p-Adic scaled space filling curve indices

Patrick Erik Bradley

Institute of Photogrammetry and Remote Sensing (IPF)
Geodetic Institute (GIK)
@ Karlsruhe Institute of Technology (KIT)

p-adics.2021

Mexico City, online
May 17 – May 28, 2021

Section

Bioinformatics, Data Analytics and Applications

Preliminaries

- joint work with Markus Jahn in geo-informatics
- Most of this presentation published in

  P.E. Bradley and M.W. Jahn. *On the Behaviour of p-Adic Scaled Space Filling Curve Indices for High-Dimensional Data*. The Computer Journal, 2020. doi: 10.1093/comjnl/bxaa036 (no volume and page numbers yet...)

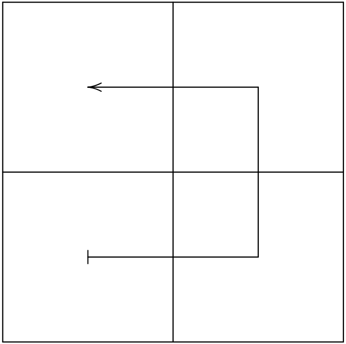
Contents

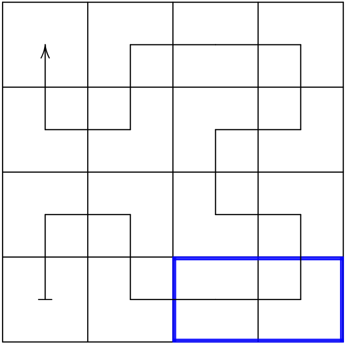
1. Introduction

- 1877. Cantor proved $|\mathbb{R}| = |\mathbb{R}^n|$, $n \geq 1$
- 1890. Peano finds continuous bijection $[0,1] \to [0,1]^2$
  - Method: iterative trisections.
  - $\rightsquigarrow$ 3-adic space-filling curve
- 1891. Hilbert obtains a binary construction of a SFC.
  - 2-adic SFC
- 20th century. Higher-dimensional realisations of Hilbert's SFC via *Gray Code*.
- Question: How many are there? $\rightsquigarrow$ Lots!

2. Hilbert Curves

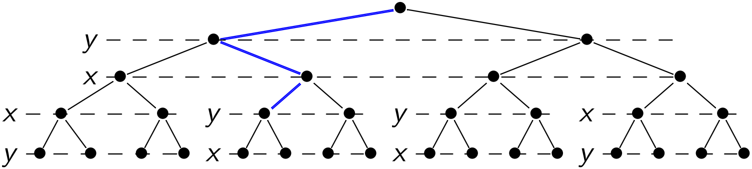
First two iterations





2. Hilbert Curves

Hilbert Tree



2. Hilbert Curves

Locality Property. Points nearby on the curve are nearby in $[0, 1]^2$.

Not vice versa!

2. Hilbert Curves

p-adic discrete curve. A bijective map

$$C\colon \{0, \ldots, p^r - 1\} \to \left(\mathbb{Z}/p^k\mathbb{Z}\right)^r$$

is a *p-adic discrete curve* of *order* k and *dimension* r, if

$$d_H(C(i), C(i+1)) = 1$$

for $i = 0, \ldots, p^r - 2$, and d_H is the Hamming distance.

$$\mathcal{C}(k, r) = \{p\text{-adic curves of order } k \text{ and dimension } r\}$$

2. Hilbert Curves

Let

$$C \in \mathcal{C}_m(1, r)$$

Then

$$\mathrm{Aut}_C \left(\mathbb{Z}/m\mathbb{Z}\right)^r = \{\sigma \in \mathrm{Sym}\left(\mathbb{Z}/m\mathbb{Z}\right)^r : \sigma \circ C \in \mathcal{C}_m(1, r)\}$$

is the *C-automorphism group*.

2. Hilbert Curves

Refinement.

$$\varpi^k\colon (\mathbb{Z}/p\mathbb{Z})^r \to \left(\mathbb{Z}/p^k\mathbb{Z}\right)^r$$
$$(x_1, \ldots, x_r) \mapsto (x_1 p^{k-1} \mod p^k, \ldots, x_r p^{k-1} \mod p^k)$$

Maps to lower left corner in finer curve.

