Embedding a snowflake metric space into Euclidean space

Marie A. Snipes
Kenyon College

joint work with Jim Skon and Preston Pennington

AMS Spring Central and Western Joint Sectional Meeting
University of Hawaii
March 22, 2019

The project

General question: How can we best represent a metric space with Euclidean coordinates?

There are metric spaces that do not embed bi-Lipschitzly in any Euclidean space. However, if the metric space is doubling, then Assouad's theorem guarantees that every snowflake of the space does embed bi-Lipschitzly in some $\mathbb{R}^n$.

The project

General question: How can we best represent a metric space with Euclidean coordinates?

There are metric spaces that do not embed bi-Lipschitzly in any Euclidean space. However, if the metric space is doubling, then Assouad's theorem guarantees that every snowflake of the space does embed bi-Lipschitzly in some $\mathbb{R}^n$.

A classical example of doubling metric space with no bi-Lip embedding into $\mathbb{R}^n$ constructed by Laakso.

We are interested in embedding the snowflaked Laakso space into $\mathbb{R}^n$. This is joint work with Jim Skon (Kenyon CS) and Preston Pennington (Kenyon '20).

Graph Metric Spaces

Recall that a *metric space* is a set X with a distance function $d : X \times X \to \mathbb{R}^+$ that satisfies

- $d(x, y) = 0 \Leftrightarrow x = y$
- $d(x, y) = d(y, x)$
- $d(x, y) \leq d(x, z) + d(z, y)$

for all points x, y, z in X.

Graph metric spaces: Given graph (V, E), define distance on V so that $d(x, y)$ is the length ($\#$ edges) of the shortest path between vertices x and y.

Can also assign positive weights to the edges for non-integer distances.

Doubling Metric Spaces

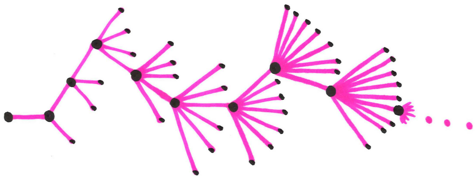
A metric space (X, d) is *doubling* if there exists a constant $C \geq 1$ so that every ball of radius r can be "covered by" at most C balls of radius at most $r/2$.

Doubling Metric Spaces

A metric space (X, d) is *doubling* if there exists a constant $C \geq 1$ so that every ball of radius r can be "covered by" at most C balls of radius at most $r/2$.

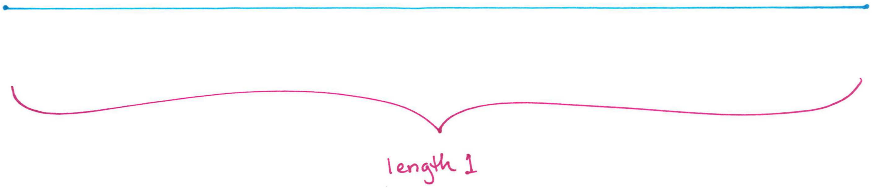
Examples

- $\mathbb{R}^n$ and subsets of $\mathbb{R}^n$, for all n: **Doubling**
- The following infinite graph with the path metric: **Not Doubling**



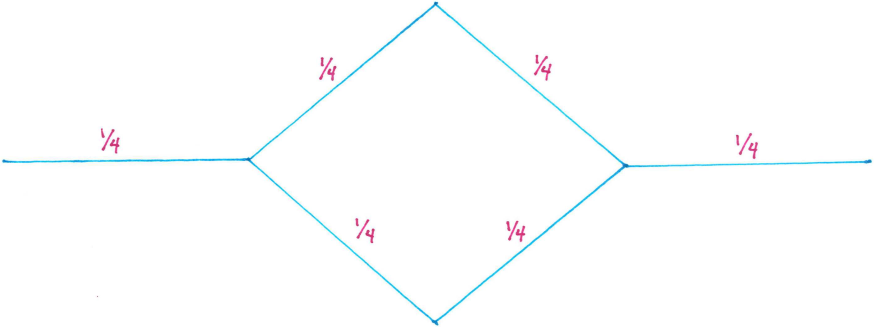
The Laakso Space (as simplified by Lang and Plaut)

