ToMAToMP: Robust and Multi-Parameter Topological Clustering

Ludo Andrianirina

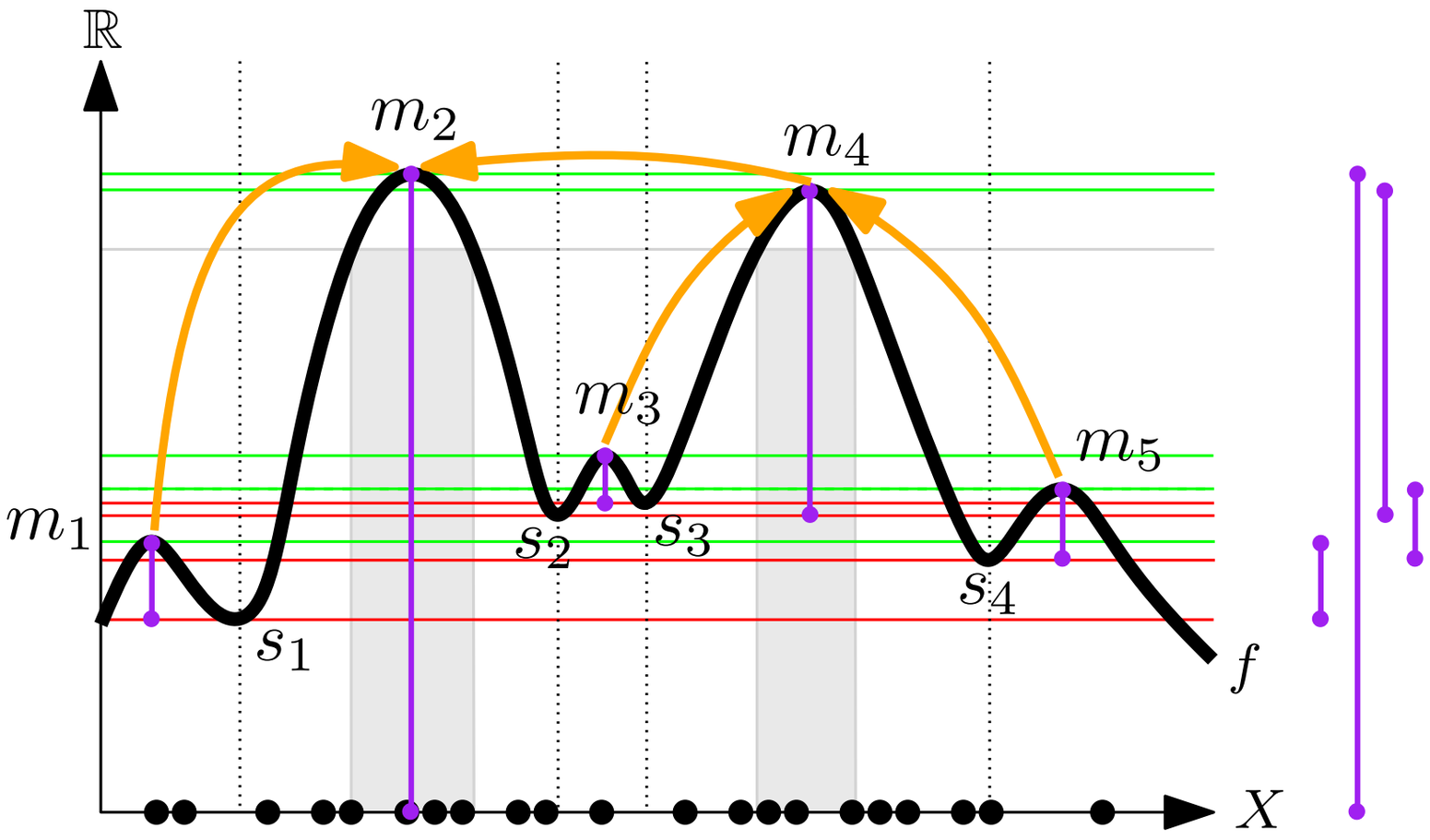
Mathieu Carrière

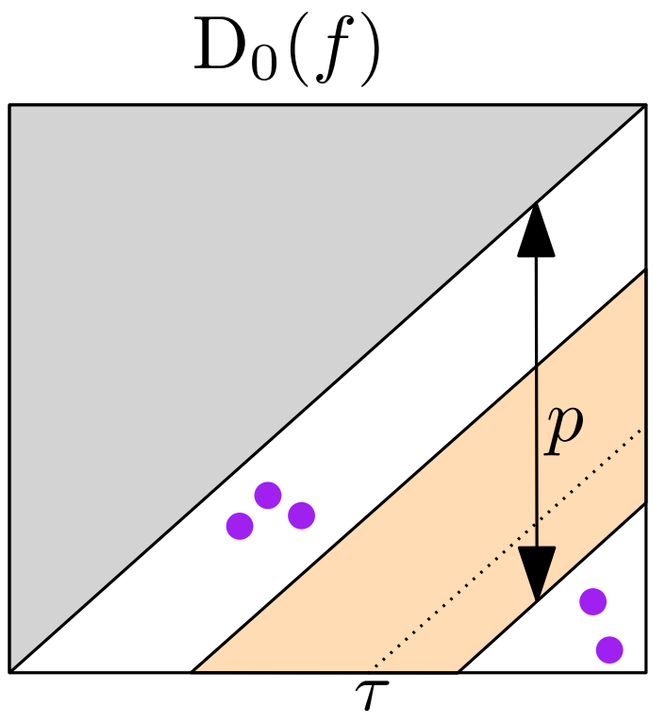
DataShape, Centre Inria d'Université Côte d'Azur



ToMAToMP

ToMATo is a *topological clustering* method: it computes the connected components of superlevel sets of functions.