2. Hilbert Curves

Projection.

$$\pi^k \colon \left(\mathbb{Z}/p^k\mathbb{Z}\right)^r \to (\mathbb{Z}/p\mathbb{Z})^r$$
$$(x_1, \ldots, x_r) \mapsto (x_1 \mod p, \ldots, x_r \mod p)$$

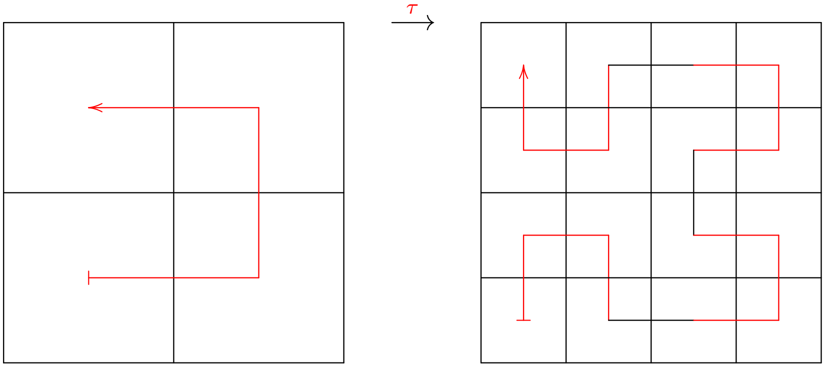
2. Hilbert Curves

- Fix $C \in \mathcal{C}_p(1, r)$.
- Let
$$\tau_0, \ldots, \tau_{p^r - 1} \in \mathrm{Aut}_C \left(\mathbb{Z}/p\mathbb{Z}\right)^r$$
- Construct iteratively: $C_0 = C, \ \ldots, \ C_{k-1} \in \mathcal{C}_p(k-1, r)$.
-
$$\begin{aligned} C_k \colon \left\{0, \ldots, p^{kr} - 1\right\} &\to \left(\mathbb{Z}/p^k\mathbb{Z}\right)^r \\ i &\mapsto \tau_{i'}\left(\pi^{k-1}\left(C_{k-1}\left(i \mod \left(p^{k-1}\right)^r\right)\right)\right) + \pi^k\left(C\left(i'\right)\right) \\ i' &= i \ \mathrm{div} \ \left(p^{k-1}\right)^r \in \{0, \ldots, p^r - 1\} \end{aligned}$$
- $C_k \in \mathcal{C}_p(k, r)$

Say: (C_k) is a *Hilbert family* generated by C and (τ_k).

2. Hilbert Curves

Glueing copies with automorphisms:



τ is a combination of reflections and rotations

2. Hilbert Curves

Lemma. Let p be odd. Then any $C \in \mathcal{C}_p(1, r)$ is the generating curve of a Hilbert family.

Idea of proof.

Any two opposite corners of the hypercube can be connected by a discrete p-adic curve, if p is odd. This leads to a refinement of a curve with a suitable automorphism. □

2. Hilbert Curves

Lemma. Let p be even. Then a discrete curve $C \in \mathcal{C}_p(1, r)$ beginning and ending in a corner satisfies

$$d_H(C(0), C(p^r - 1)) < r$$

I.e. start and endpoints are not opposite corners of the hypercube.

2. Hilbert Curves

Open Problem. How many different p-adic Hilbert families are there modulo automorphisms of the r-hypercube (i.e. reflections and rotations)?

3. p-adic Gray Code

2-**adic reflected Gray code.**

$$Z_2^n \to Z_2^n,\ x \mapsto x + (x \triangleright 1)$$

$x \triangleright k$: right shift by k bits

E.g. 3-digit binary numbers in Gray code ordering:

000	001	011	010	110	111	101	100
0	1	3	2	6	7	5	4

start: $(0, \ldots, 0)$ end: $(1, 0, \ldots, 0)$
$\rightsquigarrow$ common edge in hypercube

3. p-adic Gray Code

p-adic reflected Gray code with $p > 2$ odd.

