a space may be self-similar in several essentially different ways.

There's a restricted version of Theorem 5. Call a space **discretely self-similar** if it is homeomorphic to one of the spaces Ia arising from a self-similarity system $(\mathbb{A}, M)$ in which the category $\mathbb{A}$ is discrete (has no arrows except for identities). The Cantor set is an example: we can take $\mathbb{A}$ to be the one-object discrete category.

Theorem 6 *For topological spaces,*

$$\text{discretely self-similar} \iff \text{totally disconnected compact metrizable}.$$

Totally disconnected compact Hausdorff spaces are the same as profinite spaces, and the metrizable ones are those that can be written as the limit of a *countable* system of finite discrete spaces. For instance, the underlying space of the absolute Galois group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is discretely self-similar.

If you find the general notion of self-similarity too inclusive, you may prefer to restrict to finite categories $\mathbb{A}$, which gives the notion of **finite self-similarity**. This means that the system of equations is finite. A simple cardinality argument shows that almost all compact metrizable spaces are not finitely self-similar.

I'll finish with two conjectures. They both say that certain types of compact metrizable space *are* finitely self-similar.

Conjecture 1 *Every finite simplicial complex is finitely self-similar.*

I strongly believe this to be true. The standard simplices Δ^n are finitely self-similar, and if we glue a finite number of them along faces then the result should be finitely self-similar too. For example, by gluing two intervals together we find that the circle is finitely self-similar. Note that any manifold is as locally self-similar as could be, in the sense of the introduction: every point is locally isomorphic to every other point.

More tentatively,

Conjecture 2 *The Julia set $J(f)$ of any complex rational function f is finitely self-similar.*

This brings us full circle: it says that in the first example, we could have taken any rational function f and seen the same type of behaviour: after a finite number of decompositions, no more new spaces I_n appear. Both $J(f)$ and its complement are invariant under f, so f restricts to an endomorphism of $J(f)$ and this endomorphism is, with finitely many exceptions, a $\deg(f)$-to-one mapping. This suggests that f itself should provide the global self-similarity structure of $J(f)$, and that if $(\mathbb{A}, M)$ is the corresponding self-similarity system then the sizes of $\mathbb{A}$ and M should be bounded in terms of the degree of f.

References

[Fre]   Peter Freyd, Real coalgebra, post on categories mailing list, 22 December 1999, available via www.mta.ca/∼cat-dist

[Lam]   Joachim Lambek, A fixpoint theorem for complete categories, *Mathematische Zeitschrift* 103 (1968), 151–161.

[Lei1]   Tom Leinster, A general theory of self-similarity I, e-print math.DS/0411344.

[Lei2]   Tom Leinster, A general theory of self-similarity II: recognition, e-print math.DS/0411345.

[Mac]   Saunders Mac Lane, *Categories for the Working Mathematician*, Graduate Texts in Mathematics 5, Springer, 1971.

[Mil]   John Milnor, *Dynamics in One Complex Variable*, Vieweg, 1999.