Construct as the limit of a sequence of graphs



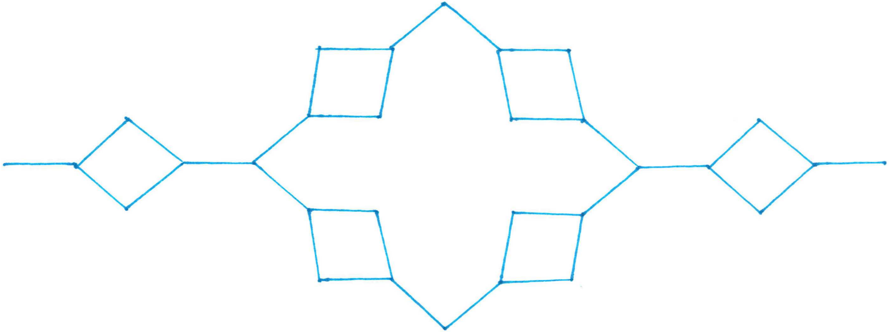
The Laakso Space (as simplified by Lang and Plaut)

Construct as the limit of a sequence of graphs



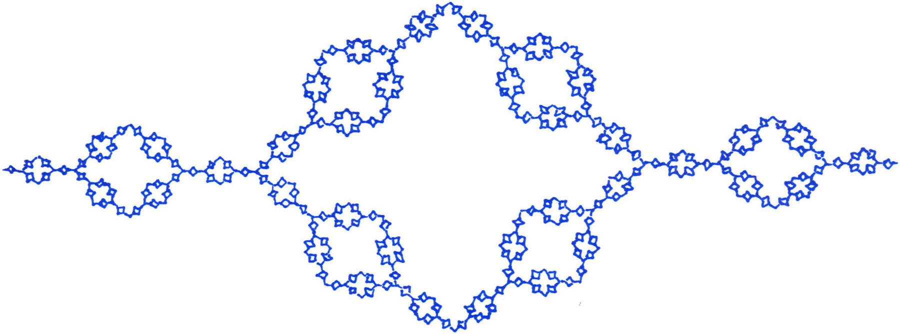
The Laakso Space (as simplified by Lang and Plaut)

Construct as the limit of a sequence of graphs



The Laakso Space (as simplified by Lang and Plaut)

Construct as the limit of a sequence of graphs



Metric Space Embeddings

A map $f : X \to Y$ is an *embedding* if it is a homeomorphism onto its image.

Competing goals:

Find an embedding into the simplest (lowest dimensional) space possible!

Also look for an embedding that doesn't distort the metric too much!

- Isometry: distances preserved exactly

$$d(x, y) = d(f(x), f(y))$$

- Bi-Lipschitz map: distances distorted by a bounded amount

$$\frac{1}{L} \cdot d(x, y) \leq d(f(x), f(y)) \leq L \cdot d(x, y)$$

Metric Space Embeddings $f : X \to \mathbb{R}^n$

Isometric embeddings into $\mathbb{R}^n$: too much to hope for.

Metric Space Embeddings $f : X \to \mathbb{R}^n$

Isometric embeddings into $\mathbb{R}^n$: too much to hope for.

bi-Lipschitz embeddings into $\mathbb{R}^n$:

Doubling is necessary (doubling property is bi-Lip invariant).

Doubling is not sufficient: Shown by Semmes, 1996; simpler example by Laakso, 2002.

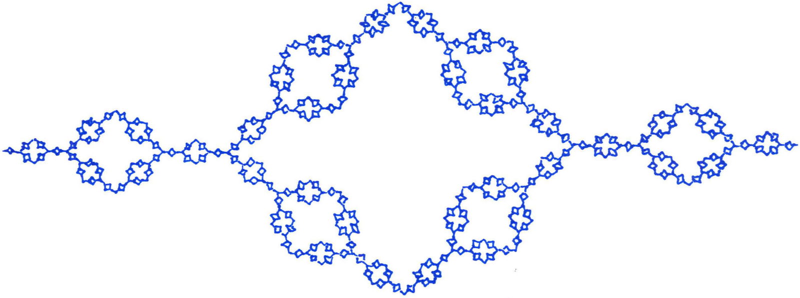
Metric Space Embeddings $f : X \to \mathbb{R}^n$

Isometric embeddings into $\mathbb{R}^n$: too much to hope for.

bi-Lipschitz embeddings into $\mathbb{R}^n$:

Doubling is necessary (doubling property is bi-Lip invariant).

Doubling is not sufficient: Shown by Semmes, 1996; simpler example by Laakso, 2002.



