Department of Computer Science, University of Copenhagen

Tree-like shape spaces

François Lauze
francois@di.ku.dk

IPAM April 1. 2019

In collaboration with!

Aasa Feragen

Marlene de Bruijne

Jens Petersen

Pechin Lo

Søren Hauberg

Megan Owen

Mads Nielsen

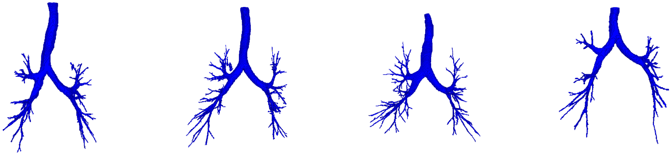
Asger Dirksen

Laura Hohwü Thomsen

Mathilde Marie Winkler

Airway shape modeling

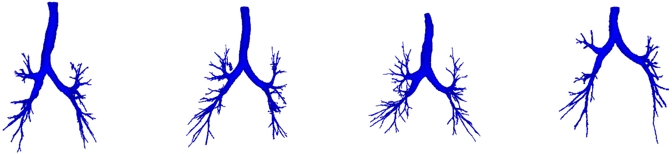
Starting point: What does the average human airway tree look like?



Wanted:

- A tree-space for geometric trees, where distances are given by geodesic (shortest path) length.
- Statistics defined using geodesics, analogous to manifold statistics.

Airway shape modeling

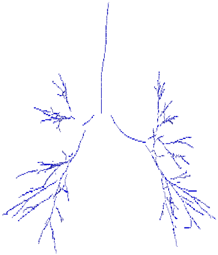


- Smoker's lung (COPD) is caused by inhaling damaging particles.
- Likely that damage made depends on airway geometry
- Reversely: COPD changes the airway geometry, e.g. airway wall thickness.
- $\rightsquigarrow$ Geometry can help diagnosis/prediction.

Airway shape modeling

Description of data:

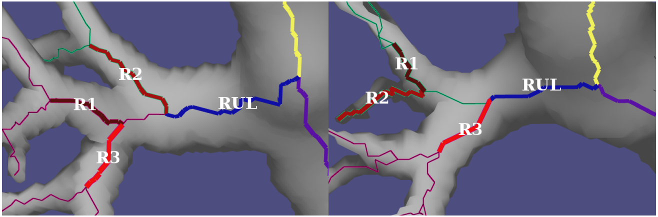
- Topology, branch shape – work with centerline trees embedded in $\mathbb{R}^3$.



Airway shape modeling

Description of data:

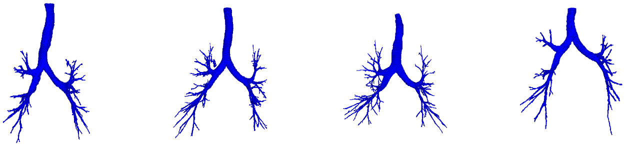
- Somewhat variable topology (combinatorics) in *anatomical* tree



Airway shape modeling

Description of data:

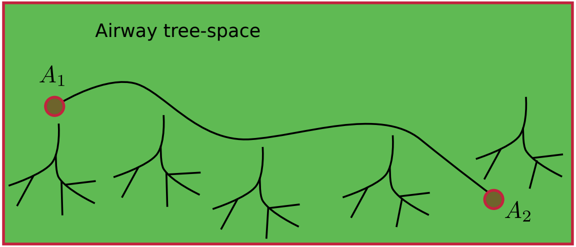
- Substantial amount of noise in *segmented* trees (missing or spurious branches), especially in COPD patients, *i.e. inherently incomplete data*



- Thus, we need a space for trees with different sizes, topologies and branch shapes.

Tree-space paths

We work with *Geodesic Metric Spaces*



A path in tree-space corresponds to a tree deformation.
A shortest path from A_1 to A_2 is a *geodesic*.

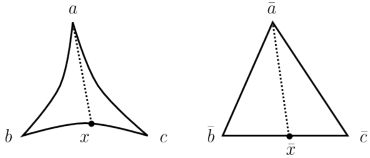
Statistics in metric spaces

- **Statistics:** Frechet/Ensemble means defined based on distance alone:

$$m = \operatorname{argmin} \sum_{i=1\ldots n} d^2(T, T_i).$$

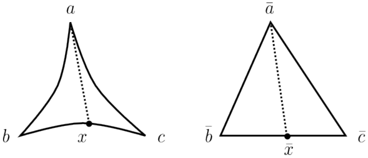
- **Computation:** Using geodesics

Curvature in metric spaces



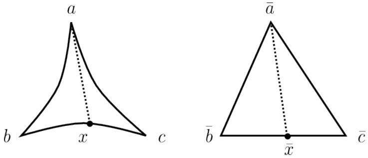
- A $CAT(0)$ space is a metric space in which geodesic triangles are "thinner" than for their comparison triangles in the plane; that is, $d(x, a) \leq d(\bar{x}, \bar{a})$.

