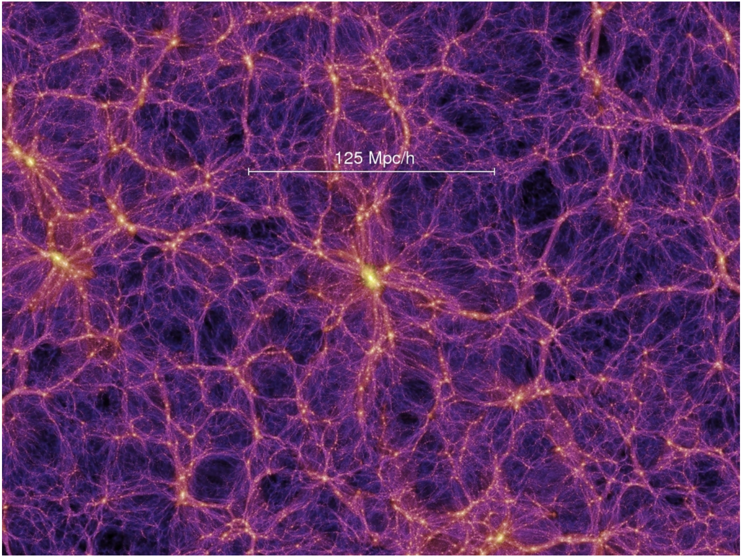
When Geometry and Statistics meet Cosmology: the challenge of detecting cosmic webs.

Yen-Chi Chen

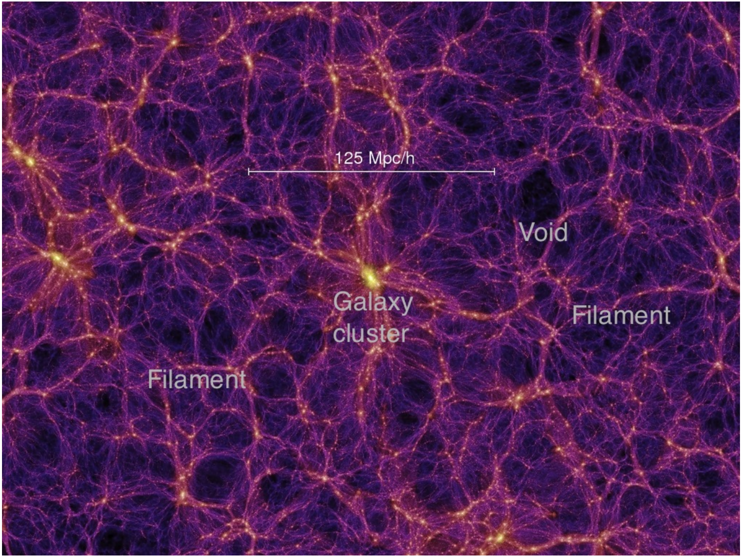
Department of Statistics
University of Washington
◦ Supported by NSF DMS - 195278 and DMS - 2112907 and DMS - 2141808.
Joint work with Yikun Zhang

Cosmic Web: What Does Our Universe Look Like



Credit: Millennium Simulation

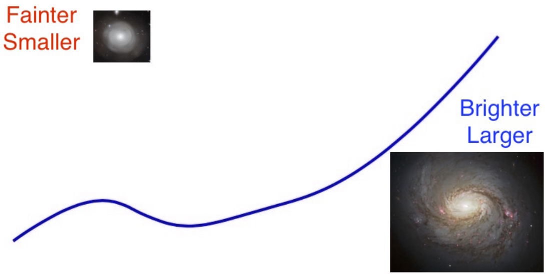
Cosmic Web: What Does Our Universe Look Like



Credit: Millennium Simulation

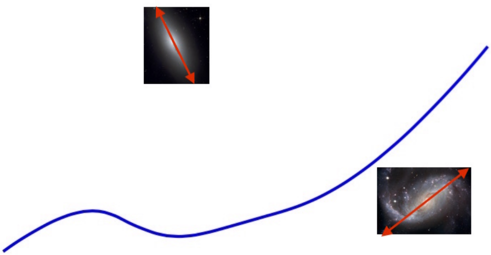
The Importance of Filaments

- A galaxy's brightness, size, and mass are associated with the distance to filaments.



The Importance of Filaments

- A galaxy's brightness, size, and mass are associated with the distance to filaments.
- A galaxy's alignment is associated with filaments.



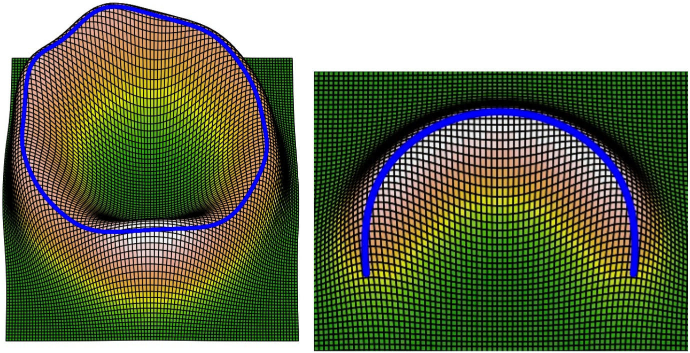
Density Ridges

We formalize the notion of filaments as *density ridges*.

