CLIC

Common Lines Implied Clustering

Donovan Webb

eBIC/University of Bath

March 23rd, 2020

Table of Contents

1 Single Lines

2 Finding Common Lines

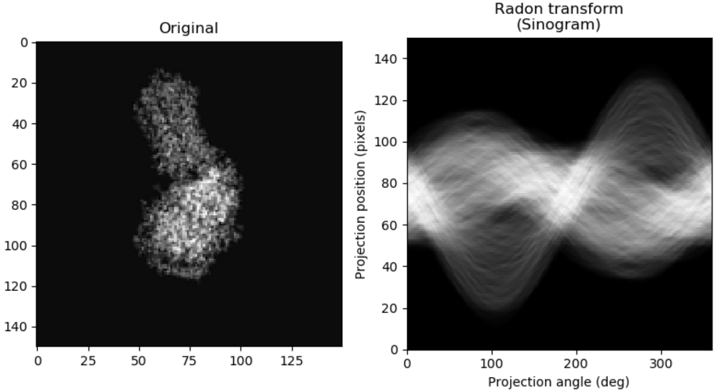
3 Clustering

4 Reconstruction

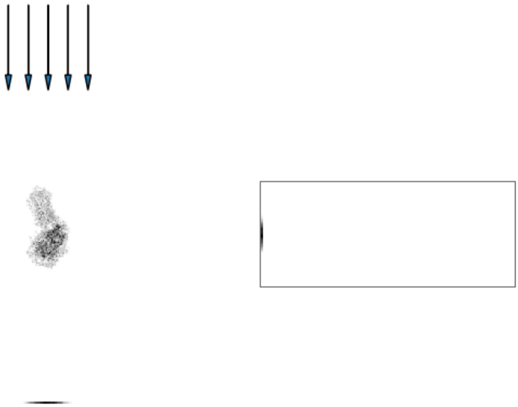
5 Full Pipeline

Single Lines

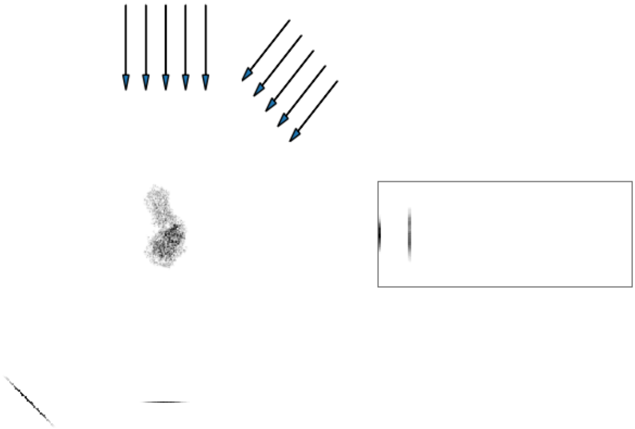
The Radon Transform



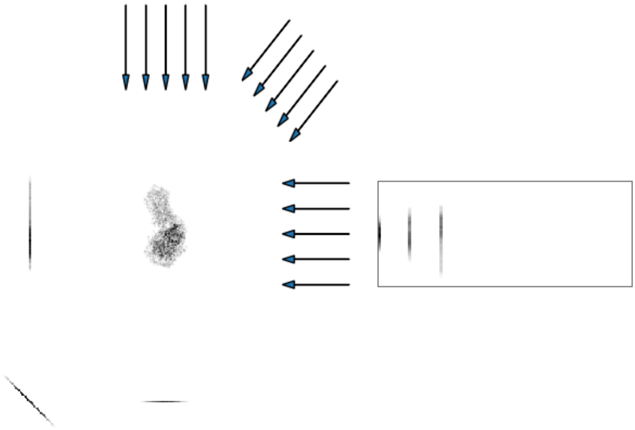
The Radon Transform



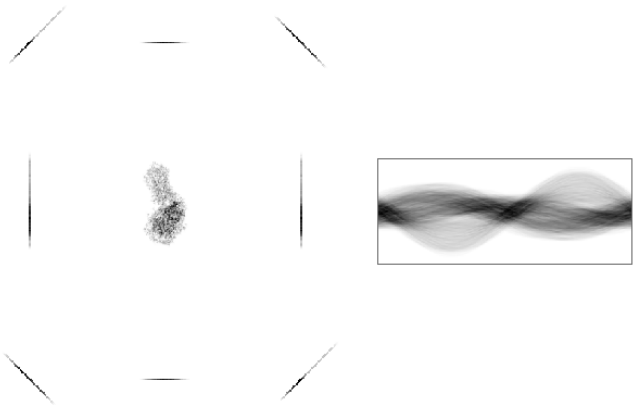
The Radon Transform



The Radon Transform

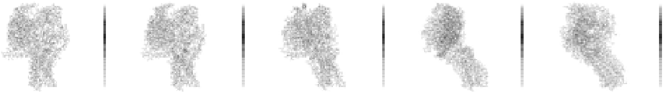


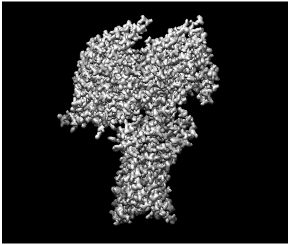
The Radon Transform



Common Lines

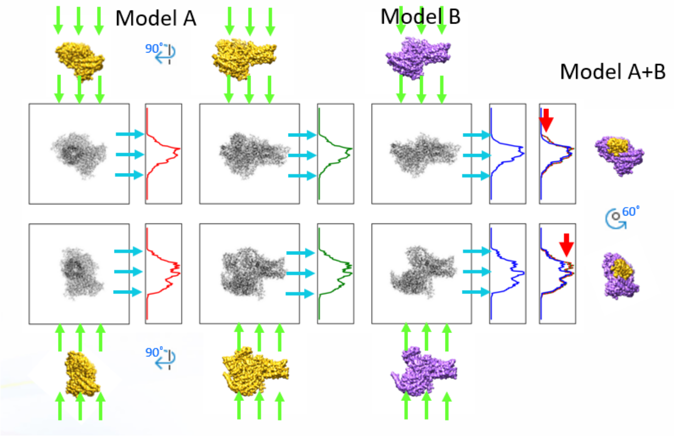