Curvature in metric spaces



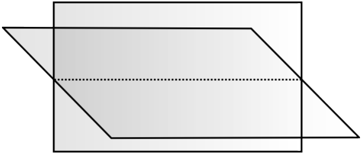
- A $CAT(0)$ space is a metric space in which geodesic triangles are "thinner" than for their comparison triangles in the plane; that is, $d(x, a) \leq d(\bar{x}, \bar{a})$.
- A space has non-positive curvature if it is locally $CAT(0)$.

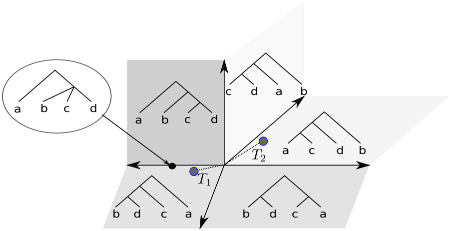
Curvature in metric spaces



- A $CAT(0)$ space is a metric space in which geodesic triangles are "thinner" than for their comparison triangles in the plane; that is, $d(x, a) \leq d(\bar{x}, \bar{a})$.
- A space has non-positive curvature if it is locally $CAT(0)$.
- (Similarly define curvature bounded by κ by using comparison triangles in hyperbolic space or spheres of curvature κ.)

Example





Theorem (see e.g. Bridson-Haefliger)

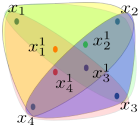
Let (X,d) be a $CAT(0)$ space; then all pairs of points have a unique geodesic joining them. The same holds locally in $CAT(\kappa)$ spaces, $\kappa \neq 0$. □

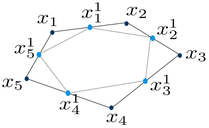
Statistics in metric spaces and the *CAT(0)* property

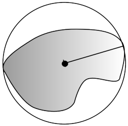
Theorem

In *CAT(0)* spaces:

- Frechet means are unique - $m = \operatorname{argmin}_{x \in X} \sum_{i=1}^{N} d^2(x, x_i)$[1]
- centroids are unique [2]
- Birkhoff shortening converges [3]
- circumcenters are unique [4]







[1] Sturm 2003

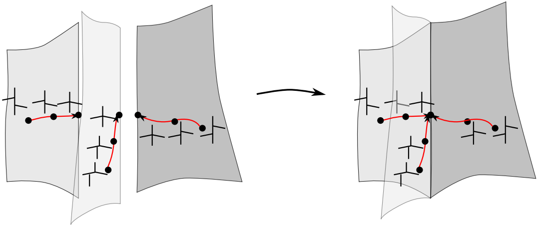
[2] Billera, Holmes, Vogtmann 2001

[3] Feragen, Hauberg, Nielsen, Lauze 2011

[4] Bridson-Haefliger 1999

A space of geometric trees: Intuition

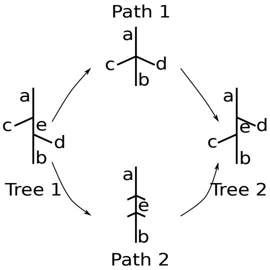
What would a path-connected space of deformable trees look like?

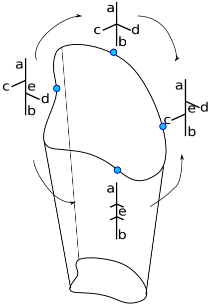


- Easy: Trees with same topology in their own "component"
- Harder: How are the components connected?
- Solution: glue collapsed trees, deforming one topology to another
- $\rightsquigarrow$ Stratified space, self intersections

A space of geometric trees: Intuition

The tree-space has conical "bubbles"



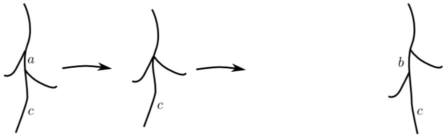


Classical example: Tree edit distance (TED)

- TED is a classical, algorithmic distance
- Tree-space with TED is a nonlinear metric space
- dist(T_1, T_2) is the minimal total cost of changing T_1 into T_2 through three basic operations:
- Remove edge, add edge, deform edge.

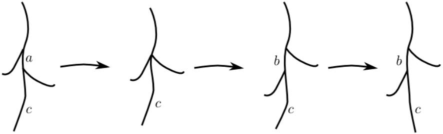
Classical example: Tree edit distance (TED)

- TED is a classical, algorithmic distance
- Tree-space with TED is a nonlinear metric space
- dist(T_1, T_2) is the minimal total cost of changing T_1 into T_2 through three basic operations:
- Remove edge, add edge, deform edge.



