Cosmic Filament Detection Through Directional Density Ridges

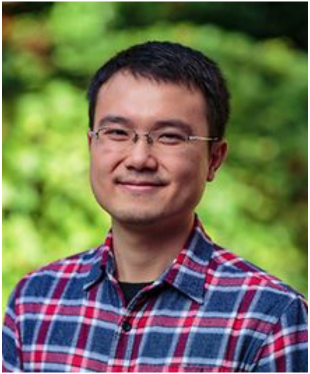
Yikun Zhang
(Joint work with *Yen-Chi Chen*)

Department of Statistics,
University of Washington

July 19, 2021 at Shanghai
Astronomical Observatory

Contributors

Yikun Zhang



Professor Yen-Chi Chen

Introduction

What Is Cosmic Web?

Cosmic Web is a large-scale network structure revealing that the matter in our Universe is not uniformly distributed.

What Is Cosmic Web?

Cosmic Web is a large-scale network structure revealing that the matter in our Universe is not uniformly distributed.

- It is caused by the anisotropic collapse of matter in gravitational instability scenarios at the early stage of the Universe (Zel'Dovich, 1970; Shandarin and Zeldovich, 1989; Bond et al., 1996).

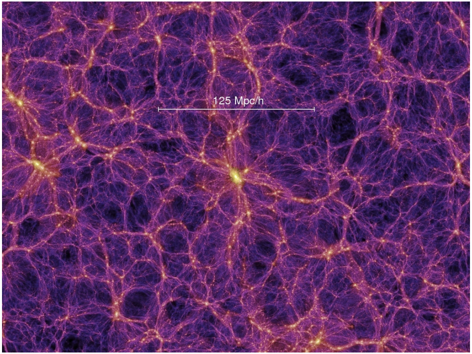


Figure 1: Visualization of *Cosmic Web* (credited to the millennium simulation project (Springel et al., 2005))

Characteristics of Cosmic Web

Cosmic web consists of four distinct components (Libeskind et al., 2018):

Characteristics of Cosmic Web

Cosmic web consists of four distinct components (Libeskind et al., 2018):

$\left.\begin{array}{l}\bullet \text{ Massive compact galaxy } \textit{clusters}, \\ \bullet \text{ Interconnected } \textit{filaments}, \\ \bullet \text{ Two-dimensional tenuous } \textit{sheets/walls}, \end{array}\right\}$ on which matter concentrates.

around • Vast and near-empty *voids*.

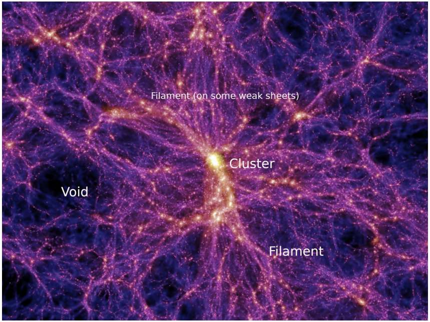


Figure 2: Characteristics of *Cosmic Web* (credited to the millennium simulation project (Springel et al., 2005))

Significance of Cosmic Filaments (I)

Why do we focus on cosmic filaments in this talk?

Significance of Cosmic Filaments (I)

Why do we focus on cosmic filaments in this talk?

- They connect complexes of super-clusters (Lynden-Bell et al., 1988).

Why do we focus on cosmic filaments in this talk?

- They connect complexes of super-clusters (Lynden-Bell et al., 1988).
- They contain information about the global cosmology and the nature of dark matter (Zhang et al., 2009; Tempel et al., 2014).

Significance of Cosmic Filaments (I)

Why do we focus on cosmic filaments in this talk?

- They connect complexes of super-clusters (Lynden-Bell et al., 1988).
- They contain information about the global cosmology and the nature of dark matter (Zhang et al., 2009; Tempel et al., 2014).
- Some properties of nearby galaxies, such as stellar masses, intrinsic alignments, and luminosity, are influenced by cosmic filaments (Zhang et al., 2013; Clampitt et al., 2016; Poudel et al., 2017; Chen et al., 2017).

Significance of Cosmic Filaments (I)

Why do we focus on cosmic filaments in this talk?

- They connect complexes of super-clusters (Lynden-Bell et al., 1988).
- They contain information about the global cosmology and the nature of dark matter (Zhang et al., 2009; Tempel et al., 2014).
- Some properties of nearby galaxies, such as stellar masses, intrinsic alignments, and luminosity, are influenced by cosmic filaments (Zhang et al., 2013; Clampitt et al., 2016; Poudel et al., 2017; Chen et al., 2017).
- Cosmic filaments also serve as carriers of quiescent galaxies and hot gas (emitting in radio) (Bonjean et al., 2018; Govoni et al., 2019).
- ...

Significance of Cosmic Filaments (II)

Further, the trajectory of cosmic microwave background (CMB) light is shown to be distorted due to cosmic filaments, creating an effect known as weak lensing.