Two projections of the same 3D volume share at least one common line in the Radon transform





Common Lines

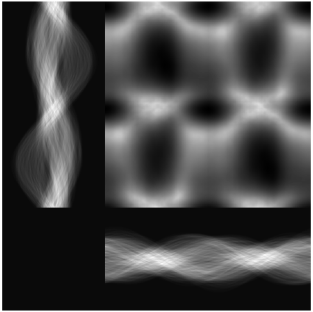
What about two different 3D volumes?



Finding Common Lines

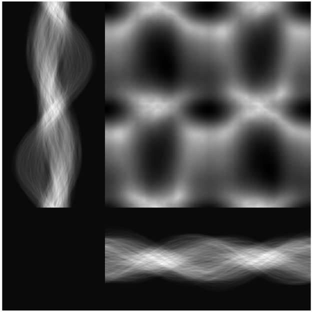
Sinogram Cross Correlation

Finding the common line between two sinograms



Sinogram Cross Correlation

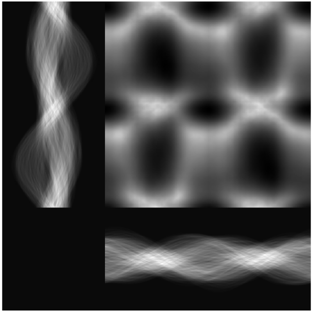
Finding the common line between two sinograms



But what about N sinograms?

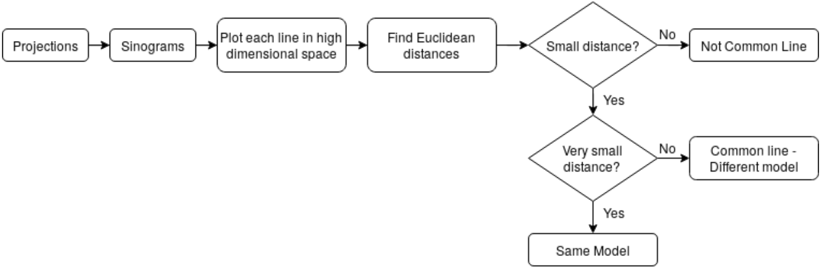
Sinogram Cross Correlation

Finding the common line between two sinograms

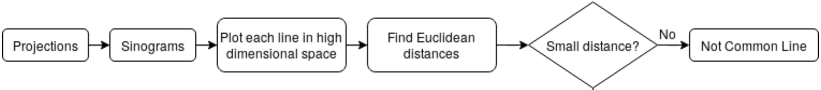


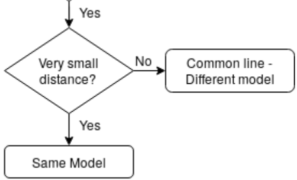
But what about N sinograms?
What about N sinograms from a heterogenous dataset?