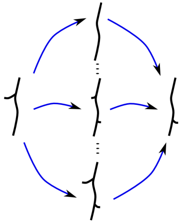
Classical example: Tree edit distance (TED)

- TED is a classical, algorithmic distance
- Tree-space with TED is a nonlinear metric space
- dist(T_1, T_2) is the minimal total cost of changing T_1 into T_2 through three basic operations:
- Remove edge, add edge, deform edge.



Classical example: Tree edit distance (TED)

- Tree-space with TED is a geodesic space, but almost all geodesics between pairs of trees are non-unique (infinitely many)



- Then what is the average of two trees? Many!
- Tree-space with TED has everywhere unbounded curvature.
- TED is *not* suitable for statistics.

Classical example: Tree edit distance (TED)

The problems can be "solved" by choosing specific geodesics[5].
OBS! Geometric methods can no longer be used for proofs, and one risks choosing problematic paths.

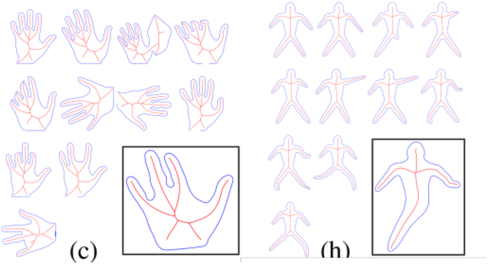
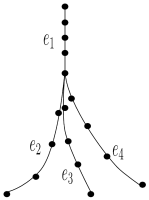


Figure: Trinh and Kimia compute average shock graphs using TED with the simplest possible choice of geodesics.

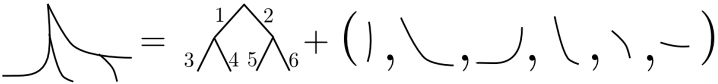
[5] Trinh, Kimia, CVPR workshops 2010

Tree representation [6]



Tree parametrization $(\mathscr{T}, x)$

- $\mathscr{T} = (V, E, r, <)$ finite rooted, ordered/planar binary tree, describing the tree topology (combinatorics)
- $x \in \bigoplus_{e \in E} (\mathbb{R}^d)^n$, $d = 3$, $n = \sharp$ landmarks/edge

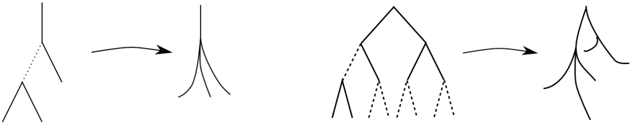


[6] Feragen, Lo, de Bruijne, Nielsen, Lauze, 2010

Tree representation [6]

We are allowing collapsed edges, which means that

- we can represent higher order vertices
- we can represent trees of different sizes using the same combinatorial tree $\mathscr{T}$



(dotted line = collapsed edge = zero/constant attribute)

[6] Feragen, Lo, de Bruijne, Nielsen, Lauze, 2010

The space of tree-like preshapes

Let $\mathscr{T}$ be a finite ordered (planar), rooted binary tree

Definition

Define the space of tree-like *pre*-shapes as the direct sum

$$\bigoplus_{e \in E} (\mathbb{R}^d)^n$$

where $(\mathbb{R}^d)^n$ is the edge shape space.

This is just a space of *pre-shapes*.

The space of tree-like preshapes

Let $\mathscr{T}$ be a finite ordered (planar), rooted binary tree

Definition

Define the space of tree-like *pre*-shapes as the direct sum

$$\bigoplus_{e \in E} (\mathbb{R}^d)^n$$

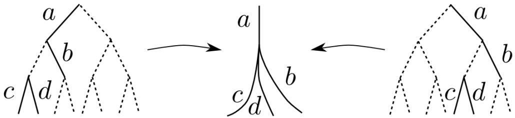
where $(\mathbb{R}^d)^n$ is the edge shape space.

This is just a space of *pre-shapes*.

Remove the ordered tree (vertices) hypothesis to represent non planar trees.

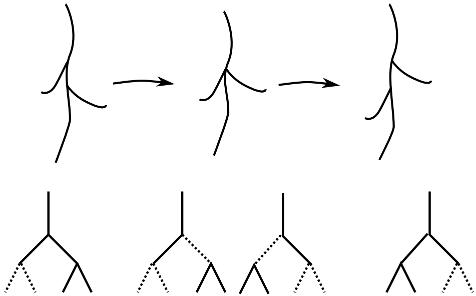
From pre-shapes to shapes

Many shapes have more than one representation



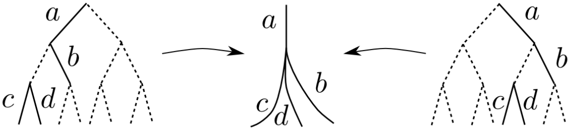
From pre-shapes to shapes

Not all shape deformations can be recovered as natural paths in the pre-shape space:



Shape space definition

- Start with the pre-shape space $X = \bigoplus_{e \in E} (\mathbb{R}^d)^n$.
- Define an equivalence $\sim$ by identifying points in X that represent the same ordered/unordered tree-shape.