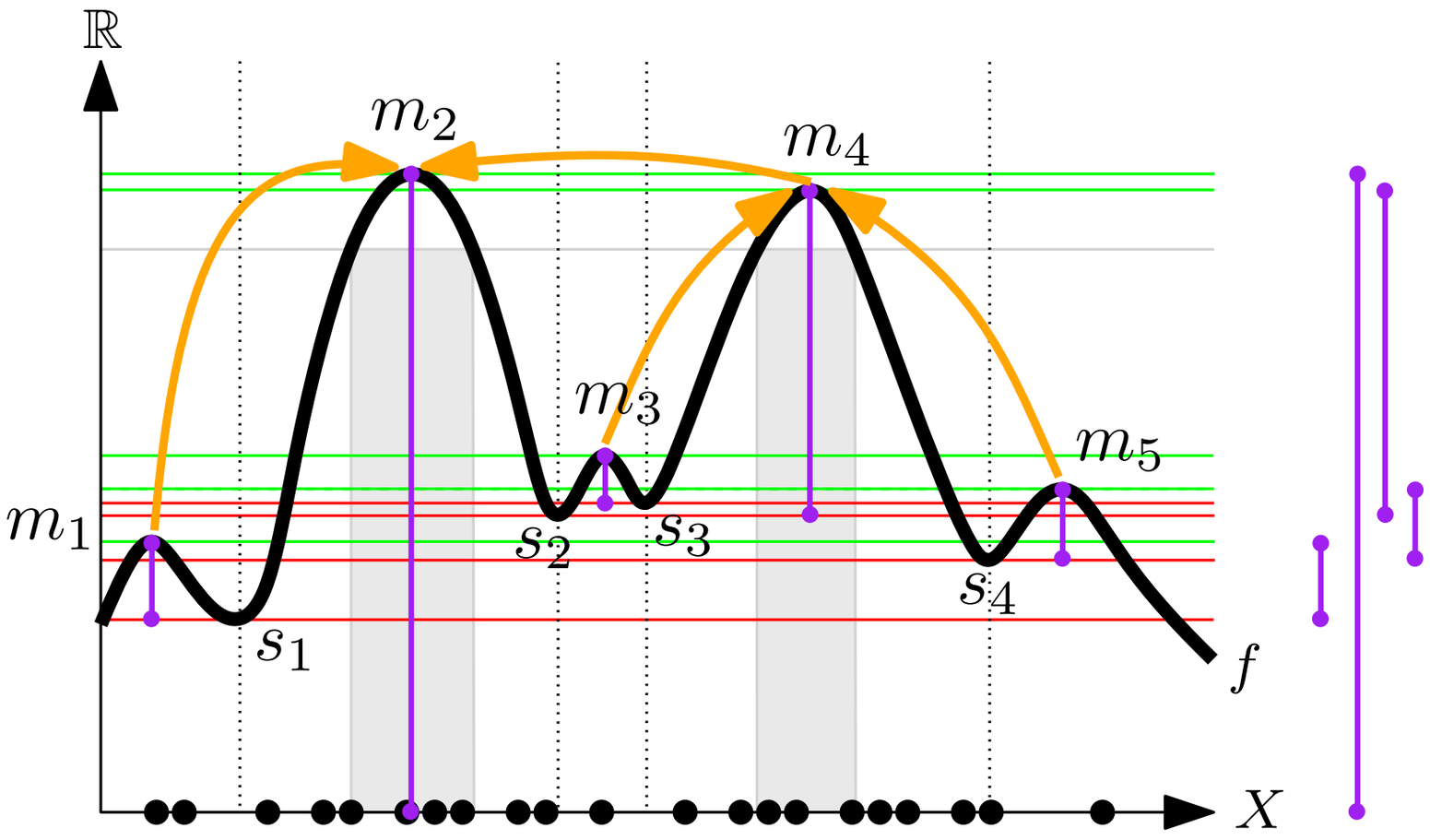


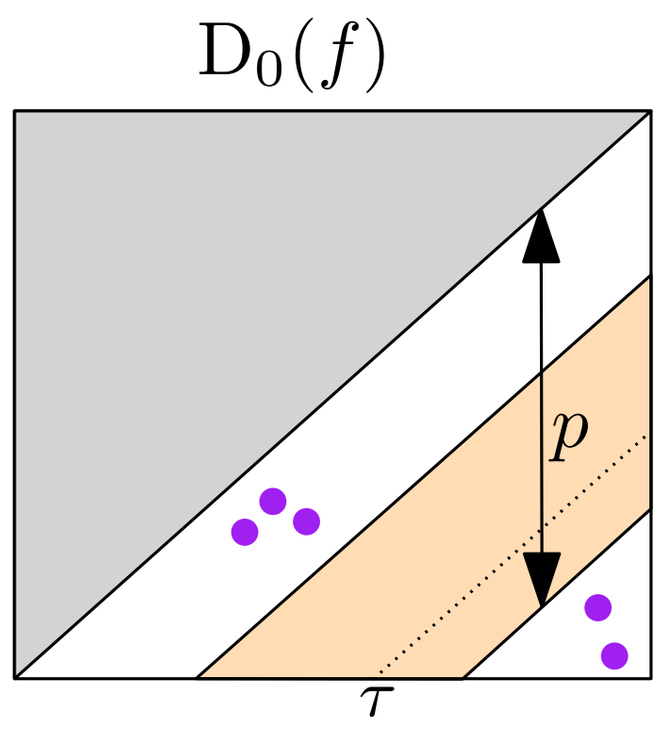


Persistence-Based Clustering in Riemannian Manifolds, Chazal et al., *J. ACM*, 2013

ToMAToMP

ToMATo is a *topological clustering* method: it computes the connected components of superlevel sets of functions.

