

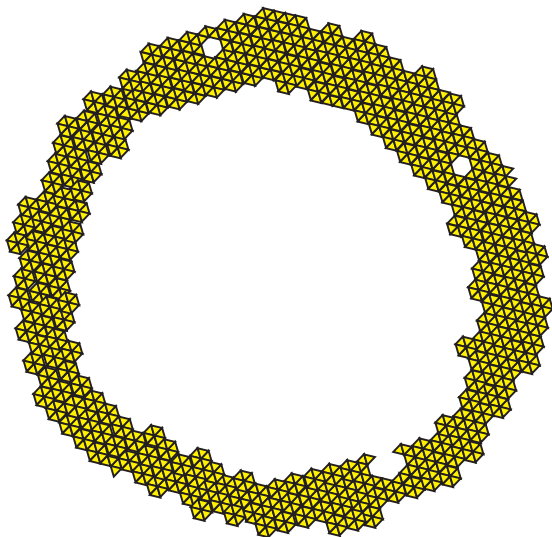
Topological data analysis

Mikael Vejdemo-Johansson

21 January 2011

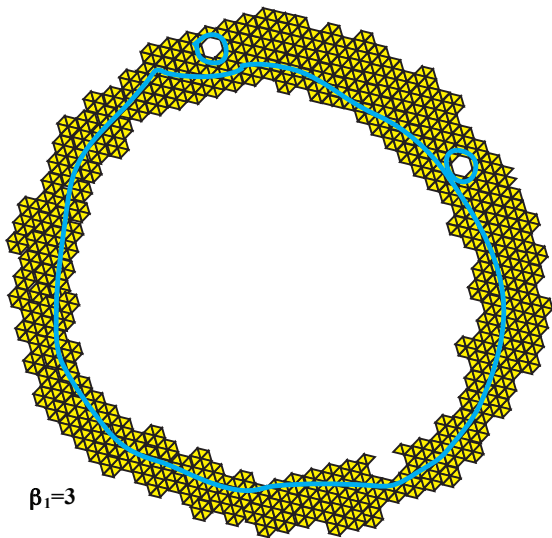
Example: pointcloud of a circle

First idea: pick some nice ϵ to work with $VR_\epsilon(X)$.



Example: pointcloud of a circle

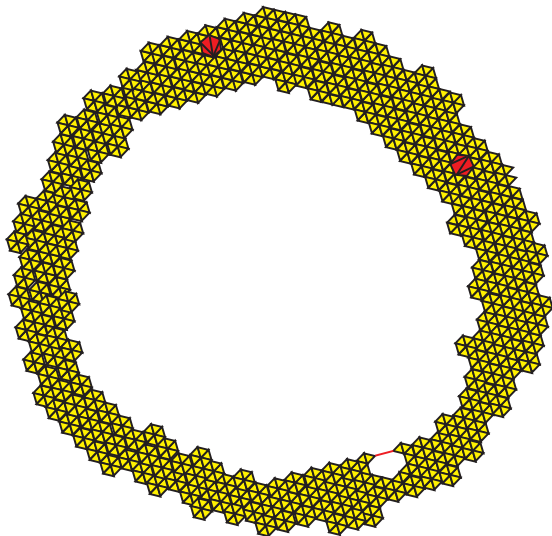
First idea: pick some nice ϵ to work with $VR_\epsilon(X)$.



$\beta_1=3$

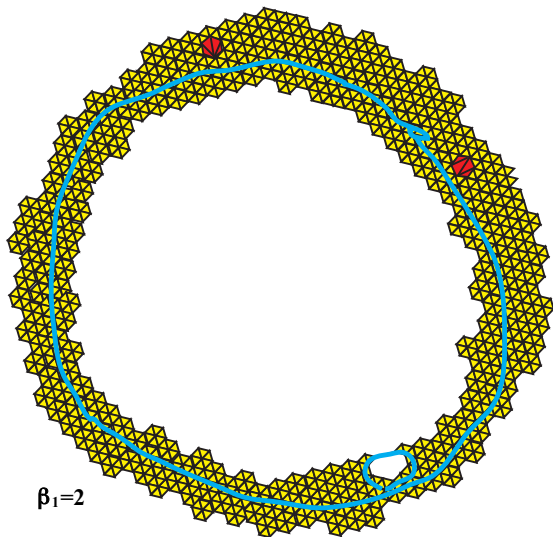
Example: pointcloud of a circle

First idea: pick some nice ϵ to work with $VR_\epsilon(X)$.