Snowflaking a Metric Space

Given a metric space (X, d) and $\alpha \in (0, 1]$, set

$$d^\alpha(x, y) := (d(x, y))^\alpha.$$

(X, d^α) is a metric space, called the α-snowflake of (X, d).

Snowflaking a Metric Space

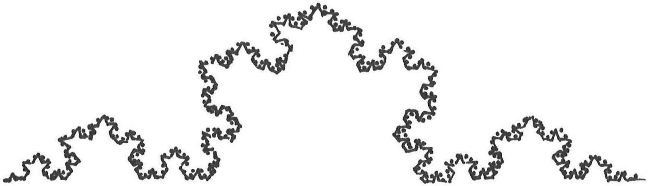
Given a metric space (X, d) and $\alpha \in (0, 1]$, set

$$d^\alpha(x, y) := (d(x, y))^\alpha.$$

(X, d^α) is a metric space, called the α-snowflake of (X, d).

Why do we call it snowflaking?

$$[0, 1]^\alpha \quad \overset{bi-Lip}{\hookrightarrow} \quad \mathbb{R}^2, \;\; \alpha = \log 3 / \log 4$$



Assouad's Theorem

Theorem (Assouad, 1983)

Each snowflaked version of a doubling metric space admits a bi-Lipschitz embedding in some Euclidean space. In particular, the distortion L of the embedding and dimension N of the target space each depend on both the snowflaking constant and on the doubling constant.

Assouad's Theorem

Theorem (Assouad, 1983)

Each snowflaked version of a doubling metric space admits a bi-Lipschitz embedding in some Euclidean space. In particular, the distortion L of the embedding and dimension N of the target space each depend on both the snowflaking constant and on the doubling constant.

Theorem (Naor-Neiman, 2012)

The dimension N can be chosen independent of the snowflaking constant!

Assouad's Theorem

Theorem (Assouad, 1983)

Each snowflaked version of a doubling metric space admits a bi-Lipschitz embedding in some Euclidean space. In particular, the distortion L of the embedding and dimension N of the target space each depend on both the snowflaking constant and on the doubling constant.

Theorem (Naor-Neiman, 2012)

The dimension N can be chosen independent of the snowflaking constant!

Key ingredients in the proof: random embeddings at different scales and a version of the "Lovász local lemma."

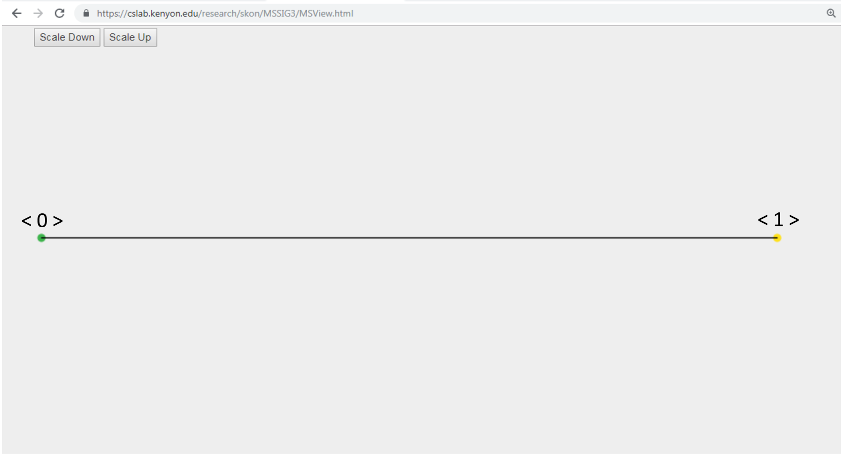
An improvement to Assouad's Theorem

Non-Probabilistic Proof (David-Snipes, 2013).
Big picture idea of the construction:

- Choose a sequence of scales r_k (powers of a small parameter τ).
- For each scale choose a maximal r_k-separated set of "grid points" in the metric space.
- Color the grid points at every level.
- Define the embedding based on the colorings of all the grid points.
- Scales $\leftrightarrow$ digits, and colors $\leftrightarrow$ coordinate directions (coordinate subspaces) of Euclidean space.