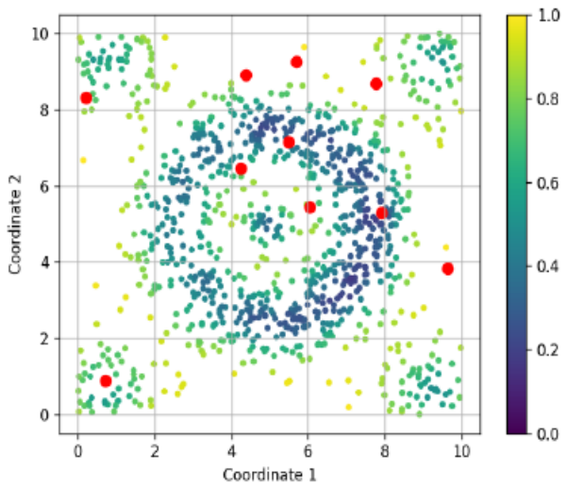




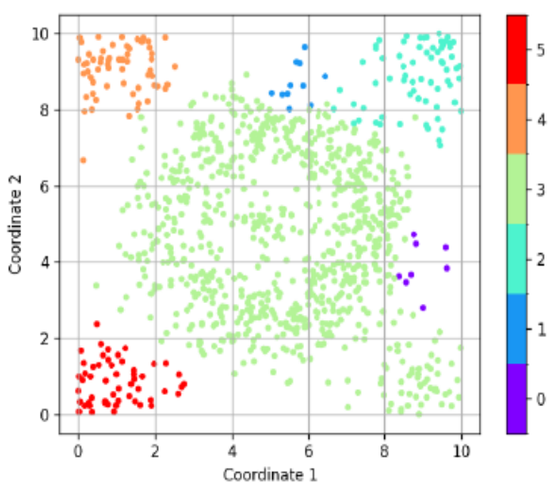
Persistence-Based Clustering in Riemannian Manifolds, Chazal et al., *J. ACM*, 2013

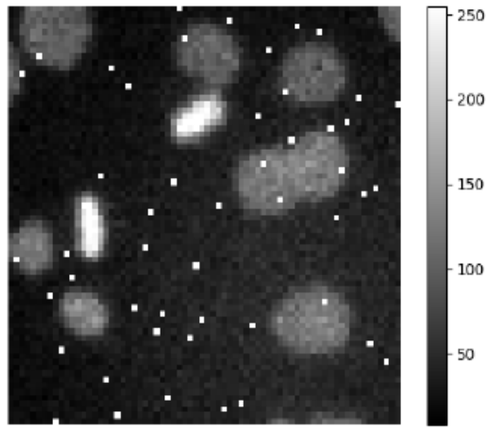
However, this method is, by design, highly sensitive to outliers.