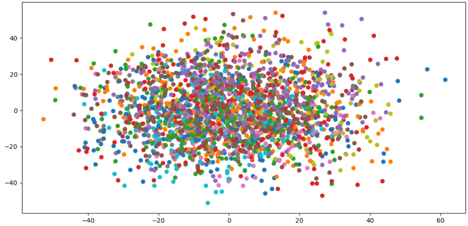
Pipeline


Projections
Sinograms
Plot each line in high dimensional space
Find Euclidean distances
Small distance?
No
Not Common Line
Yes
Very small distance?
No
Common line - Different model
Yes
Same Model

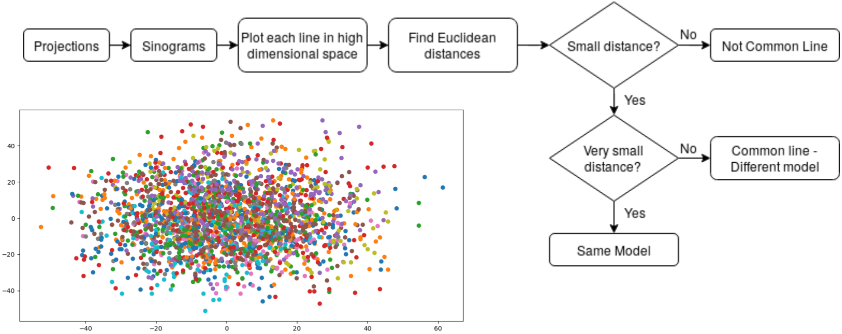
Pipeline


Projections
Sinograms
Plot each line in high dimensional space
Find Euclidean distances
Small distance?
No
Not Common Line


Yes
Very small distance?
No
Common line - Different model
Yes
Same Model

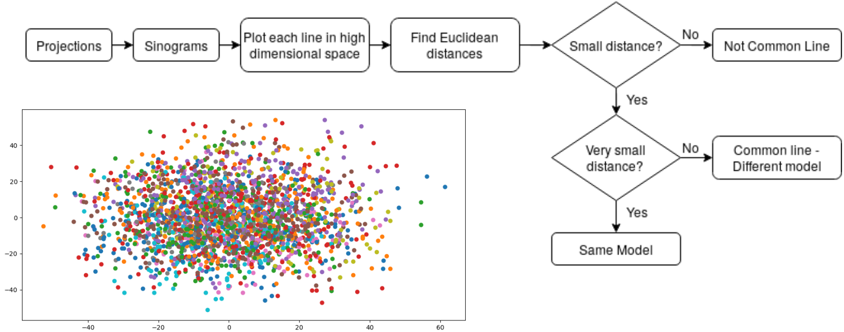

40
20
0
−20
−40
−40
−20
0
20
40
60

Pipeline



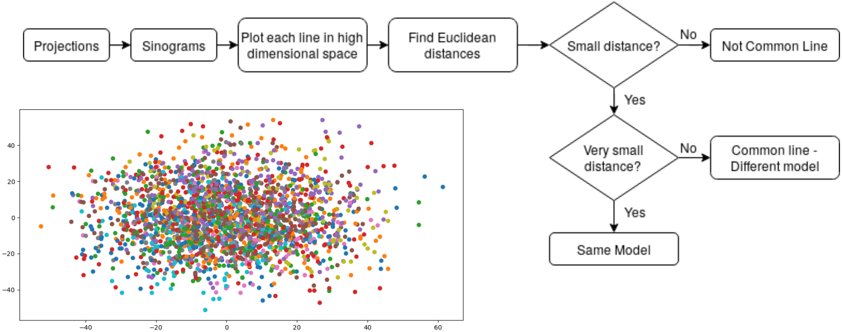
Slow.