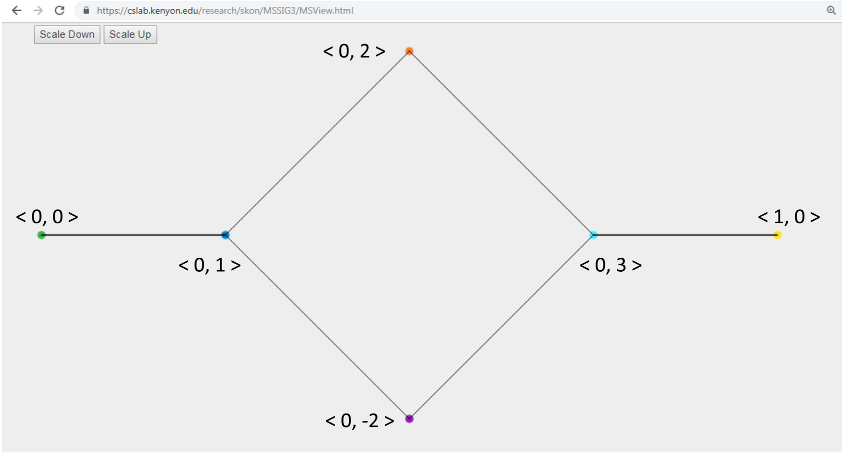
Embedding the Snowflaked Laakso Space

Assign an address (signature) to each point:



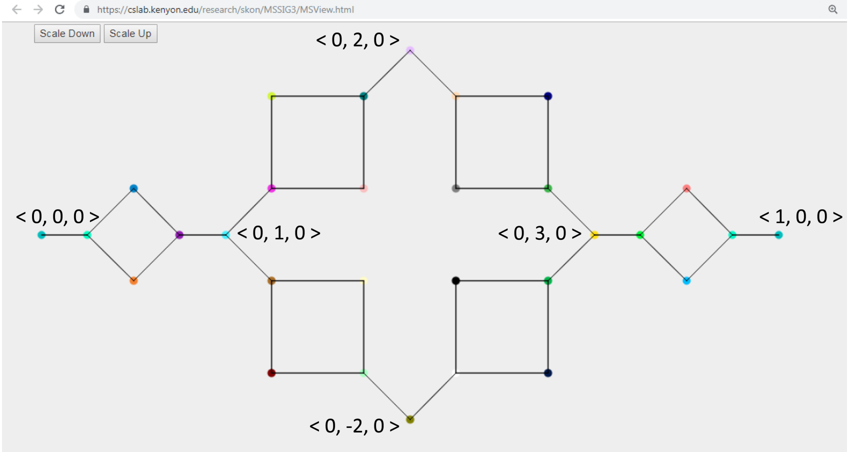
Embedding the Snowflaked Laakso Space

Assign an address (signature) to each point:



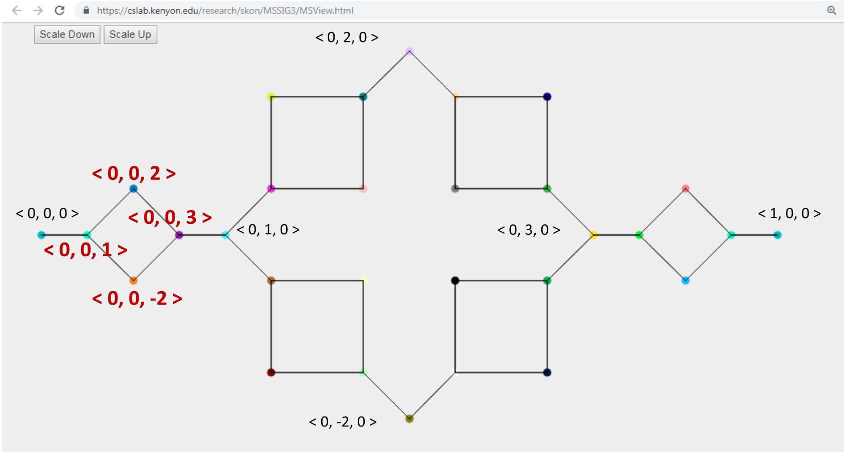
Embedding the Snowflaked Laakso Space

Assign an address (signature) to each point:



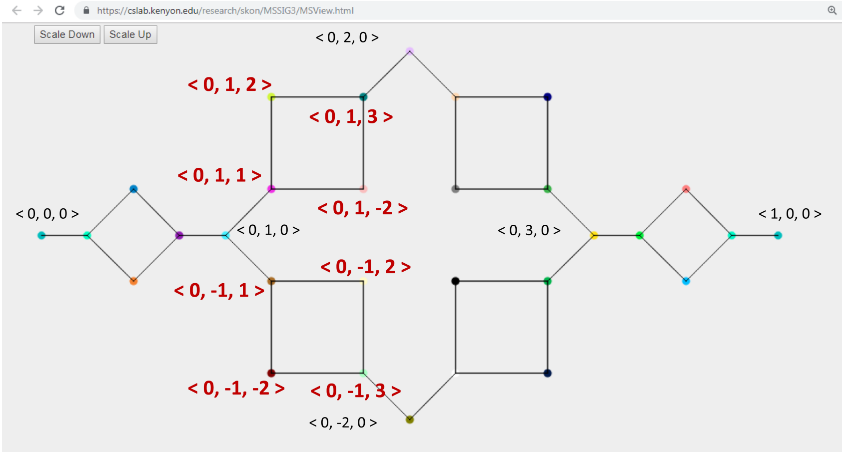
Embedding the Snowflaked Laakso Space

Assign an address (signature) to each point:



Embedding the Snowflaked Laakso Space

Assign an address (signature) to each point:



Embedding the Snowflaked Laakso Space

- Choose constants
  - Snowflaking constant $\alpha > 2/3$: Set $\alpha = \log 3/ \log 4$.
  - Small parameter $\tau < 1 - \alpha$ that gives a sequence of scales: Set $\tau = 1/64$; then scales are $r_k = \tau^{2k}$.
- For each scale r_k, choose a maximal r_k-separated set of "grid points" in the metric space.
  Since $\tau = 1/4^3$, the kth set of grid points is just the $6k$-th stage in the construction of the space.
- Color the grid points at every level. No two points within $10r_k$ of each other can share the same color.

Coloring the r_k-separated sets

Greedy algorithm:

- Enumerate the set of colors
- Enumerate the set of grid points
- Each grid point gets smallest possible color

A priori, number of colors needed is large: $C^5 = 6^5 = 7776$.

Coloring the r_k-separated sets

Greedy algorithm:

- Enumerate the set of colors
- Enumerate the set of grid points
- Each grid point gets smallest possible color

A priori, number of colors needed is large: $C^5 = 6^5 = 7776$.

We implemented this coloring algorithm and found that the maximum number of colors needed is just 31.

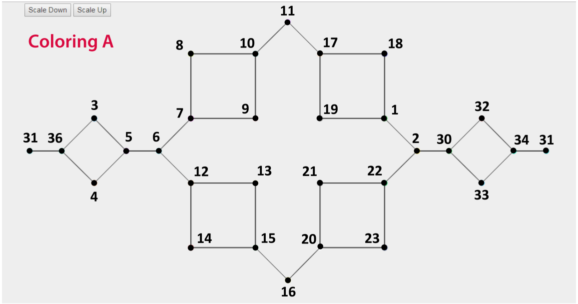
Problem: Greedy is expensive!

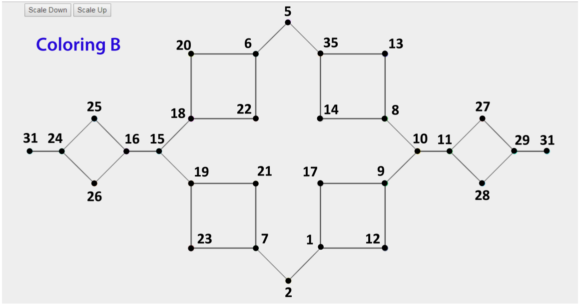
Coloring the r_k-separated sets

A smarter algorithm (Pennington, 2018): Based on stage 3

Total of 36 colors.

The two colorings can be appended to themselves or each other without violating proximity rules.