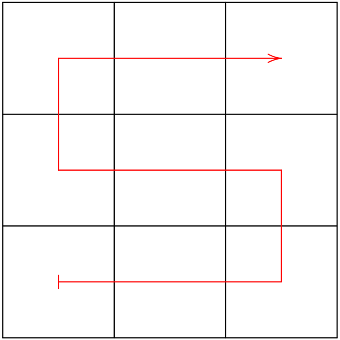
$$G(1,p) = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ p-1 \end{pmatrix}$$

$$G(n,p) = \begin{pmatrix} 0 \times G(n-1,p) \\ 1 \times G(n-1,p)^R \\ \vdots \\ p-2 \times G(n-1,p)^R \\ p-1 \times G(n-1,p) \end{pmatrix}$$

$(\cdots)^R$ means in reverse order!

p-adic Gray Code

Example. $G(2,3)$.



In General. Start and end points are opposite corners!

3. p-adic Gray Code

p-adic reflected Gray code with $p > 2$ even.

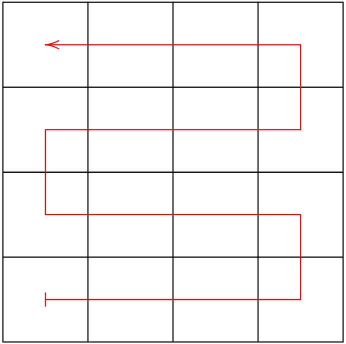
$$G(1,p) = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ p-1 \end{pmatrix}$$

$$G(n,p) = \begin{pmatrix} 0 \times G(n-1,p) \\ 1 \times G(n-1,p)^R \\ \vdots \\ p-2 \times G(n-1,p) \\ p-1 \times G(n-1,p)^R \end{pmatrix}$$

$(\cdots)^R$ means in reverse order!

3. p-adic Gray Code

Example. $G(2, 4)$.



In General. Start and end points are corners with common edge!

4. p-adic Gray Hilbert Curves

Apply transformations of the type

$$T\colon Z_p^n \to Z_p^n,\ x \mapsto x^\tau + e$$

- coordinate permutation $\tau \in S_n$
- translation with $e \in Z_p^n$

to construct *affine p-adic Gray Hilbert Curves*.

4. *p*-adic Gray Hilbert Curves

Construction.

- Let a, b be sub-hypercubes of $[0, 1]^n$ w.r.t. to a p-adic subdivision
- Assume a, b are consecutive in given p-adic reflected Gray Code.

4. p-adic Gray Hilbert Curves

Case: p odd.

- Choose an ordering of coordinates for a.
- f_a: endpoint of $G(n, p)$ w.r.t. a.
- f_a is opposite to start point $\Rightarrow$ f_a neighbours b

4. p-adic Gray Hilbert Curves

Case: p even.

- f_a: corner of a neighbouring b s.t.
  f_a different with origin of a in only one coordinate
- Choose ordering of remaining coordinates.
- Corresponding Gray code ends in f_a.

4. p-adic Gray Hilbert Curves

Both cases.

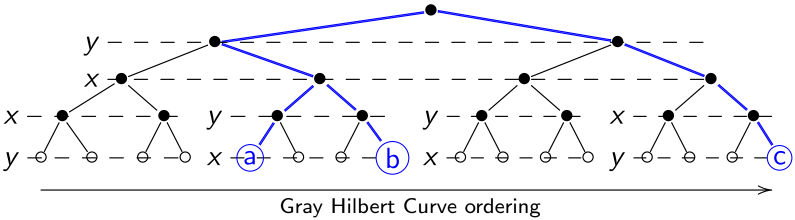
- e_b: corner in $a \cup b$ neighbouring f_a
- Bring origin of Z_p^n to e_b via translation
- Continue the whole process with b, c
- c: next hypercube in Gray code order for b.

Result. Affine p-adic Gray Hilbert curve.

Remark. Number of possibilities exponential in n.