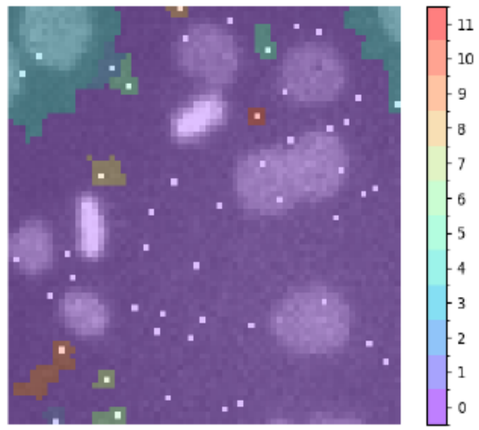
ToMAToMP



ToMATo



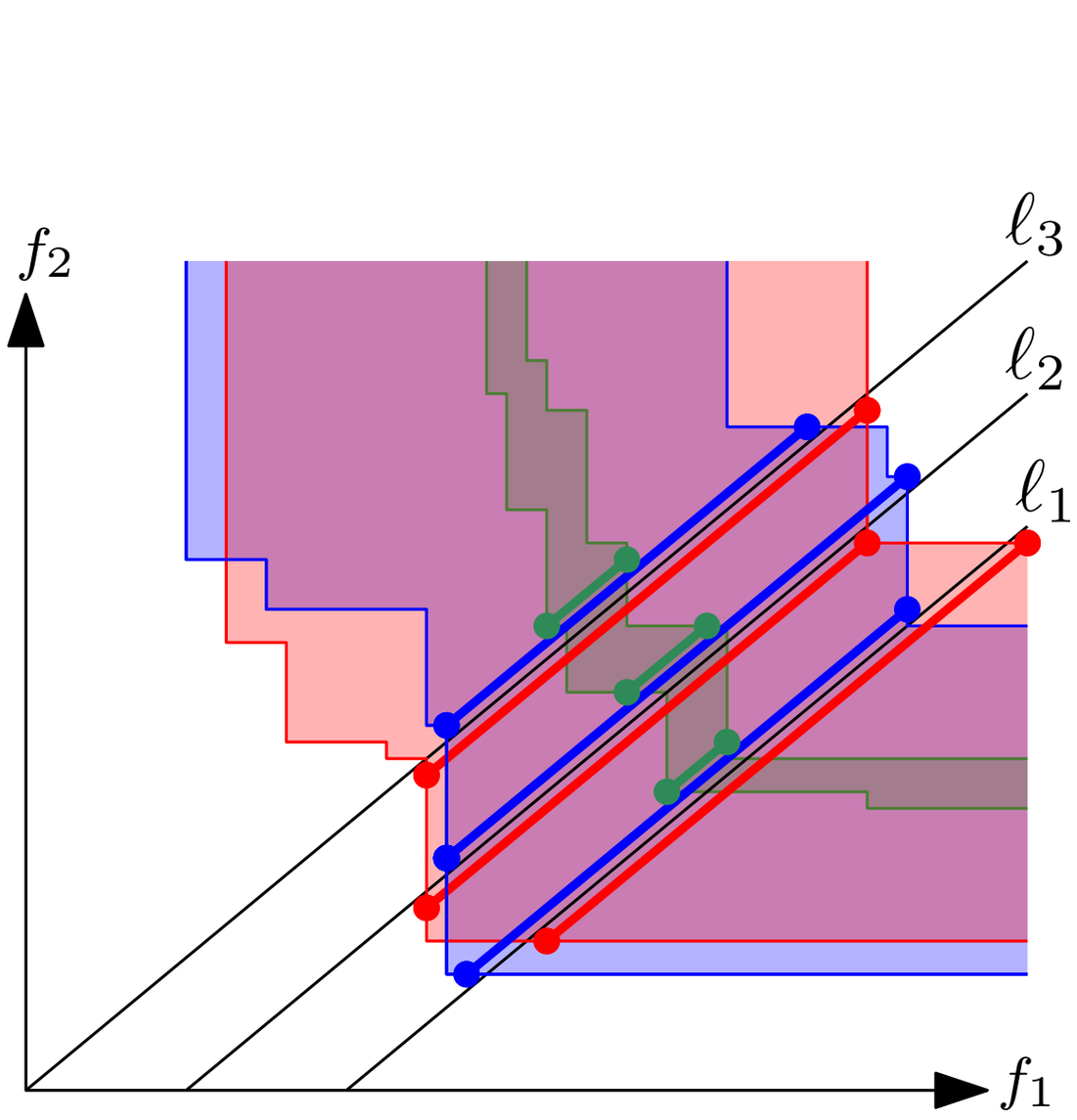


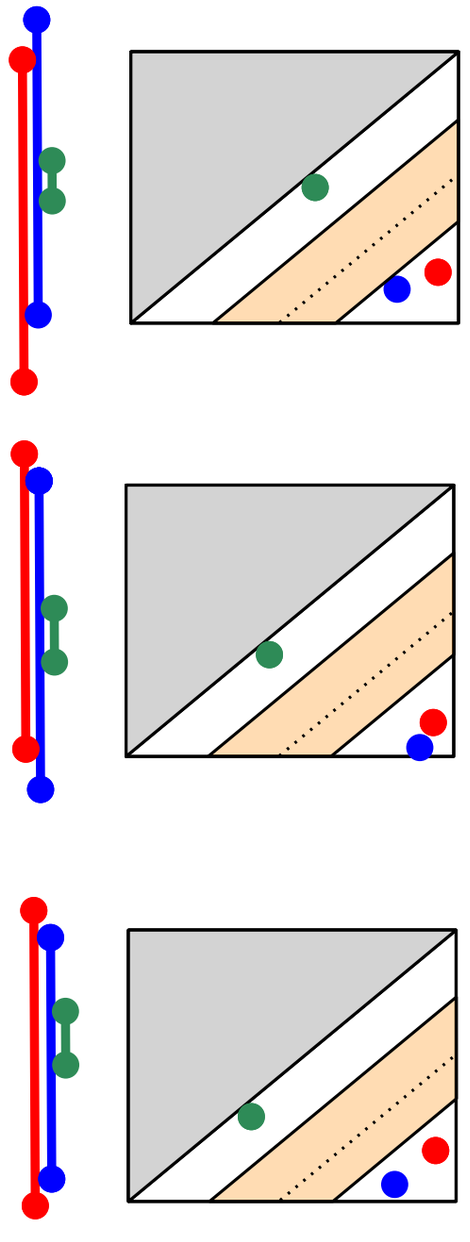