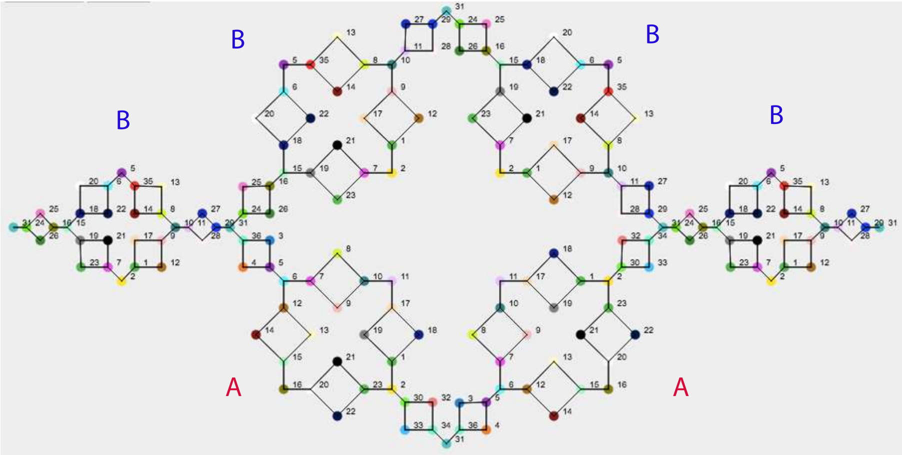




Coloring the r_k-separated sets

A smarter algorithm (Pennington, 2018): Combine 2 colorings for Stage 3 to color any higher level.

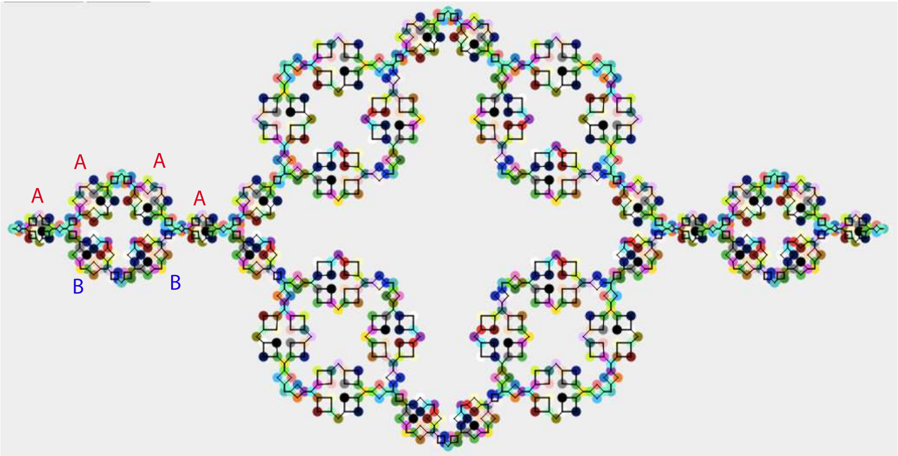
Level 4:



Coloring the r_k-separated sets

A smarter algorithm (Pennington, 2018): Combine 2 colorings for Stage 3 to color any higher level.

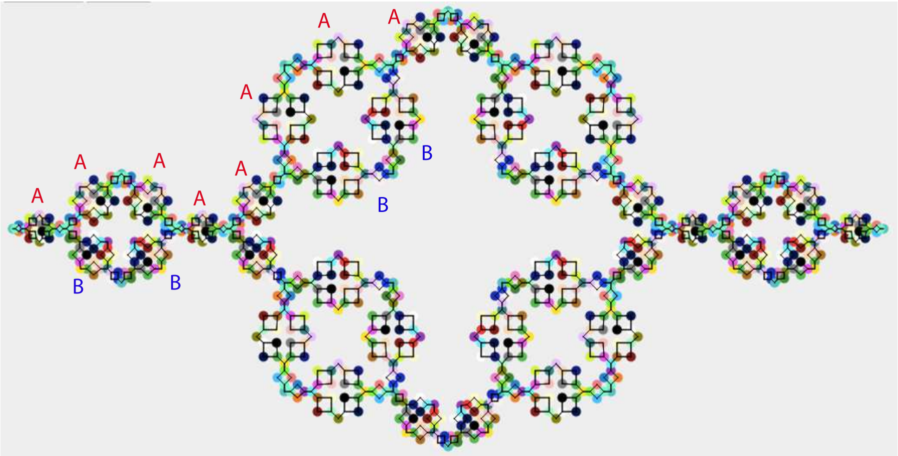
Level 5:



Coloring the r_k-separated sets

A smarter algorithm (Pennington, 2018): Combine 2 colorings for Stage 3 to color any higher level.