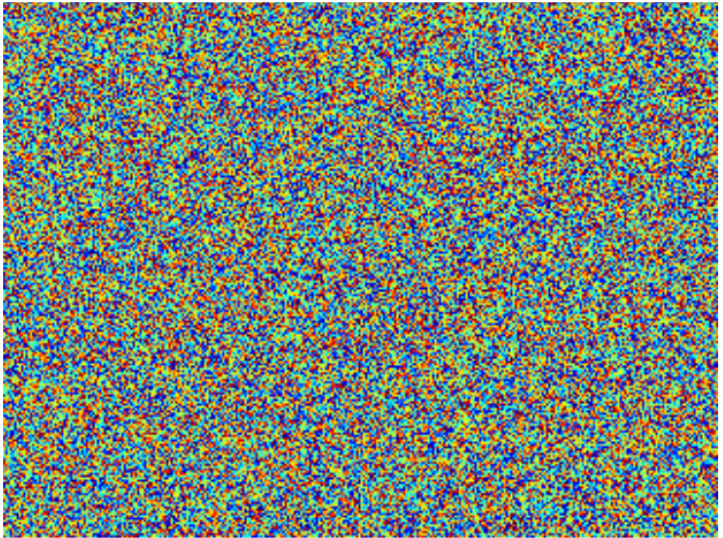


Example: pointcloud of a circle

First idea: pick some nice ϵ to work with $VR_\epsilon(X)$.



$\beta_1=2$





Example: natural images

Lee-Mumford-Pedersen investigated whether a statistically significant difference exists between natural and random images.

Natural images form a “subspace” of all images. Dimension of ambient space e.g. $640 \times 480 = 307\,200$.

This space of natural images should have:

- ▶ high dimension: there are many different images.
- ▶ high codimension: random images look nothing like natural ones.

Natural 3x3 patches

Instead of studying entire images, we consider the distribution of 3×3 pixel patches.

Most of these will be approximately constant in natural images. Allowing these drowns out structure.

Lee-Mumford-Pedersen chose 8 500 000 patches with high contrast from a collection of black-and-white images used in cognition research. Each 3×3 -patch is considered a vector in \mathbb{R}^9 .

Normalised brightness: $\mathbb{R}^9 \rightarrow \mathbb{R}^8$. Normalised contrast: $\mathbb{R}^8 \rightarrow \mathcal{S}^7$.

Subsequent topological analysis by Carlsson–de Silva–Ishkanov–Zomorodian.

Pixel patches in S^7

The resulting patches are dense in S^7 – so we consider high-density regions.

Pick out 25% densest points. We can pick a parametrised method to measure density:

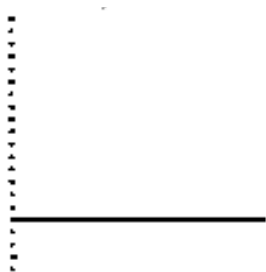
Definition

k -codensity $\delta_k(x)$ of a point x is the distance to its k th nearest neighbour.

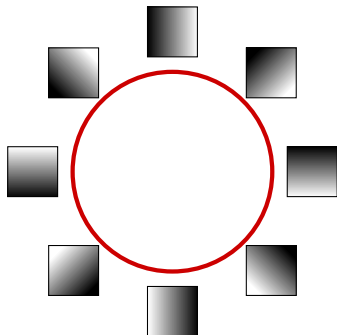
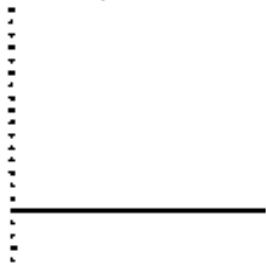
k -density $d_k(x)$ is $1/\delta_k(x)$.

High k yields a smoothly changing density measure capturing global properties. Low k yields a wilder density measure capturing local properties. k acts as a kind of focus control.

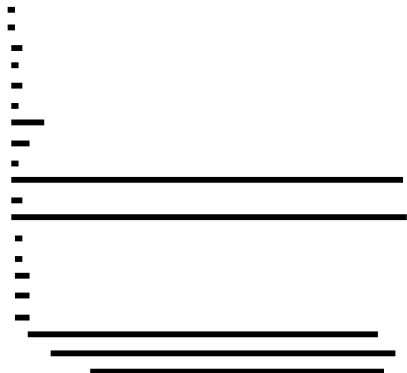
300-density



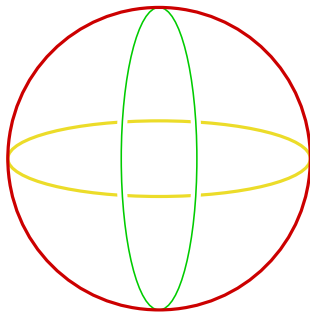
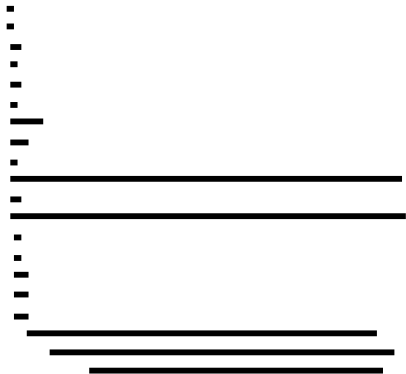
300-density