Example: Ridges in Mountains

Credit: Google

Example: Ridges in Smooth Functions



Formal Definition of Density Ridges

- $p : \mathbb{R}^d \mapsto \mathbb{R}$, the density function.
- $(\lambda_j(x), v_j(x))$: jth eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.
- $V(x) = [v_2(x), \cdots, v_d(x)]$: matrix of the 2nd eigenvector to the last eigenvector.
- $V(x)V(x)^T$: a projection.
- Ridges:

$$R = \mathrm{Ridge}(p) = \{x : V(x)V(x)^T \nabla p(x) = 0, \lambda_2(x) < 0\}.$$

- Local modes:

$$\mathrm{Mode}(p) = \{x : \nabla p(x) = 0, \lambda_1(x) < 0\}.$$

Dimension of Ridges

The dimension of a ridge is 1.

This is because ridges are points satisfying $V(x)V(x)^T \nabla p(x) = 0$.

$V(x)V(x)^T$ has rank $d-1$, so there are $d-1$ effective constraints.

By the Implicit Function Theorem, ridges have dimension 1.

We use the plug-in estimate:

$$\widehat{R}_n = \text{Ridge}(\widehat{p}_n),$$

where $\widehat{p}_n(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$ is the kernel density estimator (KDE) and $X_1, \cdots, X_n$ are the locations of galaxies.

[1]Ozertem, Umut, and Deniz Erdogmus. "Locally defined principal curves and surfaces." JMLR (2011).

Estimator and Algorithm

We use the plug-in estimate:

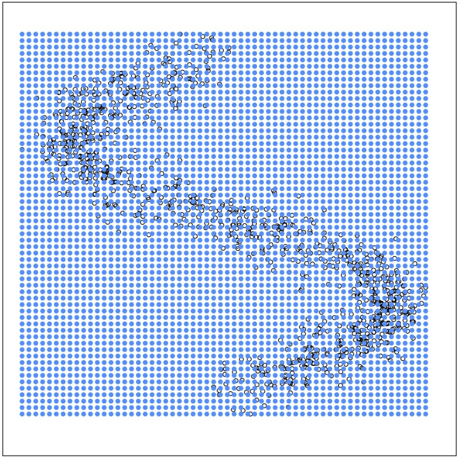
$$\widehat{R}_n = \text{Ridge}(\widehat{p}_n),$$

where $\widehat{p}_n(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$ is the kernel density estimator (KDE) and $X_1, \cdots, X_n$ are the locations of galaxies.

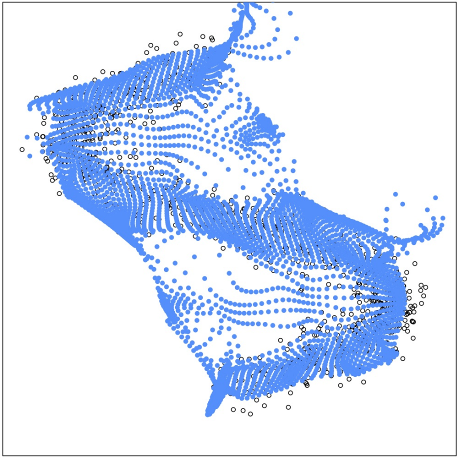
- In general, finding ridges from a given function is hard.
- The Subspace Constraint Mean Shift[1] (SCMS) algorithm allows us to find $\widehat{R}_n$, ridges of the KDE.

[1]Ozertem, Umut, and Deniz Erdogmus. "Locally defined principal curves and surfaces." JMLR (2011).