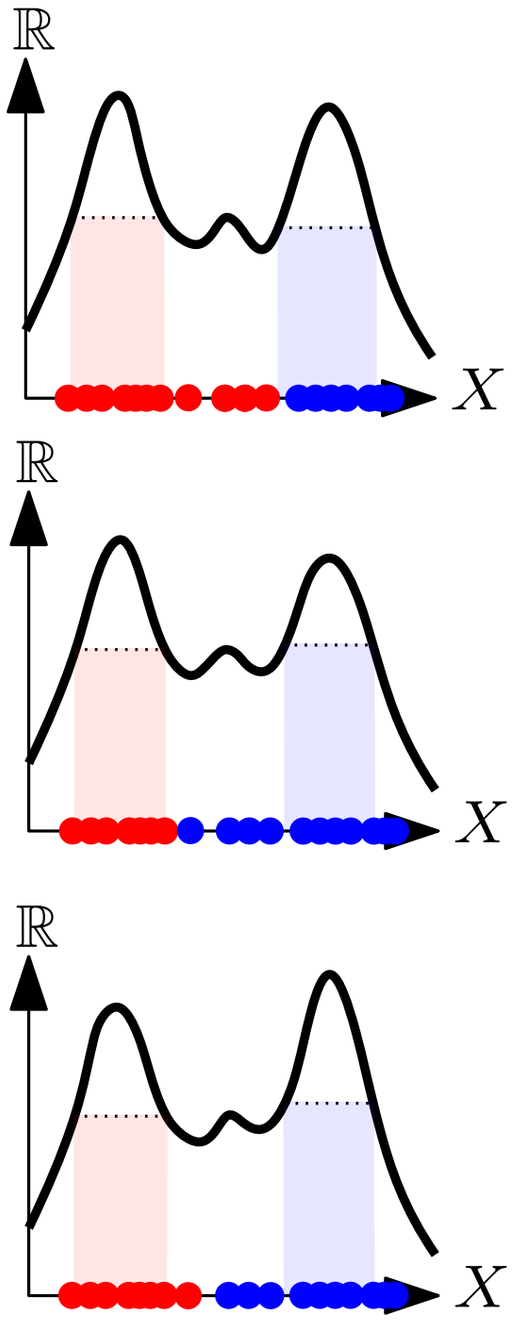
However, this method is, by design, highly sensitive to outliers.

ToMAToMP

Idea: ToMAToMP: apply ToMATo on lines and pick majoritary labels.