- Space of ordered/unordered tree-like shapes $\tilde{X} = X/\sim$: folded vector space.
- Corresponds to identifying, or gluing together, subspaces $\{x \in X | x_e = 0 \text{ if } e \notin E_1\}$ and $\{x \in X | x_e = 0 \text{ if } e \notin E_2\}$ in X.

Shape space definition

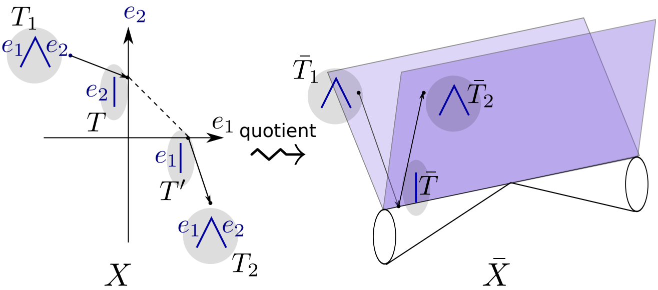


Figure: Space of ordered trees with at most 2 edges

Some details

- Preshape $x \in X = \bigoplus_E \left(\mathbb{R}^d\right)^n$, $E_x = \{e \in E | x_e \neq 0\}$
- Identification: $x \sim y \in E$ there exists a bijection

$$\phi : E_x \to E_y, \quad y_{\phi_e} = x_e$$

- Equivalence relation. Shape space: $\tilde{X} = X/\sim$
- Quotient pseudo-metric: d metric on X,

$$\bar{d}(\bar{x}, \bar{y}) = \inf \left\{ \sum_{i=1}^{k} d(x_i, y_i), x_1 \in \bar{x}, y_k \in \bar{y}, y_i \sim x_{i+1} \right\}$$

- $d = \| - \|_1$: *Tree-Edit-Distance*,
- $d = \| - \|_2$: *Quotient-Euclidean-Distance*,

Geodesics in tree-space

Theorem (Feragen, Lo, de Bruijne, Nielsen, Lauze, 2013)

- The quotient pseudometric $\bar{d}$ is a metric on $\tilde{X}$.
- $(\tilde{X}, \bar{d})$ is a geodesic space.
- Geodesics are not generally unique, neither for ordered nor unordered trees
- $\bar{d}$=QED: for a generic tree T_1 (of any size), for a generic second tree T_2 (of any size), there is a unique geodesic connecting them.
- At generic points, the space is locally $CAT(0)$.
- Its geodesics are locally unique at generic points.
- At non-generic points, the curvature is unbounded.

Metric and Geodesics

- From the distance

$$\bar{d}(\bar{x},\bar{y}) = \inf\left\{\sum_{i=1}^{k} d(x_i,y_i), x_1 \in \bar{x}, y_k \in \bar{y}, y_i \sim x_{i+1}\right\}$$

- k in formula above: topological transition the geodesics / in the combinatorics of trees.
- Out of these transitions: straight line segments. So what is the smallest k?
- When sufficiently unique geodesics can be computed, a lot more can.

$CAT(0)$, Geometric triangles

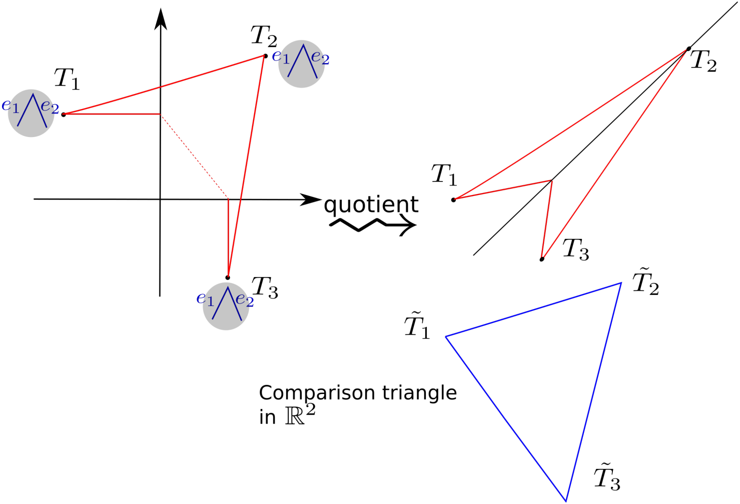


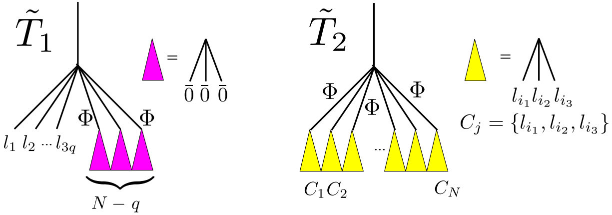
Figure: Geodesic Triangles in Tree-Shape Space

Computational Complexity of tree-space geodesics [7]

Assume edge attributes have dimension > 1
(for dim $= 1$, S. Provan).