Pipeline



Slow. Exhaustive.

Pipeline



Slow. Exhaustive. Doesn't handle noise well.

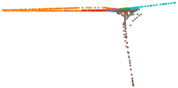
Dimensional Reduction

Find features - Reduce noise

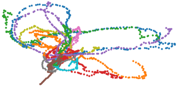
Linear (PCA)

Non-Linear

a)



b)



c)

d)

a: LLE, b: ISOMAP, c: TSNE, d: UMAP

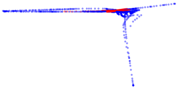
Dimensional Reduction - ground truths

Find features - Reduce noise

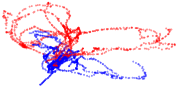
Linear (PCA)

Non-Linear

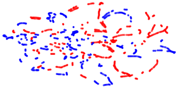
a)



b)



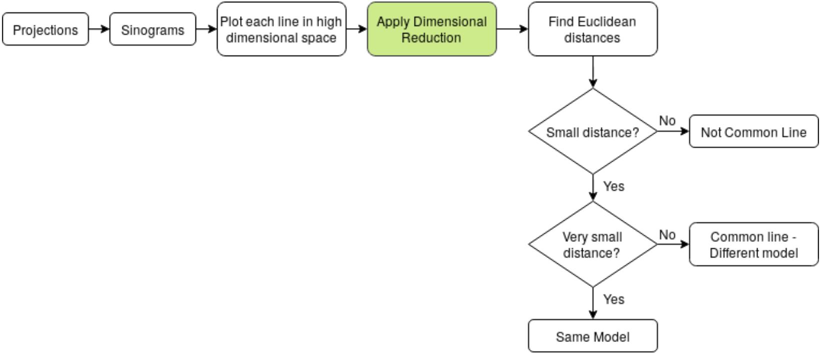
c)



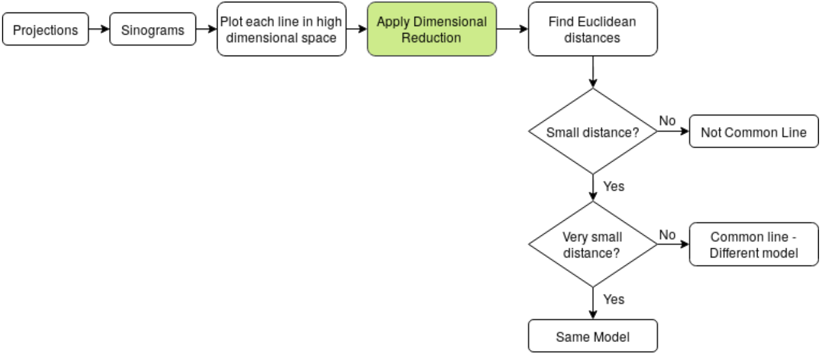
d)

a: LLE, b: ISOMAP, c: TSNE, d: UMAP

Pipeline


Projections
Sinograms
Plot each line in high dimensional space
Apply Dimensional Reduction
Find Euclidean distances
Small distance?
No
Not Common Line
Yes
Very small distance?
No
Common line - Different model
Yes
Same Model

Pipeline



But how do we assign clusters?