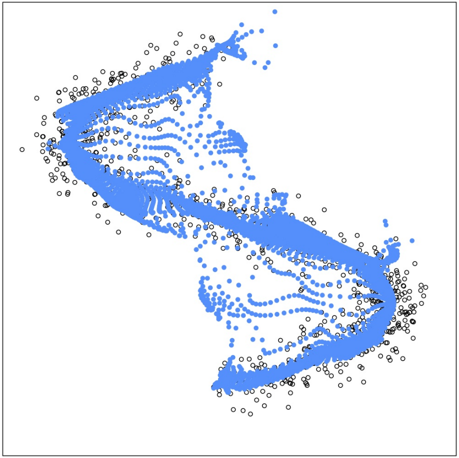
SCMS: Ridge Recovery Algorithm



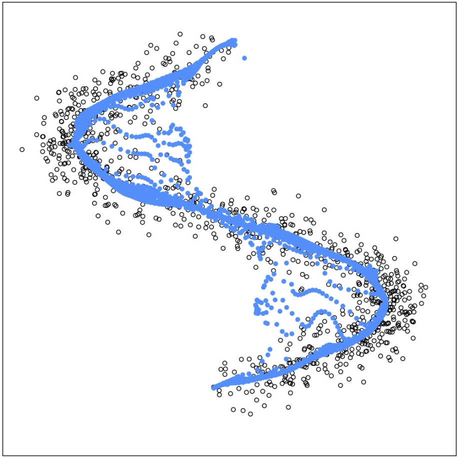
SCMS: Ridge Recovery Algorithm



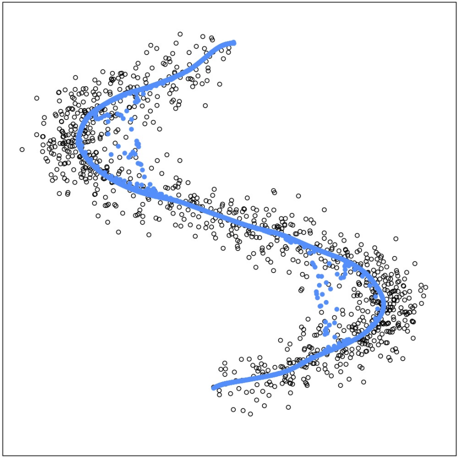
SCMS: Ridge Recovery Algorithm



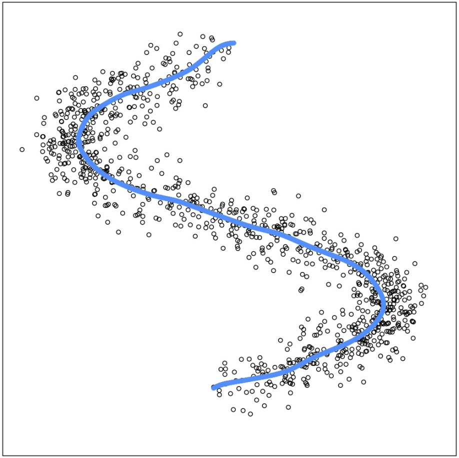
SCMS: Ridge Recovery Algorithm



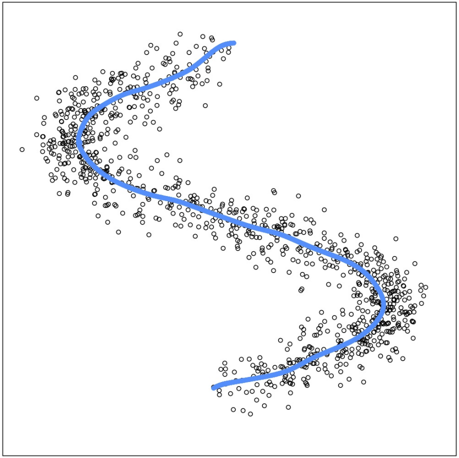
SCMS: Ridge Recovery Algorithm



SCMS: Ridge Recovery Algorithm



SCMS: Ridge Recovery Algorithm



SCMS moves blue mesh points by gradient ascent and a projection.

Formal definition of the SCMS algorithm

- Starting at an initial point $x^{(0)}$, the SCMS algorithm generates a sequence of points $x^{(1)}, x^{(2)}, \cdots$ via the following updating procedure:

  $$x^{(t+1)} = x^{(t)} + \eta \widehat{V}(x^{(t)}) \widehat{V}(x^{(t)})^T \nabla \widehat{p}_n(x^{(t)})$$

  for $t = 0, 1, 2, 3, \cdots$.
- The tuning parameter $\eta > 0$ is the step size.

Convergence of the SCMS algorithm

- Let $x^{(\infty)} \in \widehat{R}_n$ be its destination.

Theorem (Linear convergence of SCMS)

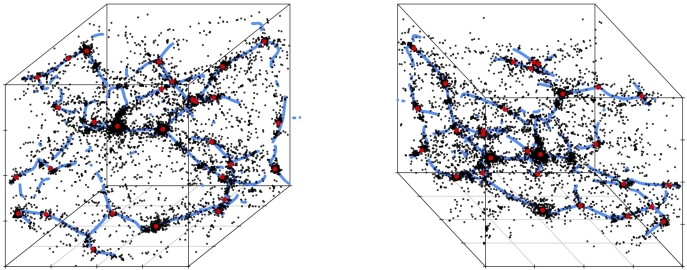
Under suitable conditions and $\|x^{(0)} - x^{(\infty)}\|_2 \leq r_0$, we have

$$\|x^{(t)} - x^{(\infty)}\|_2 \leq \Gamma^t \|x^{(0)} - x^{(\infty)}\|_2,$$

where $\Gamma \in (0, 1)$.

- We provide an explicit description of Γ, r_0 in our paper.
- Technical challenge: the projection matrix $\widehat{V}(x^{(t)})\widehat{V}(x^{(t)})^T$ also depends on the current location $x^{(t)}$, so we have to bound this difference as well.

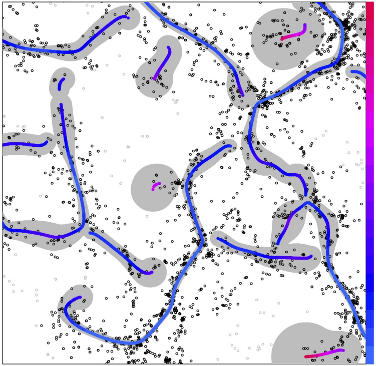
3D Example for Estimated Ridges



Blue curves: density ridges.

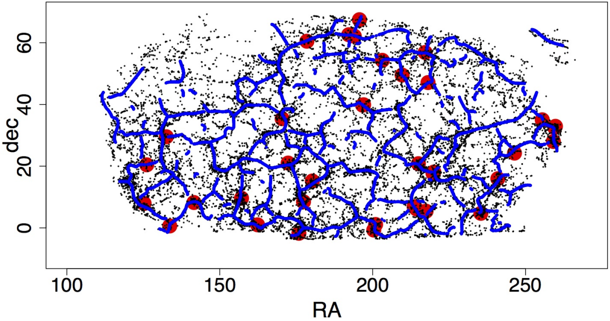
Red points: density local modes.