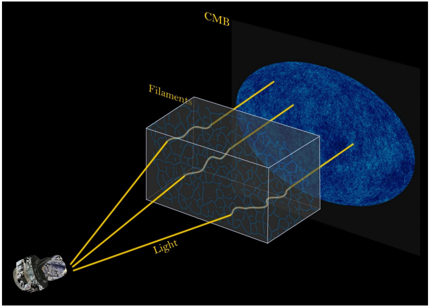


Figure 3: Illustration of the bending trajectory of CMB lights (credit to Siyu He, Shadab Alam, Wei Chen, and Planck/ESA; see He et al. (2018) for details)

Challenges in Detecting Cosmic Filaments

- The filamentary structures are overwhelmingly complex (Cautun et al., 2013):
  - Lack of structural symmetries.
  - How to measure its connectivity.
  - Intrinsic multi-scale nature.
  - ...

Challenges in Detecting Cosmic Filaments

- The filamentary structures are overwhelmingly complex (Cautun et al., 2013):
  - Lack of structural symmetries.
  - How to measure its connectivity.
  - Intrinsic multi-scale nature.
  - ...
- There exist no universal and mathematically rigorous definitions about cosmic filaments!

1. A brief review on existing methods in filament detection.

1. A brief review on existing methods in filament detection.
2. Detailed discussions on our methodology.

Highlights of the Today's Talk

1. A brief review on existing methods in filament detection.
2. Detailed discussions on our methodology.
   - Formulate cosmic filaments as the directional density ridges of the underlying galaxy density function.
   - Estimate the directional density ridges via our *Directional Subspace Constrained Mean Shift* (DirSCMS) algorithm.

1. A brief review on existing methods in filament detection.

2. Detailed discussions on our methodology.
   - Formulate cosmic filaments as the directional density ridges of the underlying galaxy density function.
   - Estimate the directional density ridges via our *Directional Subspace Constrained Mean Shift* (DirSCMS) algorithm.

3. Applications on SDSS-IV galaxy data (Ahumada et al., 2020) with some further analysis.

Previous Works in Filament Detection

The 3D methods identify filaments within a cubic section of the Universe or in some N-body simulations; see Cautun et al. (2014) for a complete literature review.

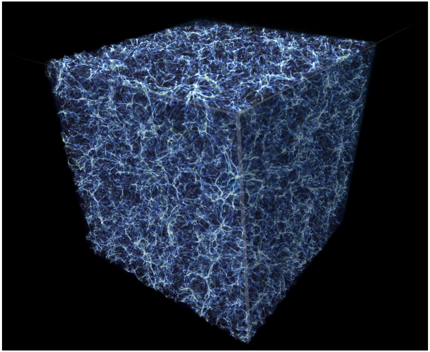


Figure 4: Matter distribution in a cubic section of the Universe (credit to NASA, ESA, and E. Hallman at University of Colorado, Boulder)

Existing Research on Filament Detection
(3D Methods with Survey Data)

In astronomical survey data, such as SDSS or the Dark Energy Survey, the positions of galaxies/particles are recorded as

$$\{(z_1, \phi_1, \eta_1), ..., (z_n, \phi_n, \eta_n)\},$$

where, for $i = 1, ..., n$,

Existing Research on Filament Detection
(3D Methods with Survey Data)

In astronomical survey data, such as SDSS or the Dark Energy Survey, the positions of galaxies/particles are recorded as

$$\{(z_1, \phi_1, \eta_1), ..., (z_n, \phi_n, \eta_n)\},$$

where, for $i = 1, ..., n$,

- $z_i \in (0, \infty)$ is the *redshift* value,
- $\phi_i \in [0, 2\pi)$ is the *right ascension* (RA), i.e., celestial longitude,
- $\eta_i \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ is the *declination* (DEC), i.e., celestial latitude.

Existing Research on Filament Detection
(3D Methods with Survey Data)

In astronomical survey data, such as SDSS or the Dark Energy Survey, the positions of galaxies/particles are recorded as

$$\{(z_1, \phi_1, \eta_1), ..., (z_n, \phi_n, \eta_n)\},$$

where, for $i = 1, ..., n$,

- $z_i \in (0, \infty)$ is the *redshift* value,
- $\phi_i \in [0, 2\pi)$ is the *right ascension* (RA), i.e., celestial longitude,
- $\eta_i \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ is the *declination* (DEC), i.e., celestial latitude.

One may convert angular coordinates to Cartesian coordinates as

$$\begin{aligned} X_i &= d(z_i) \cos\phi_i \cos\eta_i, \\ Y_i &= d(z_i) \sin\phi_i \cos\eta_i, \\ Z_i &= d(z_i) \sin\eta_i, \end{aligned}$$

where $d(\cdot)$ is a distance transforming function; see Tempel et al. (2014) for details.

There are some potential drawbacks of detecting filaments with survey data in the 3D space:

There are some potential drawbacks of detecting filaments with survey data in the 3D space:

- The determination of $d(\cdot)$ may rely on some complicated cosmological models.

There are some potential drawbacks of detecting filaments with survey data in the 3D space:

- The determination of $d(\cdot)$ may rely on some complicated cosmological models.
- The galaxy distribution is elongated along the line of sight due to the peculiar velocities of galaxies (i.e., the so-called *finger-of-god* effect).