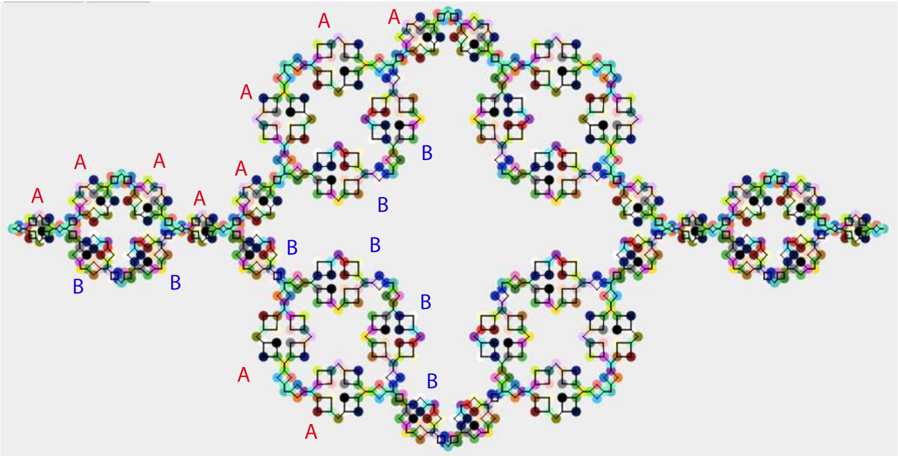
Level 5:



Coloring the r_k-separated sets

A smarter algorithm (Pennington, 2018): Combine 2 colorings for Stage 3 to color any higher level.

Level 5:



Coloring the r_k-separated sets

Run-time comparison for coloring

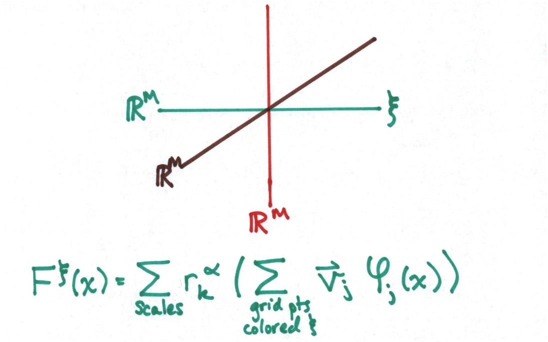
Level	Greedy Algorithm	New Algorithm
4	360 ms	340 ms
6	2 sec	410 ms
8	26 min	5 sec
10	n/a	6 min

https://cslab.kenyon.edu/research/skon/metricspace1.0/MSView.html

Embedding the Snowflaked Laakso Space

- Choose a sequence of scales r_k (powers of a small parameter τ).
- For each scale choose a maximal r_k-separated set of "grid points" in the metric space.
- Color the grid points at every level.
- Define the embedding based on the colorings of all the grid points.
- Scales $\leftrightarrow$ digits, and colors $\leftrightarrow$ coordinate directions (coordinate subspaces) of Euclidean space.

Defining the embedding



Fix a color ξ.

Assign each ξ-colored grid point a vector v_J in the ball of radius τ^2 in $\mathbb{R}^M$.

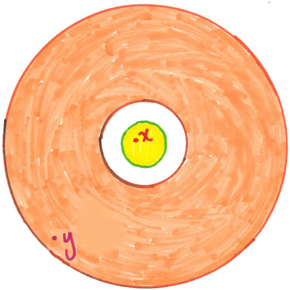
The function $F^\xi : X \to \mathbb{R}^m$ is a double weighted sum (weighted by level and proximity to grid points).

Choosing the vector v_J

Choose v_J successively. For each choice, consider

- the weighted partial sum of previously chosen values in the annulus
- the weighted partial sum of previously chosen values in the ball together with the new v_J

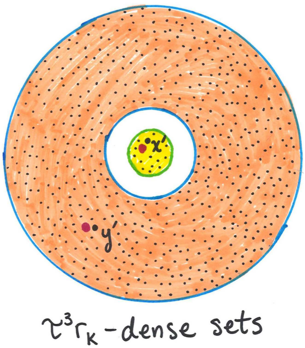
The difference between these, for all choices of pairs of points, should be large.

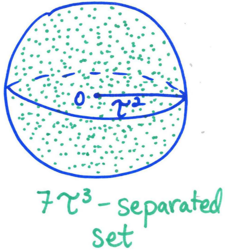


Choosing the vector $\vec{v}_J$

Discretize. For fixed x', y', at most one of the discrete vectors in the sphere doesn't work as a choice of v_J.

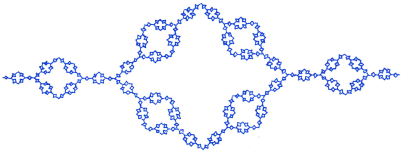
For small τ and large M, there are more discrete vectors in the sphere than pairs x', y' so one of the vectors in the sphere works for all x' in the ball and y' in the annulus.