Uncertainty of Ridges from the Bootstrap

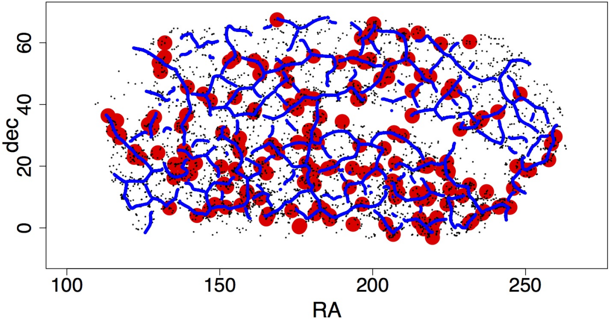


SDSS: Comparing to Clusters

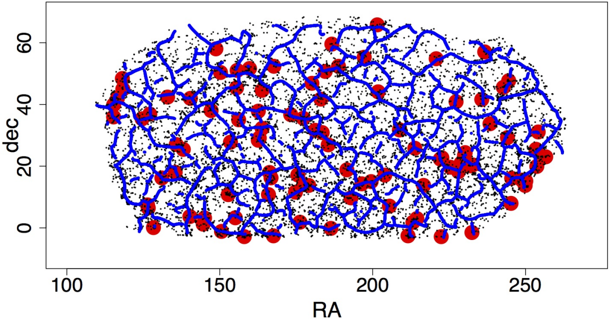
- Blue: filaments. Red: galaxy clusters (redMaPPer).



- Blue: filaments. Red: galaxy clusters (redMaPPer).



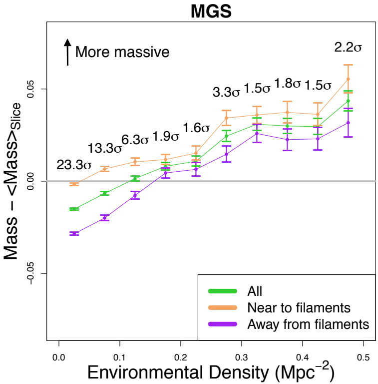
- Blue: filaments. Red: galaxy clusters (redMaPPer).



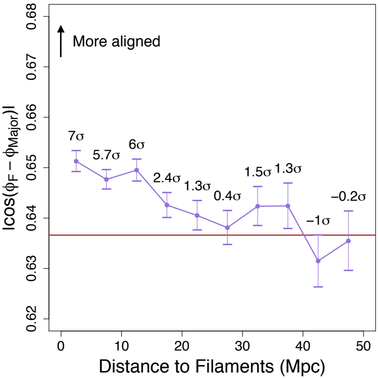
SDSS: Filament Effects VS Environments

Do filaments have an extra effect other than environments?

⟶Yes!



SDSS: Alignment

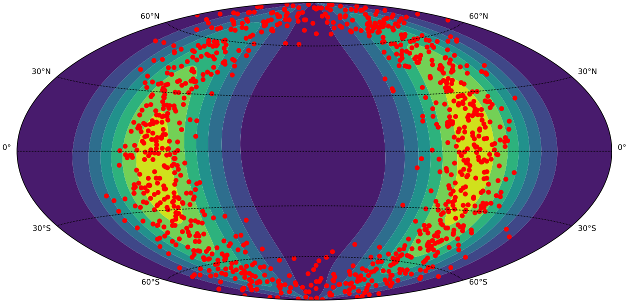


Accounting for the spherical geometry

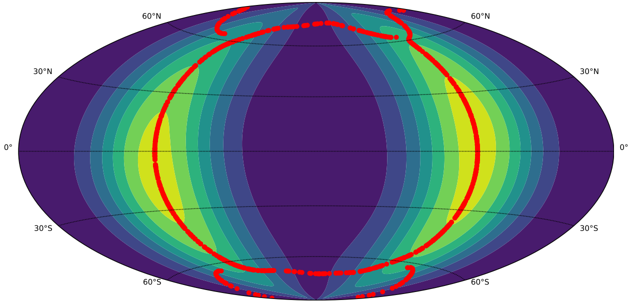
Directional data

- While the above results seem to be good, it has a severe problem: our data (locations of galaxies) is not in Euclidean space.
- In particular, we use (RA, dec) to represent the location of a galaxy.
- (RA, dec) are spherical coordinate!
- The Euclidean ridge finding algorithm may lead to a severe bias.

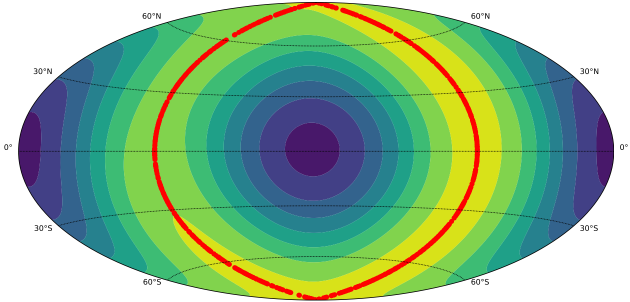
Failure of usual SCMS



Failure of usual SCMS



Failure of usual SCMS



Directional ridges - 1

- Let $X_1, \cdots, X_n \in \Omega_q$, where $\Omega_q = \{x \in \mathbb{R}^{q+1} : \|x\|_2 = 1\}$ be the directional data on q-dimensional sphere.
- To define ridges on Ω_q, we need to use gradient on a Riemannian manifold.
- Luckily, in this case, we have a simple representation of the gradient on Riemannian manifold grad using the usual gradient operator ∇ (in $(q+1)$-dimension):