5. p-adic Scaled Hilbert Indexing Methods

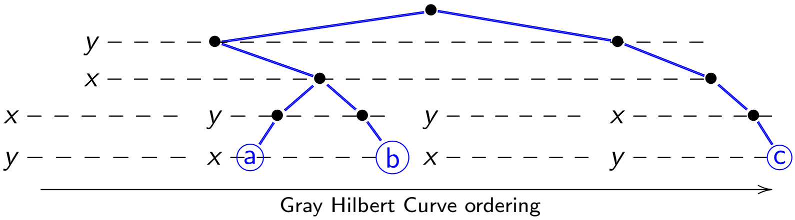
Traditional (Static).



- High depth = lots of empty leaf nodes

5. p-adic Scaled Hilbert Indexing Methods

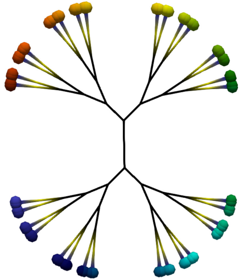
New (dynamic).



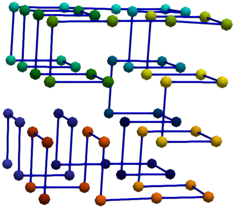
- Take smallest sub-tree containing leaf nodes with data.
- Order those leaf nodes along Gray Hilbert curve.
- $\rightsquigarrow$ most efficient index sub-structure (after removing non-fork nodes)
- efficient insert/delete/retrive methods

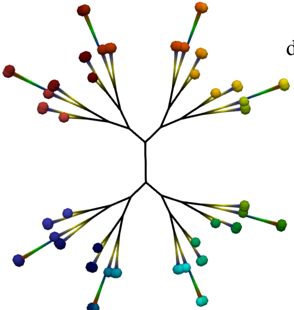
6. Locality Properties

$p = 2$, normally distributed, “bubble”

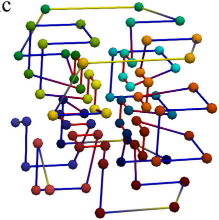


static



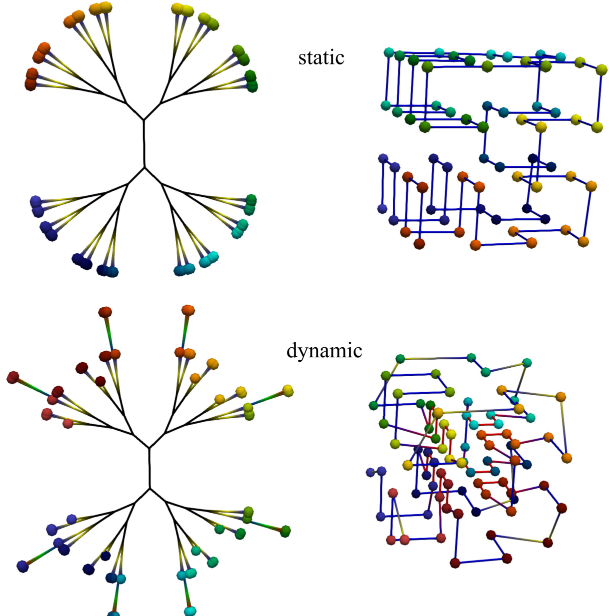


dynamic



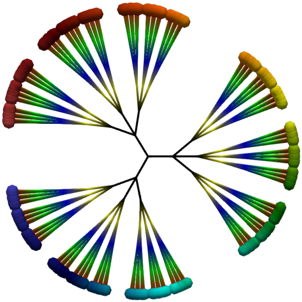
6. Locality Properties

$p = 2$, normally distributed, "ring"

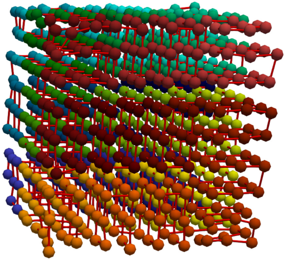


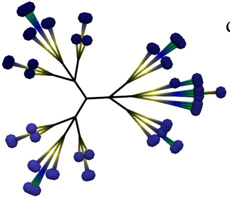
6. Locality Properties

$p = 3$, normally distributed, “bubble”