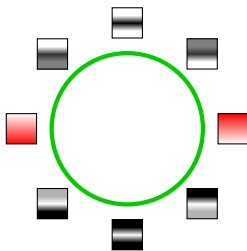
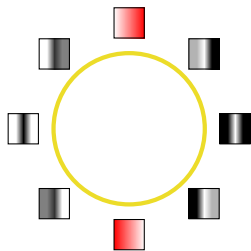
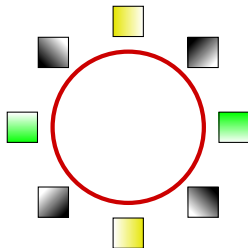
15-density



15-density



Three circles

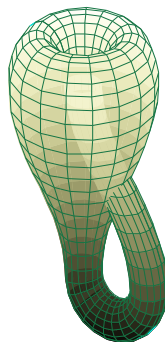
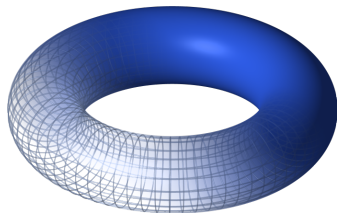


Identifying the subspace of natural pixel patches

Raising the cut-off bar yields, with coefficients in \mathbb{F}_2

$$\beta_0 = 1 \quad \beta_1 = 2 \quad \beta_2 = 1$$

Assuming the shape is a surface, this corresponds to one of

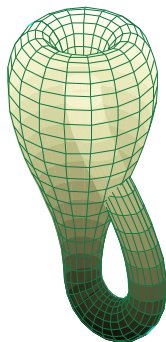


Identifying the subspace of natural pixel patches

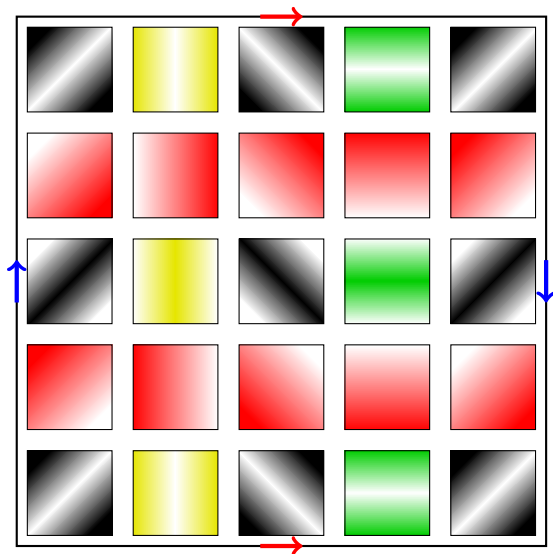
Raising the cut-off bar yields, with coefficients in \mathbb{F}_3

$$\beta_0 = 1 \quad \beta_1 = 1$$

Thus, the relevant shape is:



Klein bottle of pixel patches



Applications of this analysis

Image compression

A 3×3 -cluster may be described using 4 values:

- ▶ Position of its projection onto the Klein bottle
- ▶ Original brightness
- ▶ Original contrast

Texture analysis

Textures yield distributions of occurring patches on the Klein bottle. Rotating the texture corresponds to translating the distribution. [J Perea]

Coordinatization methods

My own work is on automating the above process by finding topological methods to recover intrinsic coordinate maps.

Idea

- ▶ Starting from a dataset X : compute its persistent homology $H(VR_*(X); k)$
- ▶ Guess a simplicial complex Y with corresponding homology
- ▶ Find maps $X \rightarrow Y$ or $VR_*(X) \rightarrow Y$ that lift to the expected correspondance.

First results

Joint with Vin de Silva and Dmitriy Morozov.

We can use that the circle is an Eilenberg-Mac Lane space, and thus

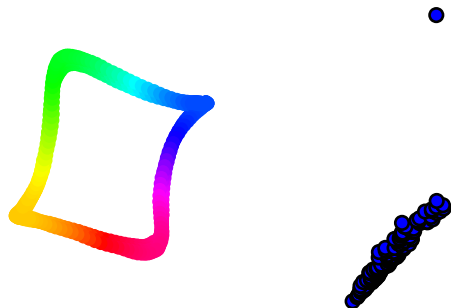
$$H^1(X; \mathbb{Z}) = [X, S^1]$$

We have established a definition of persistent cohomology, and produced techniques, algorithms and software for computing circle-valued coordinate functions using cohomology and a smoothing step.

Future directions

- ▶ Approach more generic coordinatizations by studying optimal chains in $H_0(\text{hom}(CX, CY)) = \bigoplus_p \text{hom}(H_p X, H_p Y)$.
- ▶ Apply the circular coordinates work to periodic and recurrent systems and signals. Currently looking at data sets from: meteorology, climate research, gait research, music.
- ▶ Use circular coordinates for quality control on existing analysis methods for periodic signals.

Questions?



Delay embedding of a window from a clarinet tone, using circular coordinates and a persistence diagram to quality control the delay embedding.