$$\mathsf{grad} f(x) = (I_{q+1} - xx^T)\nabla f(x),$$

  where $x \in \mathbb{R}^{q+1}$ and $I_{q+1} = \mathsf{diag}(1, 1, \cdots, 1) \in \mathbb{R}^{(q+1)\times(q+1)}$.
- In the SDSS data, we convert (RA, dec) into a point $x \in \Omega_2 \subset \mathbb{R}^3$ such that $\|x\| = 1$.

Directional ridges - 2

- With the above representation, the Hessian on Riemannian manifold can be expressed as

$$\mathscr{H} f(x) = (I_{q+1} - xx^T)\nabla\nabla f(x)(I_{q+1} - xx^T)$$

when $x \in \Omega_q$.

- The directional ridges are then defined as

$$\underline{R} = \text{Ridge}(p) = \{x : \underline{V}(x)\underline{V}(x)^T \nabla p(x) = 0, \underline{\lambda}_2(x) < 0\},$$

where $\underline{V}(x)$ is the matrix of the smallest (q-1) eigenvectors and $\underline{\lambda}_2(x)$ is the second largest eigenvalue of $\mathscr{H} p(x)$.

Directional ridges - 3

- In practice, we estimate p by the directional KDE:

$$\widehat{p}_{\mathsf{dir}}(x) = \frac{c_{L,q}(h)}{n} \sum_{i=1}^{n} L\left(\frac{1 - x^T X_i}{h^2}\right),$$

where $c_{L,q}(h) = O(h^{-q})$ is the normalizing constant and L is the directional kernel.

- A popular choice is the von-Mises kernel, i.e., $L(r) = e^{-r}$.
- This leads to $\widehat{\mathcal{H}f}(x)$ and $\underline{\widehat{V}}(x)$ and $\underline{\widehat{\lambda}}_2(x)$ and $\underline{\widehat{R}}$.

Directional SCMS

- The SCMS algorithm can be generalized to a directional SCMS with some modifications.
- We showed that the directional SCMS can be expressed as the following fixed-point iteration (starting at $x^{(0)}$):

$$x^{(t+1)} = \frac{\underline{\widehat{V}}(x^{(t)})\underline{\widehat{V}}(x^{(t)})^T \nabla \widehat{p}_{\mathsf{dir}}(x^{(t)}) + \|\nabla \widehat{p}_{\mathsf{dir}}(x^{(t)})\|_2 \cdot x^{(t)}}{\left\|\underline{\widehat{V}}(x^{(t)})\underline{\widehat{V}}(x^{(t)})^T \nabla \widehat{p}_{\mathsf{dir}}(x^{(t)}) + \|\nabla \widehat{p}_{\mathsf{dir}}(x^{(t)})\|_2 \cdot x^{(t)}\right\|_2},$$

for $t = 0, 1, 2, 3, \cdots$.

Convergence of the directional SCMS algorithm

- Let $x^{(0)}$ be an initial point of the SCMS on Ω_q and let $x^{(\infty)} \in \underline{\widehat{R}}$ be its destination.

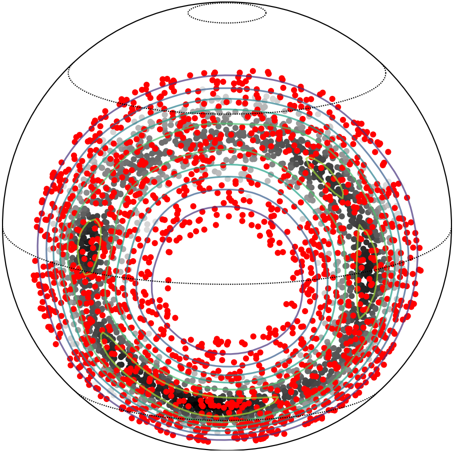
Theorem (Linear convergence of directional SCMS)

Under suitable conditions and $\|x^{(0)} - x^{(\infty)}\|_2 \leq r_{\mathsf{dir}}$, *we have*

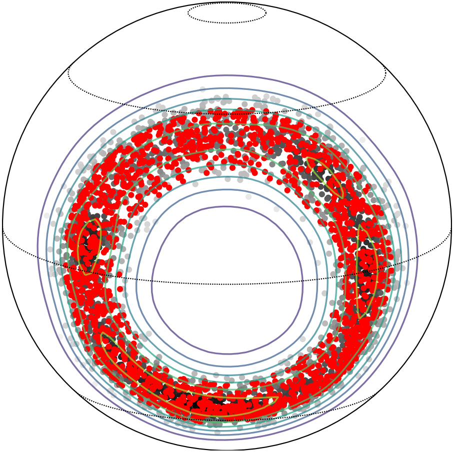
$$\|x^{(t)} - x^{(\infty)}\|_2 \leq \Gamma_{\mathsf{dir}}^t \|x^{(0)} - x^{(\infty)}\|_2,$$

where $\Gamma_{\mathsf{dir}} \in (0, 1)$.