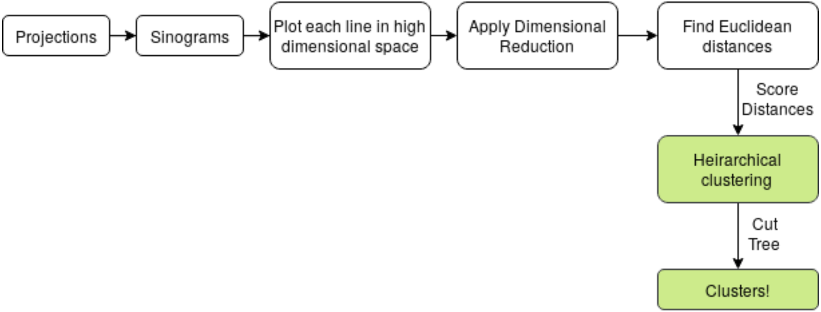
Clustering

The data

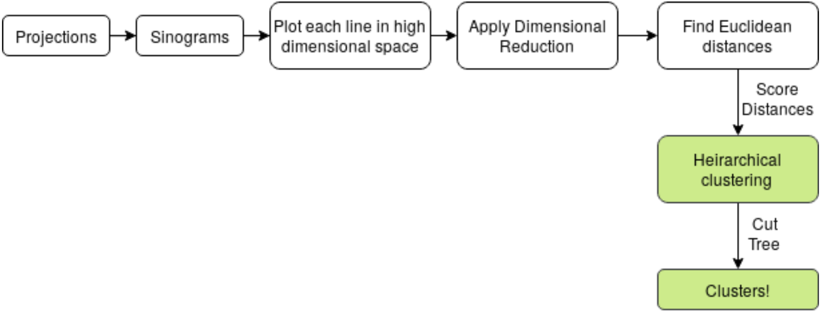
Ground truth: Good seperatation between two classes - but discontinuous

Heirarchical clustering

Pipeline


Projections
Sinograms
Plot each line in high dimensional space
Apply Dimensional Reduction
Find Euclidean distances
Score Distances
Heirarchical clustering
Cut Tree
Clusters!

Pipeline

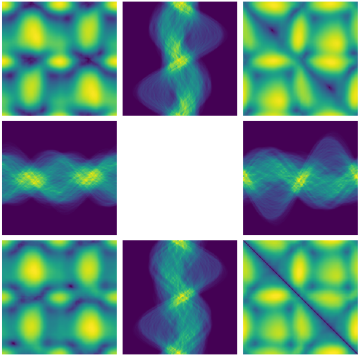


Just a model left!

Reconstruction

Angular recovery from 3 Common lines

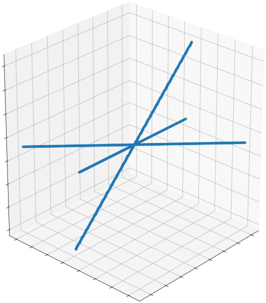
Common line gives axis of rotation. Three common lines gives 2 unique solutions for 3D orientation (One mirror of other)



Angular Reconstitution: A Posteriori Assignment of Projection Directions for 3d Reconstruction. Van Heel 1987

Angular recovery from 3 Common lines

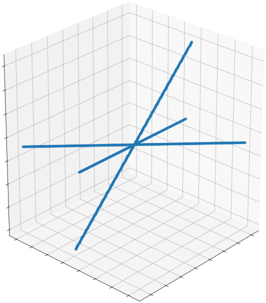
Common line gives axis of rotation. Three common lines gives 2 unique solutions for 3D orientation (One mirror of other)



Angular Reconstitution: A Posteriori Assignment of Projection Directions for 3d Reconstruction. Van Heel 1987

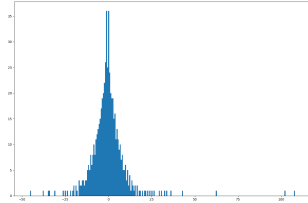
Angular recovery from 3 Common lines

Common line gives axis of rotation. Three common lines gives 2 unique solutions for 3D orientation (One mirror of other)

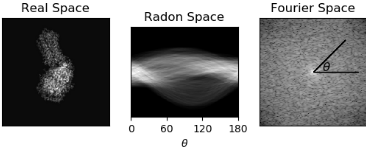


Angular Reconstitution: A Posteriori Assignment of Projection Directions for 3d Reconstruction. Van Heel 1987

Eigenvector Relaxation



Aim: Given all common lines c for projection
matrices R for each P to give greatest conse



$$cij = (cos(\theta_{ij}), sin(\theta_{ij}), 0), \quad cji = (cos(\theta_{ji}), sin(\theta_{ji}), 0) \qquad (1)$$

$$max \sum_{i \neq j} R_i c_{ij} \cdot R_j c_{ji} \qquad (2)$$

Maths*! Make large ($2N \times 2N$) symmetric matrix S. Can recover R for each P from top 3 eigenvectors of S that maximise (2)!

*Three Dimensional Structure Determination from Common Lines in Cryo-EM by Eigenvectors and Semidefinite Programming. A.Singer and Y.Shkolnisky

Full Pipeline

A full pipeline of the procedure. 2d projs ¿ 2d sins ¿ 1d lines ¿ TSNE ¿ agglo ¿ clusters ¿ split into sep datasets ¿ find common lines ¿ eigenvector relaxation ¿ Models