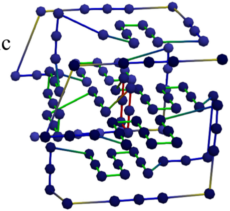


static



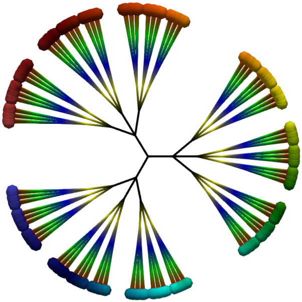


dynamic

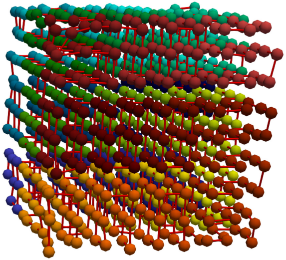


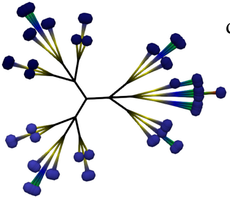
6. Locality Properties

$p = 3$, normally distributed, "ring"

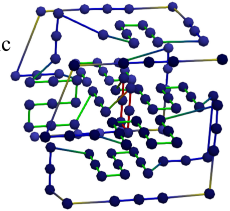


static



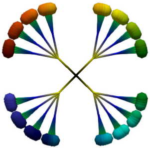


dynamic

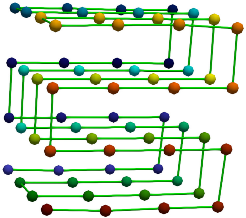


6. Locality Properties

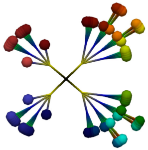
$p = 4$, normally distributed, "bubble"

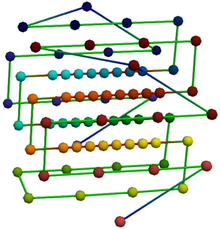


static



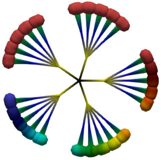
dynamic



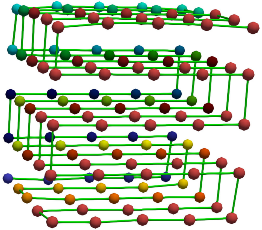


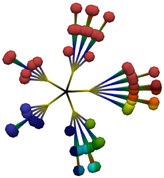
6. Locality Properties

$p = 5$, normally distributed, “bubble”

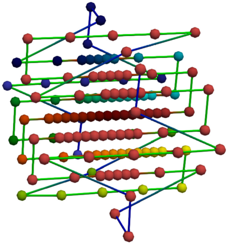


static





dynamic



6. Locality Properties

- $S \subset [0,1]^n$ (finite) point cloud
- nodes of GH tree: *buckets*
- buckets are: *empty*, *filled*, *underfilled*, or *overfilled* w.r.t. *bucket capacity* $s > 0$
- $\omega(T,S) = \frac{\#\text{ overfilled leaf nodes of } T}{\#\text{ non-empty leaf nodes of } T}$
- $L(T) = \{\text{leaf nodes of } T\}$

$$\Omega(T,S) = (1 + \omega(T,S)) \cdot |L(T)|$$

capacity of T for data S

6. Locality Properties

Capacity ratio.

$$R(T_1, T_2; S) = \frac{\Omega(T_1, S)}{\Omega(T_2, S)}$$

- T_{static}: static GH tree with maximal iteration number

$$k = \left\lceil \frac{\log_p \frac{|S|}{s}}{n} \right\rceil$$

- T_{scaled}: GH tree for S

6. Locality Properties

Motivation. Want to have

$$p^{nk} \approx \frac{|S|}{s}$$

as number of leaf nodes in T_{static}.

$\rightsquigarrow$ on average no leaf node overfilled.

6. Locality Properties

$$k = \frac{\log_p \frac{|S|}{s}}{n} + \epsilon_p$$

with $\epsilon_p \in [0, 1)$.

Theorem. If $\frac{\log_p s}{n} < \epsilon_p$, then T_{scaled} is more storage efficient than T_{static}, and relative efficiency increases with increasing dimension n. For fixed n, this is the case, if bucket capacity s is sufficiently small.

Idea of Proof.