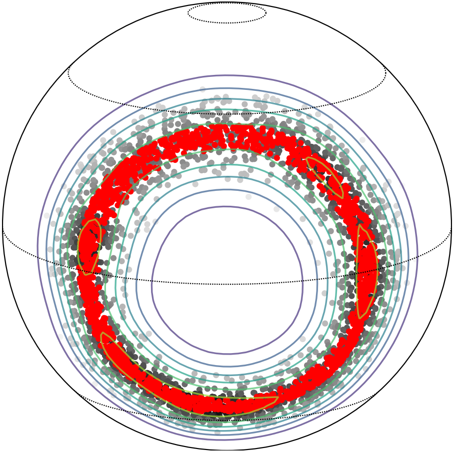
- We provide bounds on Γ_{dir} and r_{dir} in the paper.



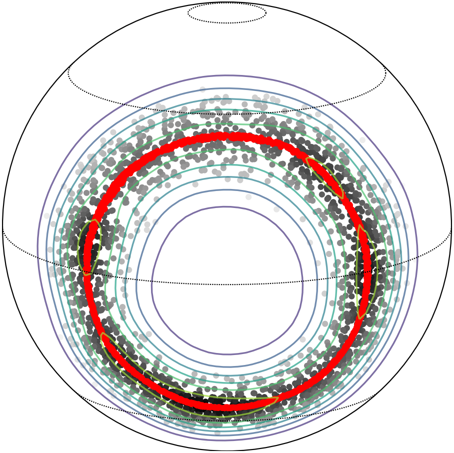
DirSCMS in action



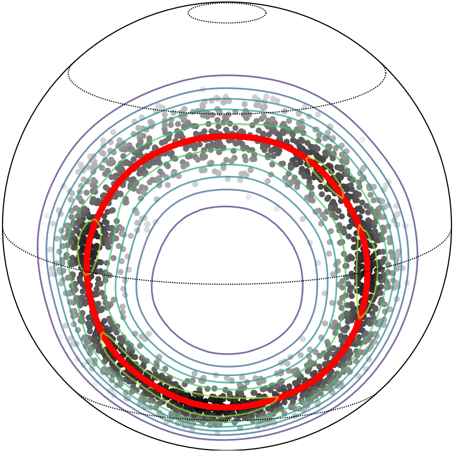
DirSCMS in action



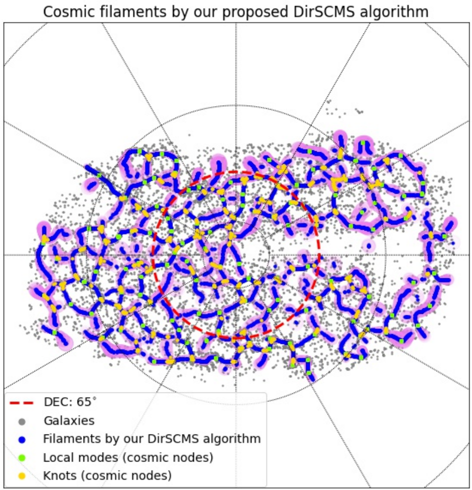
DirSCMS in action



DirSCMS in action

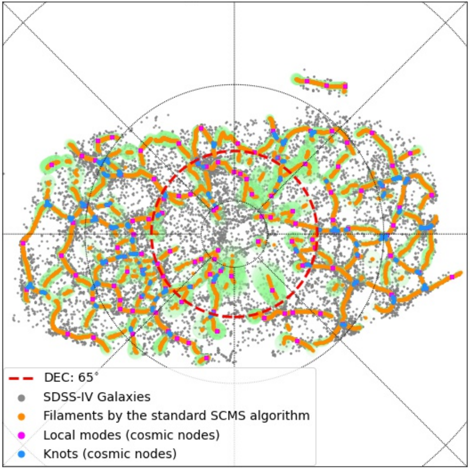


Applying to the SDSS data



Applying to the SDSS data

Cosmic filaments by the standard SCMS algorithm



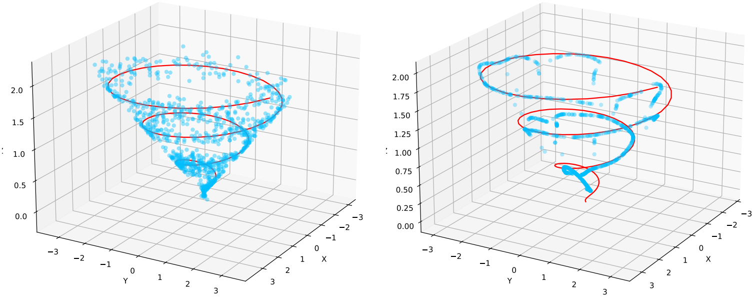
Incorporating the redshift

Incorporating the redshift information

- All the above approach is based on the idea of 'slicing the Universe'.
- Namely, we take slices based on redshift and find filaments in each slice.
- How to incorporate the information from redshift is a key problem.

