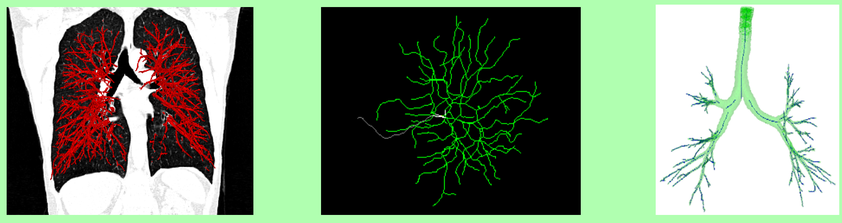
The geometry and statistics of geometric trees

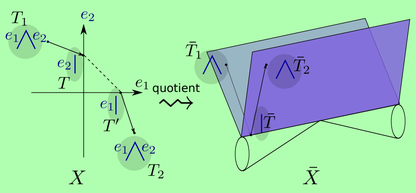
Aasa Feragen

Machine Learning and Computational Biology Research Group,
MPI for Developmental Biology and MPI for Intelligent Systems
aasa@diku.dk, http://image.diku.dk/aasa

Motivation



Tree-shapes are important in image analysis, where they appear as anatomical trees such as vessel networks, dendrite trees in neurons, or airway trees.

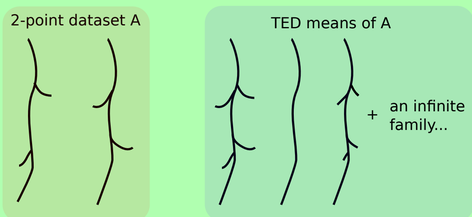


- Tree-shapes live in highly non-linear spaces
- These spaces are not manifolds
- We can still define Frechet means:

$$m(x_1, \ldots, x_n) = argmin_x \sum_{i=1}^n d(x, x_i)^2$$

- But can we compute them?

- Traditionally in computer vision and pattern recogntion, tree- and graph structures have been compared using edit distance.
- In edit distance, shortest paths are almost never unique
- As a result, Frechet means are also not unique.



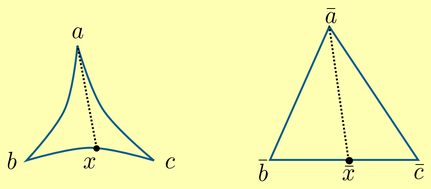
Metric geometry and statistics

Curvature in metric spaces is defined by comparison with planes, spheres and hyperbolic spaces.

Comparison is made via triangles.

CAT(0) spaces are geodesic metric spaces in which geodesic triangles are "thinner" than their planar comprison triangles.

Locally CAT(0) spaces are called non-positively curved metric spaces.



Fact

Frechet means exist and are unique in CAT(0) spaces. Thus, Frechet means exist and are unique for sufficiently dense subsets of restricted treespace.

Computation of shortest paths is generally NP hard

Assumptions on edge order and labels regularize both geometry and computational complexity

	edge order	leaf labels	unordered
L1	polynomial (Klein/ Zhang-Shasha)	? complexity Everywhere unbounded curvature	NP complete (Zhang) Everywhere unbounded curvature
L2	? complexity Locally generically CAT(0)	polynomial (Owen, Provan) (Phylo-space) CAT(0) (Billera, Holmes, Vogtmann)	NP complete Bounded trees: Locally generically CAT(0); else un-bounded curvature (in collaboration with Scott Provan)

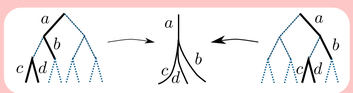
The space of tree-shapes

We study tree-shapes as defined by Feragen et al [1,2,3].

1. A pre-shape space spanned by a tree T = (V, E, r)

$$X = \prod_{e \in E} (\mathbb{R}^d)^n$$

2. An equivalence identi-fying representations of same shape.



3. Form a quotient space modding out the equivalence

$$\bar{X} = X / \sim$$

4. Endow the quotient with the quotient metric induced by the Euclidean metric on X, called the QED metric

If T is infinitely large, the space of tree-shapes has unbounded curvature everywhere.

If T is bounded, the space of tree-shapes is locally CAT(0) almost everywhere.

That is, if the data-point trees are sufficiently close together, they will live in a CAT(0) neighborhood.

The CAT(0) neighborhoods can be made larger by restructing to a particular set of trees, e.g trees with N leaves, trees with particular topologies or trees with fixed leaf label sets.

References

[1] A. Feragen, F. Lauze, P. Lo, M. de Bruijne, M. Nielsen: Geometries in spaces of treelike shapes. ACCV 2010.
[2] A. Feragen, S. Hauberg, M. Nielsen, F. Lauze, Means in spaces of treelike shapes, ICCV 2011.
[3] A. Feragen, P. Lo, M. de Bruijne, M. Nielsen, F. Lauze, Towards a theory of statistical tree-shape analysis, submitted.
[4] A. Feragen, Complexity of computing distances between geometric trees, submitted.

Geometric Graph Kernels: Efficient computation and flexible modeling

A graph kernel on a space of graphs $\mathscr{G}$ is a map

$$k \colon \mathscr{G} \times \mathscr{G} \to \mathbb{R}$$

which satisfies
- symmetry: $k(g_1, g_2) = k(g_2, g_1)$
- positive semi-definite: for any set of graphs $g_1, \ldots g_n$ the following matrix is positive semi-definite:

$$K_{ij} = k(g_i, g_j)$$

A graph kernel implicitly defines an embedding into a Hilbert space

$$h \colon \mathscr{G} \to \mathscr{H}$$

such that

$$\langle h(g_i), h(g_j) \rangle = k(g_i, g_j)$$

This gives access to a large pool of machine learning tools.

Design and application of kernels for geometric (anatomical) trees and graphs will be the focus of my research at the MPI.