Theorem

Computing QED geodesics between unordered trees is NP complete.



[7] (Feragen 2012)

Mean trees can be computed

[8] Leaf vasculature data:

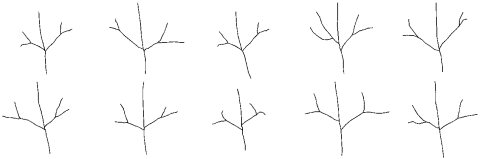


Figure: A set of vascular trees from ivy leaves form a set of planar tree-shapes.

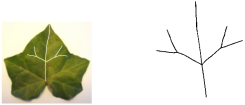


Figure: a): The vascular trees are extracted from photos of ivy leaves. b) The mean vascular tree.

[8] Feragen, Hauberg, Nielsen, Lauze, 2011

Mean trees can be computed

8

The mean upper airway tree

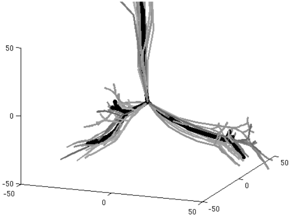


Figure: A set of upper airway tree-shapes along with their mean tree-shape.

[8] Feragen, Hauberg, Nielsen, Lauze, 2011

Mean trees can be computed

8

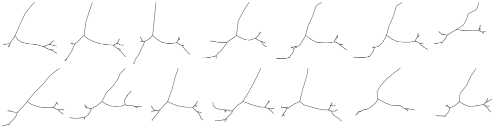


Figure: A set of upper airway tree-shapes (projected). (Fergen, Lo, de Bruijne, Nielsen, Lauze 2013)

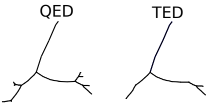


Figure: The QED and TED means (algorithm by Trinh and Kimia).

[8] Feragen, Hauberg, Nielsen, Lauze, 2011

Sturm means

Theorem (Existence of means[9])

- Means in non-positively curved spaces are unique.
- Means in non-positively curved spaces can be computed using a random infinite weighted midpoints sequence.
- $\rightsquigarrow$ Computation of mean trees [10].

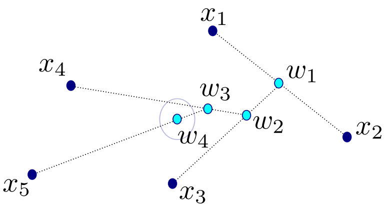
[9] Sturm 2003

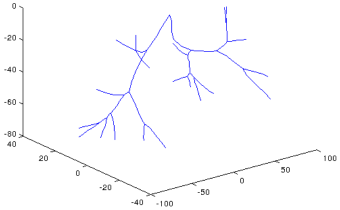
[10] Feragen, Hauberg, Nielsen, Lauze 2011; Miller et al 2012

Sturm means

Theorem (Existence of means[9])

- Means in non-positively curved spaces are unique.
- Means in non-positively curved spaces can be computed using a random infinite weighted midpoints sequence.
- $\rightsquigarrow$ Computation of mean trees [10].

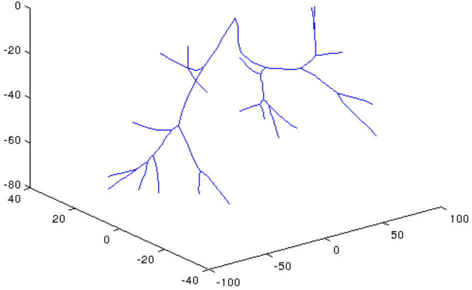




[9] Sturm 2003

[10] Feragen, Hauberg, Nielsen, Lauze 2011; Miller et al 2012

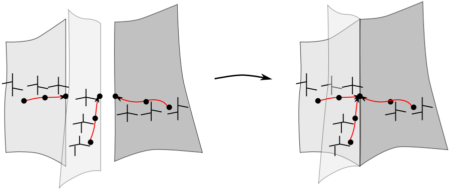
Statistics on larger trees: Mean airway [11]



[11] (Feragen, Owen and Feragen Hauberg, Nielsen, Lauze 2011)

How local are local statistics?