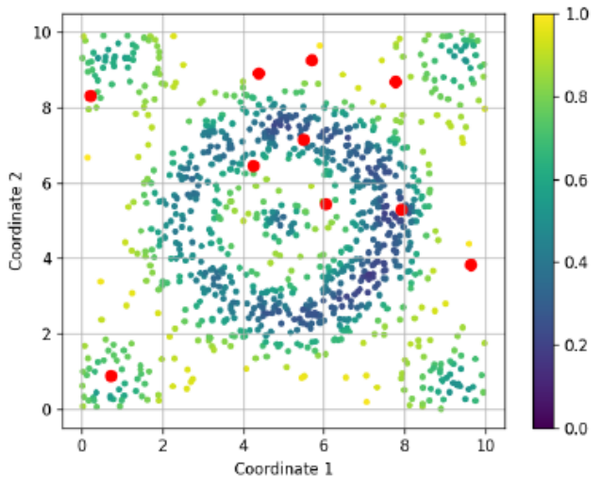




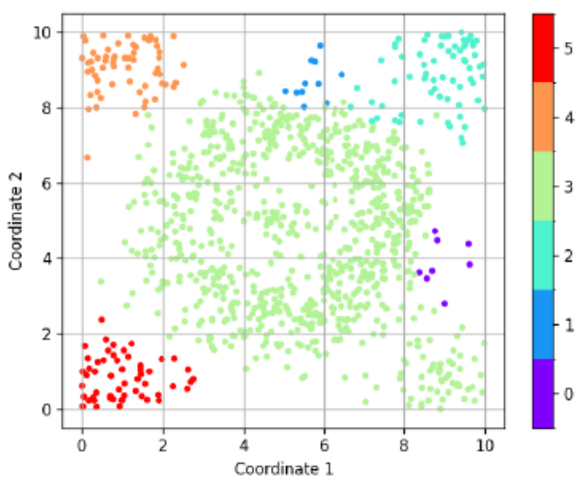


ToMATo**MP**

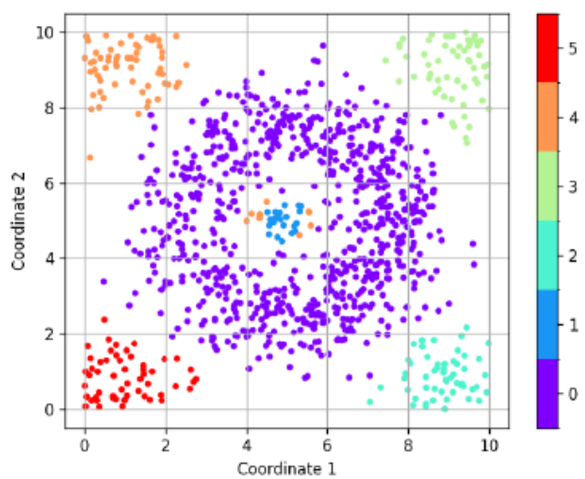
Idea: ToMATo**MP**: apply ToMATo on lines and pick majoritary labels.