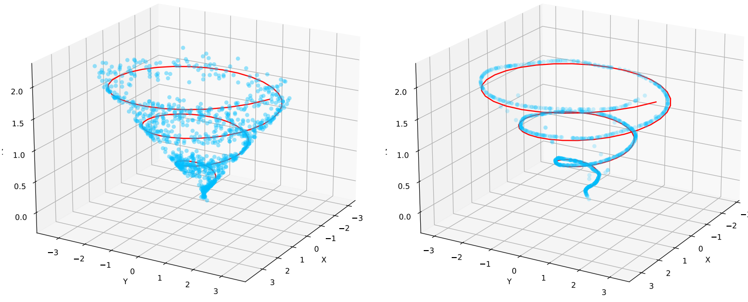
Failure of a naive idea

- Naively, one may think that we can convert (RA, dec, z) into 3-dimensional Cartesian coordinate and apply the 3D ridge finding algorithm.
- This idea may lead to unstable results. See the following simulation:



$\Omega_2 \times \mathbb{R}$ space

- To incorporate the redshift, we consider the product space $\Omega_2 \times \mathbb{R}$.
- Ω_2 is the 2-sphere, which describes the angular position (RA, dec).
- $\mathbb{R}$ is the 1-dimensional Euclidean space, which describes the redshift z.
- We attempt to find ridges in $\Omega_2 \times \mathbb{R}$.



- The right panel is the result from our directional-linear SCMS, which recover the true filament (red curve).

Density estimation in $\Omega_2 \times \mathbb{R}$

- The idea is to estimate the density in the product space directly.
- Let $x \in \Omega_2$ denotes the angular coordinate and $z \in \mathbb{R}$ denotes the redshift.
- Our data will be $(X_1, Z_1), \cdots, (X_n, Z_n) \in \Omega_2 \times \mathbb{R}$.
- We estimate the density using the product kernel:

$$\widehat{p}_{\mathsf{DL}}(x,z) = \frac{c_{L,2}(h_x)}{nh_z} \sum_{i=1}^{n} L\left(\frac{1 - x^T X_i}{h_x^2}\right) K\left(\frac{Z_i - z}{h_z}\right),$$

where $L(y) = e^{-y}$ and $K(y) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2)$ are a directional and Gaussian kernel.

Idea: Mean-Shift in $\Omega_2 \times \mathbb{R}$

- We show that a *gradient ascent* of $\widehat{p}_{\mathsf{DL}}(x, z)$ with a suitable step size can be written as follows.
- Starting at $x^{(t)}, z^{(t)}$, we compute

$$\widetilde{x}^{(t+1)} = \frac{\sum_{i=1}^{n} X_i L\left(\frac{1 - x^{(t)T} X_i}{h_x^2}\right) K\left(\frac{Z_i - z^{(t)}}{h_z}\right)}{\sum_{i=1}^{n} L\left(\frac{1 - x^{(t)T} X_i}{h_x^2}\right) K\left(\frac{Z_i - z^{(t)}}{h_z}\right)},$$

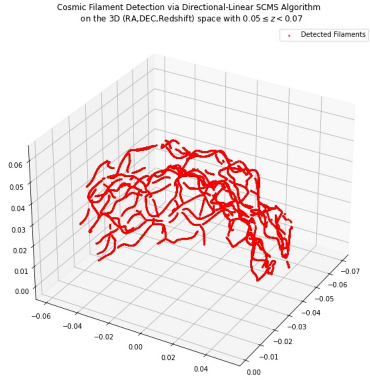
and update

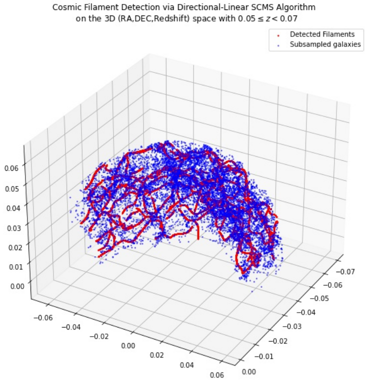
$$x^{(t+1)} = \frac{\widetilde{x}^{(t+1)}}{\|\widetilde{x}^{(t+1)}\|}.$$

Also, the location $z^{(t)}$ is updated to

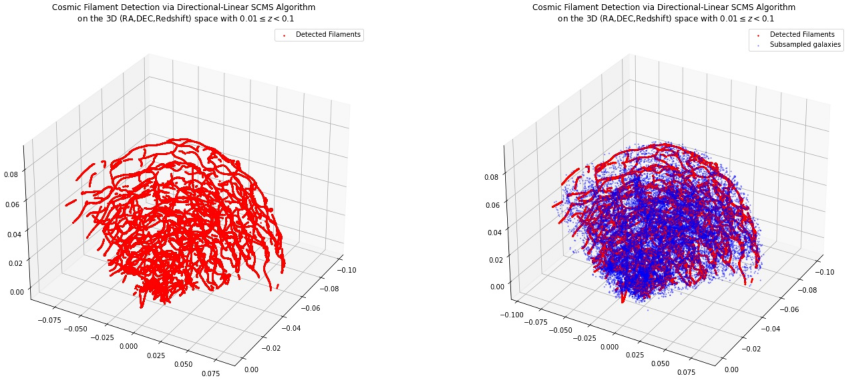
$$z^{(t+1)} = \frac{\sum_{i=1}^{n} Z_i L\left(\frac{1 - x^{(t+1)T} X_i}{h_x^2}\right) K\left(\frac{Z_i - z^{(t)}}{h_z}\right)}{\sum_{i=1}^{n} L\left(\frac{1 - x^{(t)T} X_i}{h_x^2}\right) K\left(\frac{Z_i - z^{(t)}}{h_z}\right)}.$$