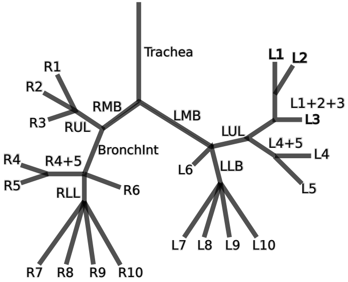
- **Restrict to:** all representations of certain restricted tree topologies.
- **Example 1:** Restrict to the set $\tilde{X}_N$ of trees with N leaves.



- **Example 2:** Restrict to all topologies occuring in airway trees.

Dealing with NP - Useful property of airways

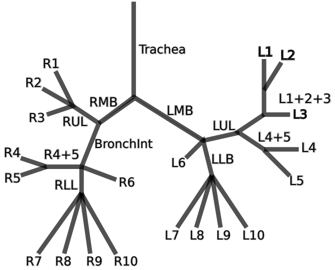
The first 6-8 generations of the airway tree are "similar" in different people.

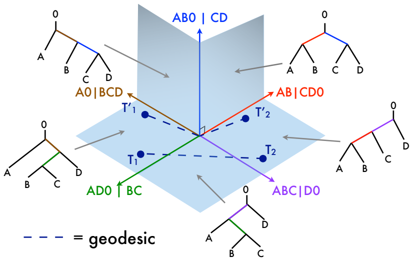


NB!: Not all present in all people; **not all present in all segmentations.**

Regularize via fixed leaf label sets[13]

- **Label** the "leaves" of your trees; constant leaf label set.
- A Variant of Billera-Holmes-Vogtmann (BHV) trees: gives a vectorized version of phylogenetic tree-space
- Polynomial time distance algorithms [12]



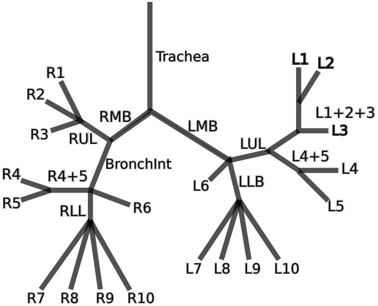


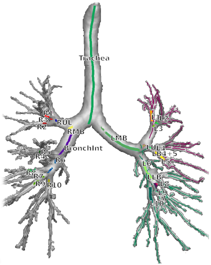
Right figure courtesy of Megan Owen.

[12] Owen, Provan, 2011

[13] Feragen et al. 2012

Geodesic airway branch labeling[14]



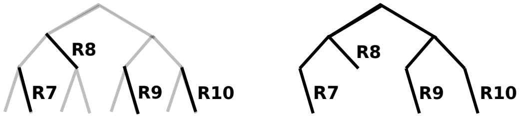


14 Feragen et al, MICCAI 2012.

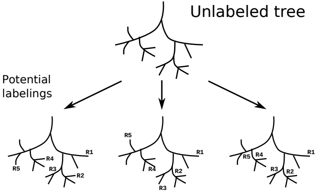
Geodesic airway branch labeling[14]

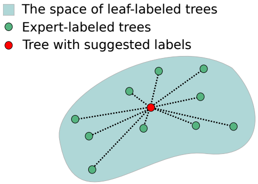
Idea:

- Generate suggested leaf label configurations and the corresponding tree spanning the labels



- **Evaluate** configuration in comparison with training data using geodesic distances between leaf-labeled airway trees



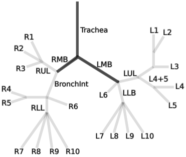


[14] Feragen et al, MICCAI 2012.

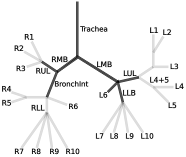
Geodesic airway branch labeling[14]

Idea:

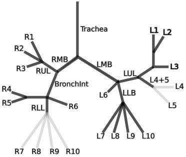
- Many possible label configurations
- Make tractable using a hierarchical labeling scheme



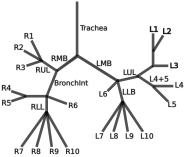
search 3 generations



search 2 and 2 generations



search 2, 2, 2 and 3 generations

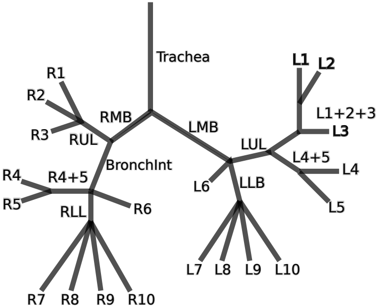


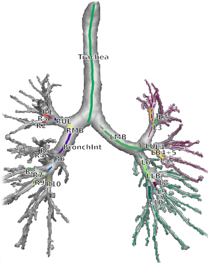
search 3 and 2 generations

[14] Feragen et al, MICCAI 2012.

Geodesic airway branch labeling[14]

- 40 airway trees from 20 subjects with different stages of COPD, hand labeled by 2 experts in pulmonary medicine.