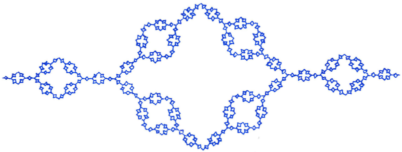
Bounding the dimension of the target space

Given our $\tau = 1/64$, we find a sufficient dimension M by determining the number of $(\tau^3 r_k)$-dense points in the ball B_J and annulus $10B_J \setminus 2B_J$.



Bounding the dimension of the target space

Given our $\tau = 1/64$, we find a sufficient dimension M by determining the number of $(\tau^3 r_k)$-dense points in the ball B_J and annulus $10B_J \setminus 2B_J$.



The discrete points can be taken to be the grid points 9 stages deeper than the r_k level.

$$r_k = (1/4)^{6k} = 4^9(1/4)^{6k+9}$$

Hence, $\#$ pts in B_J is

$$3(\#\text{ pts in level } 9) = 3\left(2 + 4\sum_{n=0}^{8} 6^n\right) = 6,046,623.$$

Bounding the dimension of the target space

Given our $\tau = 1/64$, we find a sufficient dimension M by determining the number of $(\tau^3 r_k)$-dense points in the ball B_J and annulus $10B_J \setminus 2B_J$.

Points in ball: 6×10^6
Points in annulus: 8×10^{12}

Bounding the dimension of the target space

Given our $\tau = 1/64$, we find a sufficient dimension M by determining the number of $(\tau^3 r_k)$-dense points in the ball B_J and annulus $10B_J \setminus 2B_J$.

Points in ball: 6×10^6
Points in annulus: 8×10^{12}

Note that $\tau^2/7\tau^3 = 64/7$. Hence we need dimension of $\mathbb{R}^M$ so that there are 4.8×10^{19} lattice points in the ball of radius 9.

Bounding the dimension of the target space

Given our $\tau = 1/64$, we find a sufficient dimension M by determining the number of $(\tau^3 r_k)$-dense points in the ball B_J and annulus $10B_J \setminus 2B_J$.

Points in ball: 6×10^6
Points in annulus: 8×10^{12}

Note that $\tau^2/7\tau^3 = 64/7$. Hence we need dimension of $\mathbb{R}^M$ so that there are 4.8×10^{19} lattice points in the ball of radius 9.

Use combinatorial formula for lattice points in ℓ_1 balls: $M = 319$.

Conservative estimate of the final dimension is $2 * 36 * 319 = 22968$.

(Compare to a priori estimate of $2 * 6^5 * 2423174 \approx 3.8 \times 10^{10}$)

In Summary

Results so far

- Assignment of addresses (signatures) to each point in the Laakso space.
- Can determine distance between two points just from signatures.
- A non-greedy coloring algorithm.
- Dimension estimate of the target Euclidean space.

In Summary

Results so far

- Assignment of addresses (signatures) to each point in the Laakso space.
- Can determine distance between two points just from signatures.
- A non-greedy coloring algorithm.
- Dimension estimate of the target Euclidean space.

What's left to do?

- Find the vectors v_j for $k = 1, 2$. (Very computationally expensive. Hope to leverage symmetry.)
- Calculate embedded coordinates for $k = 2$.
- Investigate distortion of this embedding. Conservatively,

$$2^{-33} d(x, y)^{\alpha} \leq |F(x) - F(y)| \leq 2023 d(x, y)^{\alpha}$$

- Create/investigate visualizations using projections.

In Summary

Results so far

- Assignment of addresses (signatures) to each point in the Laakso space.
- Can determine distance between two points just from signatures.
- A non-greedy coloring algorithm.
- Dimension estimate of the target Euclidean space.

What's left to do?

- Find the vectors v_j for $k = 1, 2$. (Very computationally expensive. Hope to leverage symmetry.)
- Calculate embedded coordinates for $k = 2$.
- Investigate distortion of this embedding. Conservatively,

$$2^{-33} d(x, y)^{\alpha} \leq |F(x) - F(y)| \leq 2023 d(x, y)^{\alpha}$$

- Create/investigate visualizations using projections.
- Vary snowflaking constant and compare embeddings.
- Generalize methods/calculations to other fractals.