- $R_p(S) = R(T_{scaled}, T_{static}; S) \to 0$ for $n \to \infty$
- $\epsilon_p \to 1$ for $n \to \infty$
- The bound leads to $R_p(S) \leq 1$.

□

6. Locality Properties

p-**adic GH local sparsity measure.**

$$\rho_p(S, s) = \frac{-\log_p(R_p(S)/s) - n\epsilon_p}{\log_p(2s)} \in [0, 1]$$

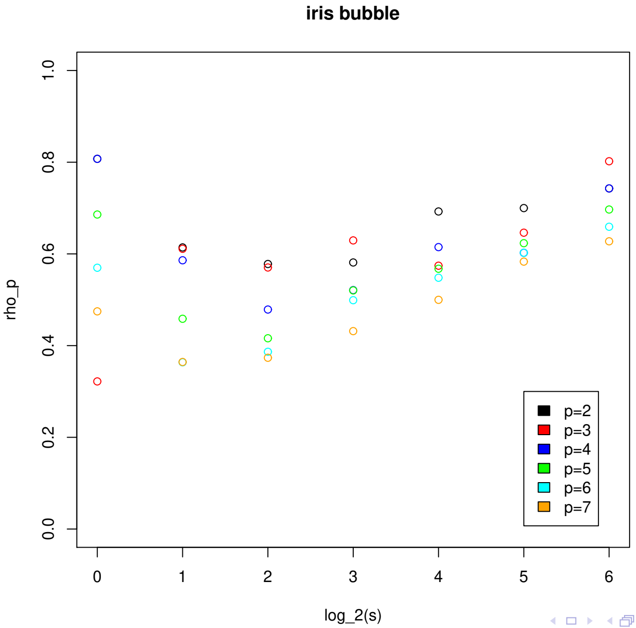
We expect this quantity to depend on distribution of S only, if s is fixed.

7. Experiments

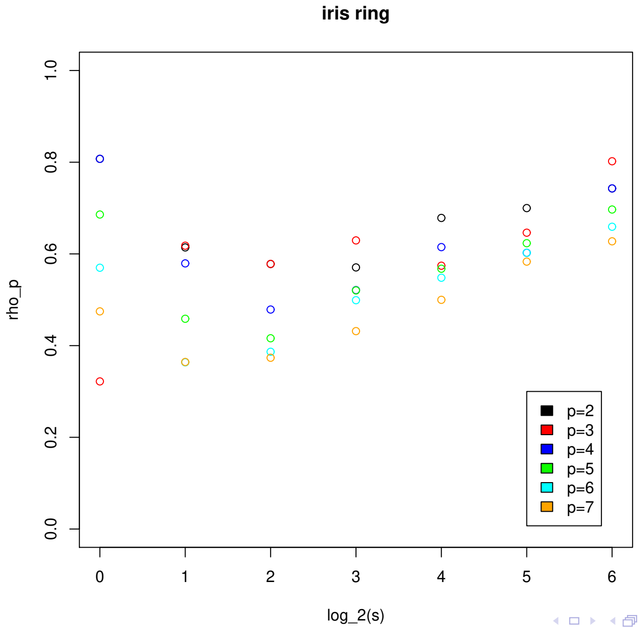
Iris data-set.

- original data embedded in four dimensions
- complete disjunctive form (cdf): embedding dimension is 400
- BCH 127 50 27 encoding: embedding dimension is 431