Directional-Euclidean SCMS on SDSS





Directional-Euclidean SCMS on SDSS



Conclusion and future direction

- We have generalized the usual ridge finding problem into directional x Euclidean data, which is better suited for Astronomy data.
- We proved both statistical and computational learning theory of our algorithm.
- We have created python library for this algorithm:

  `https://pypi.org/project/sconce-scms/`

- The catalog and associated data can be found in:

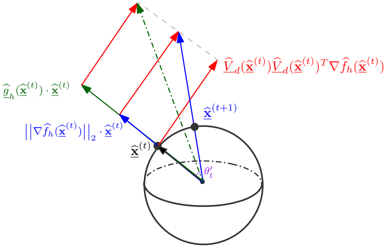
  `https://github.com/zhangyk8/sconce-scms/tree/main/examples/Theory_Method_Code`

Thank You!

More details can be found in
http://faculty.washington.edu/yenchic.

References

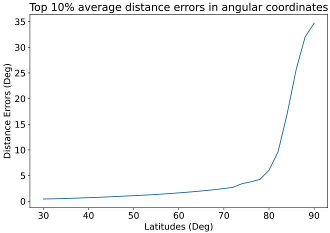
1. Zhang, Yikun, and Yen-Chi Chen. "Linear Convergence of the Subspace Constrained Mean Shift Algorithm: From Euclidean to Directional Data." arXiv preprint arXiv:2104.14977 (2021).
2. Zhang, Yikun, and Yen-Chi Chen. "Kernel smoothing, mean shift, and their learning theory with directional data." arXiv preprint arXiv:2010.13523 (2020) To appear in the Journal of Machine Learning Research.
3. Chen, Yen-Chi, et al. "Cosmic web reconstruction through density ridges: method and algorithm." Monthly Notices of the Royal Astronomical Society 454.1 (2015): 1140-1156.
4. Chen, Yen-Chi, Christopher R. Genovese, and Larry Wasserman. "Asymptotic theory for density ridges." The Annals of Statistics 43, no. 5 (2015): 1896-1928.
5. Conroy, Charlie, James E. Gunn, and Martin White. "The propagation of uncertainties in stellar population synthesis modeling. I. The relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies." The Astrophysical Journal 699.1 (2009): 486.
6. Eberly, David. Ridges in image and data analysis. Vol. 7. Springer Science & Business Media, 1996.
7. Ozertem, Umut, and Deniz Erdogmus. "Locally defined principal curves and surfaces." The Journal of Machine Learning Research 12 (2011): 1249-1286.



$$x^{(t+1)} = \frac{\underline{\widehat{V}}(x^{(t)})\underline{\widehat{V}}(x^{(t)})^T\nabla\widehat{p}_{\mathrm{dir}}(x^{(t)}) + \|\nabla\widehat{p}_{\mathrm{dir}}(x^{(t)})\|_2 \cdot x^{(t)}}{\|\underline{\widehat{V}}(x^{(t)})\underline{\widehat{V}}(x^{(t)})^T\nabla\widehat{p}_{\mathrm{dir}}(x^{(t)}) + \|\nabla\widehat{p}_{\mathrm{dir}}(x^{(t)})\|_2 \cdot x^{(t)}\|_2},$$

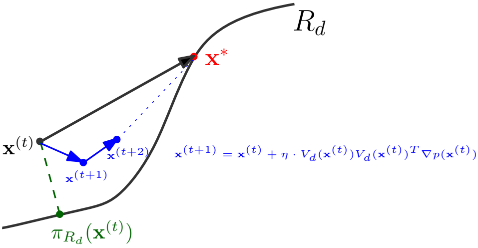
Note: $\widehat{p}_{\mathrm{dir}} = \widehat{f}_h$.

Comparison: Euclidean ridges vs directional ridges



We apply both Euclidean and directional ridge finding algorithms and study the errors of Euclidean ridges as a function of latitude.

Linear convergence: high-level idea

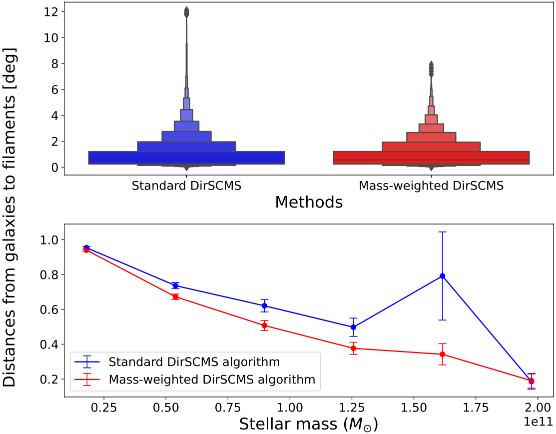


The projection matrix makes the algorithm not a conventional gradient ascent.

A key step to the proof is to bound the projection $(I_{q+1} - V_d(x^{(t)})V_d(x^{(t)})^T)(x^{(t)} - x^*)$ to be $O(\|x^{(t)} - x^*\|^2)$.

We use this decomposition to achieve that.

Weighted directional ridges: mass-distance



Errors of Euclidean method at different DEC