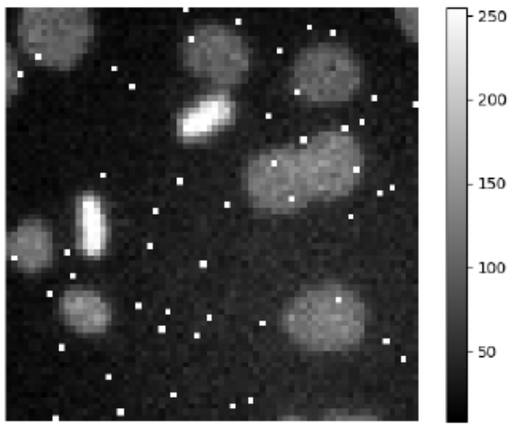


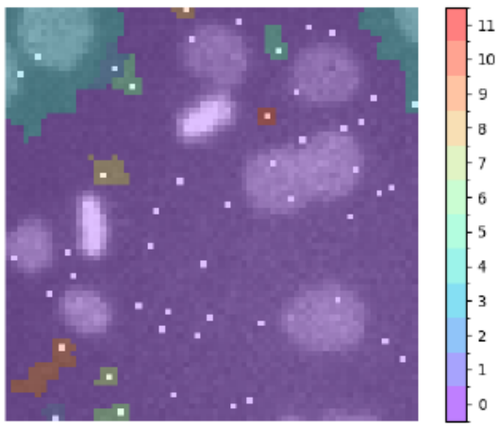
ToMATo

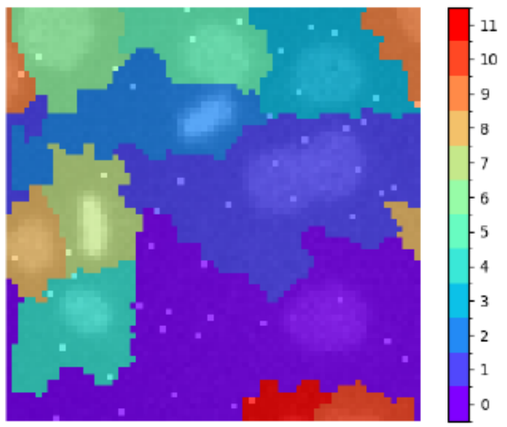


ToMATo**MP**









ToMAToMP

Idea: ToMAToMP: apply ToMATo on lines and pick majoritary labels.

Th: Assume that X is a δ-sample from a smooth manifold, and that (f_1, f_2) have K joint peaks of prominence at least $p > 0$. Then, if $2c\delta < \tau < p - 2c\delta$ and $\eta \leq \min\{(p - 4c\delta)/16, d^*\}$, where d^* is the smallest distance between fibered barcode points above p, then ToMAToMP outputs K clusters which coincides with the (cores of the) joint peaks.

ToMAToMP

Idea: ToMAToMP: apply ToMATo on lines and pick majoritary labels.

Th: Assume that X is a δ-sample from a smooth manifold, and that (f_1, f_2) have K joint peaks of prominence at least $p > 0$. Then, if $2c\delta < \tau < p - 2c\delta$ and $\eta \leq \min\{(p - 4c\delta)/16, d^*\}$, where d^* is the smallest distance between fibered barcode points above p, then ToMAToMP outputs K clusters which coincides with the (cores of the) joint peaks.

Method	`synthetic no-radius`	`3dshape no-radius`	`synthetic outlier`	`image outlier`
KMeans	0.43	0.41 $\pm$ 0.20	0.38 $\pm$ 0.03	0.69 $\pm$ 0.00
SpectralClustering	0.69	0.38 $\pm$ 0.18	0.67 $\pm$ 0.00	0.69 $\pm$ 0.00
AgglomerativeClustering	0.45	**0.42 $\pm$ 0.19**	0.43 $\pm$ 0.03	0.67 $\pm$ 0.00
DBSCAN	0.00	0.03 $\pm$ 0.11	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
DensityPeaks	0.53	0.34 $\pm$ 0.18	0.53 $\pm$ 0.01	0.63 $\pm$ 0.00
HDBSCAN	0.75	0.25 $\pm$ 0.13	0.72 $\pm$ 0.02	0.12 $\pm$ 0.01
Louvain	0.57	0.41 $\pm$ 0.15	0.56 $\pm$ 0.02	0.67 $\pm$ 0.00
Leiden	0.56	0.41 $\pm$ 0.16	0.56 $\pm$ 0.02	0.68 $\pm$ 0.00
ToMATo (min)	0.00	0.26 $\pm$ 0.21	–	–
ToMATo (mean)	0.66	0.38 $\pm$ 0.19	–	–
ToMATo	–	–	0.59 $\pm$ 0.12	0.55 $\pm$ 0.02
ToMAToMP	**0.82**	0.35 $\pm$ 0.20	**0.93 $\pm$ 0.01**	**0.74 $\pm$ 0.00**

Table 1: AMI scores for `no-radius` and `outlier` experiments. Results are averaged over all meshes (resp. frames) of the `3dshape` (resp. `image`) dataset.

Thank you!!