- **Accuracy (leave-1-subject-out)**:
  Not significantly better (paired T-test), but as good as the experts

Label	Algorithm	Expert
Avg segmental labels	72.7	71.0
Avg all 29 labels	80.5	79.5

[14] Feragen et al, MICCAI 2012.

Geodesic airway branch labeling[14]

As reproducible as a medical expert:
(not significantly better in paired T-test):

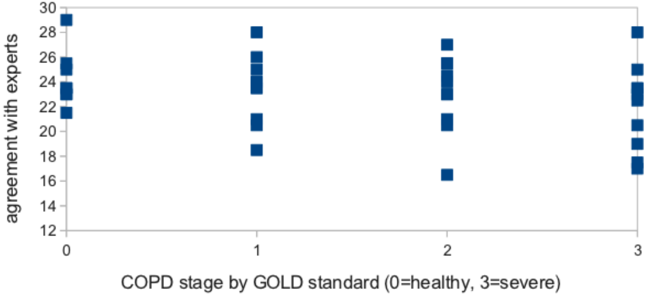
Two scans per subject, registered for label transfer. Reproducible labels / segmental labels on average, out of 20/29:

	Segmental	All
Expert 1	14.0	22.8
Expert 2	15.1	23.9
Automatic	15.2	24.0
Out of	20	29

[14] Feragen et al, MICCAI 2012.

Geodesic airway branch labeling[14]

Performance not significantly dependent on COPD level



Spearman: $(\rho = -0.22,\ p = 0.18)$

[14] Feragen et al, MICCAI 2012.

Automatically labeled databases $\rightsquigarrow$ statistics

Automatic labeling $\rightsquigarrow$ Danish Lung Cancer Screening Trial

- Database of 8016 airway trees
- 1692 unique subjects
- 732 women and 960 men

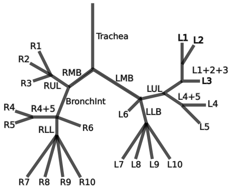
Hwo to perform statistics?

Hypothesis testing and localization of class-dependent differences[16]

- COPD/healthy samples: $A = a_1, \ldots, a_{N_1}$ and $B = b_1, \ldots, b_{N_2}$
- **Tools:** means and distances
- $\rightsquigarrow$ Permutation tests for equality of means and variance:

$$T(A,B) = d(\hat{\mu}_A, \hat{\mu}_B),$$
$$S(A,B) = \|var(A) - var(B)\|,$$
$$p_T = (1 + \sharp\{T_m \geq T_0 | m = 1, \ldots, M\})/(M+1), \ldots$$

- Consistent with clinical findings [15]



LABEL	P-VALUE mean	P-VALUE variance
full	**0.0010**	**0.0060**
RMB	**0.0020**	0.0939
RUL	0.2298	0.1668
BrInt	**0.0050**	0.1249
RLL	**0.0300**	0.0959
LMB	0.0859	**0.0210**
LUL	**0.0320**	**0.0390**
L1+2+3	**0.0260**	**0.0410**
LLB	0.5524	0.1588

[15] Hoesein et al, 2012

[16] Feragen ,Datar, Xu, Howard, Owen

Why the step to PCA is so hard

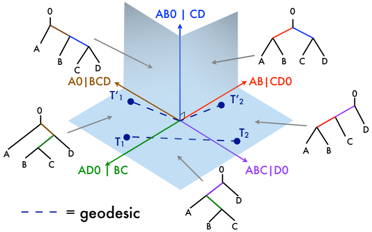
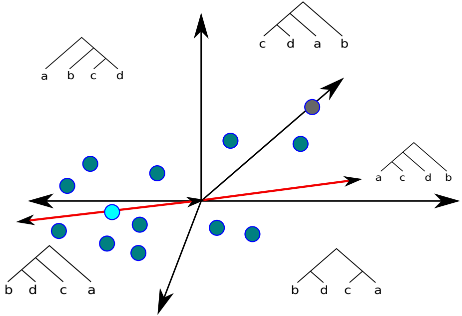


Figure: Courtesy of Megan Owen

- What is a "line"?
- How do you parametrize a "line"?
- How do you optimize over a family of "lines"?