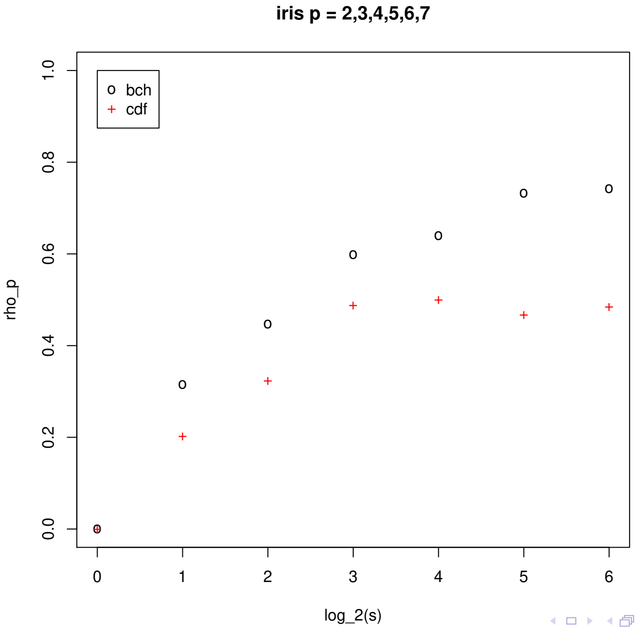
7. Experiments



7. Experiments



7. Experiments



7. Experiments

- High embedding dimension $\Rightarrow$ local sparseness
- $\rho_p(S, s)$ independent of p in this case
- Murtagh had observed high ultrametricity in this case!
- Relationship: local sparsity and ultrametricity?

8. Possible Application

p-adic pde on a dag in Euclidean space.

- Γ: dag with vertices in $[0,1]^n$
- E: edge relation of Γ
- T: p-adic GH Curve Index for vertices of Γ
- Leaf nodes of T define p-adic discs in $\mathbb{Q}_p^n$
- L: p-adic graph Laplace operator made from E à la Wilson Zúñiga-Galindo
- study dynamics on Γ in parallel using T!

8. Possible Application

Example for such a dag.

- convex polytope complex in $\mathbb{R}^n$
- dag vertices: the centres of the polytopes
- edges: boundary relationships (weighted with polytope volumes)
- *Example scenario:* city models in not too large detail
- *Example pde:* heat flow through buildings and their parts

8. Possible Application

***p*-adic graph Laplace operator.**

Γ: finite graph with vertex set $V(\Gamma)$.

- Fix an embedding $V(\Gamma) \to \mathbb{Q}_p$.
- Fix N s.t. the balls

$$B_{-N}(I) = \left\{ x \in \mathbb{Q}_p \colon \ |x - I|_p \leq p^{-N} \right\}$$

  with $I \in V(\Gamma)$ are disjoint.
-

$$K_N := \bigcup_{I \in V(\Gamma)} B_{-N}(I) \subset \mathbb{Q}_p$$

8. Possible Application

p-adic graph Laplace operator.

- $n = |V(\Gamma)|$
- $A = (A_{IJ}) \in \mathbb{C}^{n \times n}$ corresponds to function:

$$A \colon K_N \times K_N \to \mathbb{C},\ (x, y) \mapsto A(x, y)$$

$$A(x, y) = p^N \sum_{I \in V(\Gamma)} \sum_{J \in V(\Gamma)} A_{IJ} \Omega \left(p^N |x - I|_p \right) \Omega \left(p^N |y - J|_p \right)$$

- Graph Laplacian $L = A - (\gamma_I \delta_{IJ})$ corresponds to

$$\mathcal{L} \colon L^2(K_N, \mathbb{C}) \to L^2(K_N, \mathbb{C})$$
$$u(x) \mapsto \mathcal{L}u(x) = \int_{K_N} A(x, y)(u(y) - u(x))\, dy$$

Zúñiga-Galindo's *p-adic graph Laplace operator*.

9. Conclusions

- theoretical approach to p-adic index structures
- Gray Code + Hilbert Curve + any dimension + any p
- $\rightsquigarrow$ tree + linear ordering of nodes with nice locality
- plethora of index structures
- measure for sparsity of a point cloud
- *potential application:* p-adic parallel solver for p-adic pde on dag with vertices in Euclidean space
- *future work:* p-adic approach to flows on city models
- *in review:* p-adic wave equations on finite graphs and T_0-spaces

References

1. P.E. Bradley. *p-Adic Wave Equations on finite Graphs and T_0-spaces*. In review.
2. P.E. Bradley and M.W. Jahn. *On the Behaviour of p-Adic Scaled Space Filling Curve Indices for High-Dimensional Data*. The Computer Journal, 2020. doi: 10.1093/comjnl/bxaa036
3. W.A. Zúñiga-Galindo. *Reaction-diffusion equations on complex networks and Turing patterns, via p-adic analysis*. Journal of Mathematical Analysis and Applications, 491(1):124239, 2020.