First PCs [17]

- Definition:
  - Must go through *majority consensus tree*
  - PCs are "simple lines"



- Computational constraints

[17] Nye 2011

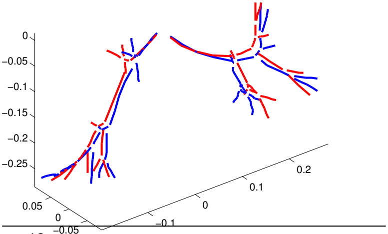
Set statistics[18]

- **Assume:** Dataset $\mathcal{X} \subset \mathcal{T}$ spans $\mathcal{T}$ well $\rightsquigarrow$ optimize over $\mathcal{X}$ ($|\mathcal{X}| = 8016$)

$$S = \arg.\min_{x \in \mathcal{T}} f(x), \qquad S = \arg.\min_{x \in \mathcal{X}} f(x)$$

- **Example:** Frechet mean and set mean

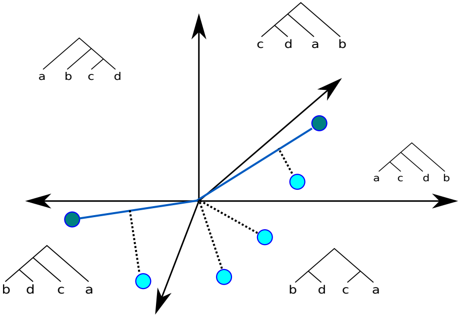
$$\mu = \arg.\min_{x \in \mathcal{T}} \sum_{i=1}^{N} d^2(x, x_i), \qquad \mu = \arg.\min_{x \in \mathcal{X}} \sum_{i=1}^{N} d^2(x, x_i)$$



[18] Feragen et al 2013

Projection onto geodesic segments[19]

- Parametrize geodesic segments γ by their endpoints
- $\mathrm{pr}_\gamma(x) = \arg.\min_{z \in \gamma(I)} d(x,z)^2$
- Non-positive curvature $\Rightarrow$ unique projection
- Computed with golden ratio search

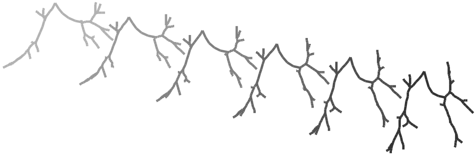


[19] Feragen et al 2013

Set PCA[20]

- First PC:
$$PC1 = \arg.\min_{x,x' \in \mathcal{T}} \sum_{i=1}^{N} d^2(x_i, \mathrm{pr}_{\gamma_{x,x'}}(x_i)),$$

- Set PC:
$$PC1 = \arg.\min_{x,x' \in \mathcal{X}} \sum_{i=1}^{N} d^2(x_i, \mathrm{pr}_{\gamma_{x,x'}}(x_i)),$$

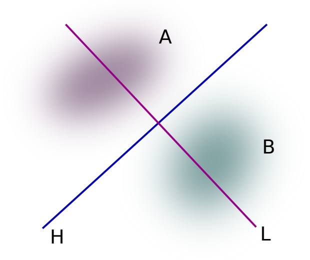


[20] Feragen et al, IPMI 2013

Fisher's LDA – the Euclidean version

- LDA as line maximizing projected class separation

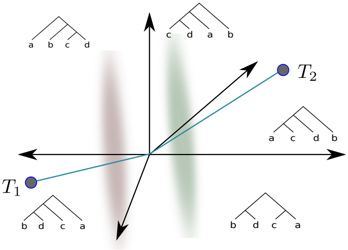
$$L_{LDA} = \arg.\max_L \frac{d^2(\hat{\mu}(\mathrm{pr}_L(A), \hat{\mu}(\mathrm{pr}_L(B))}{\hat{s}^2(\mathrm{pr}_L(A)) + \hat{s}^2(\mathrm{pr}_L(B))}.$$



Treespace LDA[21]

LDA as geodesic segment maximizing projected class separation

$$\gamma_{LDA} = \arg.\max_{x,x' \in \mathcal{T}} \frac{d^2(\hat{\mu}(\mathrm{pr}_{\gamma_{x,x'}}(A)), \hat{\mu}(\mathrm{pr}_{\gamma_{x,x'}}(B)))}{\hat{s}^2(\mathrm{pr}_{\gamma_{x,x'}}(A)) + \hat{s}^2(\mathrm{pr}_{\gamma_{x,x'}}(B))}.$$



[21] Feragen et al, IPMI 2013

Set LDA[22]

LDA as geodesic segment maximizing projected class separation

$$\gamma_{LDA} = \arg.\max_{x,x' \in \mathcal{X}} \frac{d^2(\hat{\mu}(\mathrm{pr}_{\gamma_{x,x'}}(A), \hat{\mu}(\mathrm{pr}_{\gamma_{x,x'}}(B))}{\hat{s}^2(\mathrm{pr}_{\gamma_{x,x'}}(A)) + \hat{s}^2(\mathrm{pr}_{\gamma_{x,x'}}(B))}.$$

22 23

Wrap-up

- We have: A tree-space framework for analysis of geometric trees
- Nice geometric properties for statistical analysis
- Leaf label assignment gives computational advantages at a modeling cost
- Even in the simpler space of leaf-labeled trees, many statistical problems remain open.
- For tree-shape spaces, curvature questions linked to combinatorial complexity: locality of statistics?