Existing Research on Filament Detection
(Drawbacks of 3D Methods)

There are some potential drawbacks of detecting filaments with survey data in the 3D space:

- The determination of $d(\cdot)$ may rely on some complicated cosmological models.
- The galaxy distribution is elongated along the line of sight due to the peculiar velocities of galaxies (i.e., the so-called *finger-of-god* effect).
- The number of galaxies varies dramatically across different redshift values, so applying 3D approaches will be computationally intensive.

Goal: We hope to transform the filament detection problem with survey data from the 3D space to the 2D space.

Slicing the Universe

Goal: We hope to transform the filament detection problem with survey data from the 3D space to the 2D space.

Solution: we partition the range of redshift values into several small intervals with $\Delta z = 0.005$, namely, slices of the Universe.

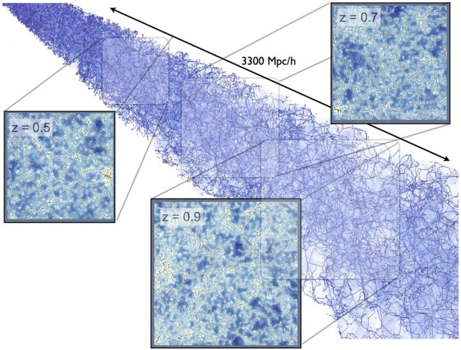


Figure 5: Illustration of slicing the Universe (credit to Laigle et al. 2018)

With each slice, says $z = 0.470$–0.475,

- the redshift values of galaxies/particles are thought to be identical.
- the locations of galaxies/particles are given by their (RA, DEC).

Existing Research on Filament Detection
(2D Methods with Survey Data)

With each slice, says $z = 0.470$–0.475,

- the redshift values of galaxies/particles are thought to be identical.
- the locations of galaxies/particles are given by their (RA, DEC).

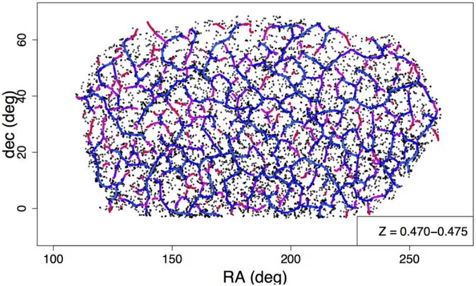


Figure 6: Cosmic filaments via density ridges on a 2D slice (Chen et al., 2015b, 2016)

[1] Y.-C. Chen, S. Ho, P. E. Freeman, C. R. Genovese, and L. Wasserman. Cosmic web reconstruction through density ridges: method and algorithm. *Monthly Notices of the Royal Astronomical Society*, 454(1):1140–1156, 2015.

[2] Y.-C. Chen, S. Ho, J. Brinkmann, P. E. Freeman, C. R. Genovese, D. P. Schneider, and L. Wasserman. Cosmic web reconstruction through density ridges: catalogue. *Monthly Notices of the Royal Astronomical Society*, 461(4):3896–3909, 2016.

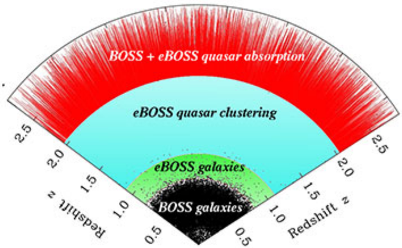
Problems of Existing Methods on 2D Slices

The slices ($\Delta z = 0.005$) in our observational studies are not some flat 2D planes, but some **spherical shells**!

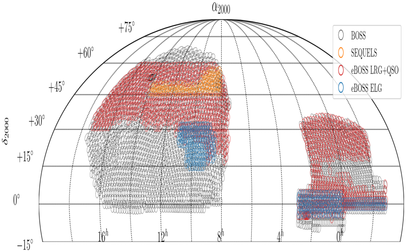
The slices ($\Delta z = 0.005$) in our observational studies are not some flat 2D planes, but some **spherical shells**!

In other words, the galaxies/particles in each slice indeed lie on (the surface of) a sphere with *nonlinear* curvature.

- Recall that the locations of galaxies/particles are recorded by $\{(z_i, \phi_i, \eta_i)\}_{i=1}^n$.



(a) Planned eBOSS coverage of the Universe (credit to M. Blanton and SDSS)



(b) BOSS/eBOSS Spectroscopic Footprint as of DR16 (credit to SDSS)

Why can't we ignore the spherical geometry? (I)

Setup: Suppose that we want to recover the true ring/filament structure across the North and South pole of a unit sphere given some noisy data points from it.

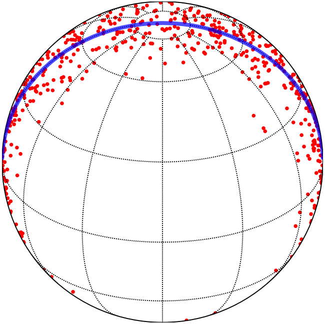


Figure 8: Noisy observations (red points) and the underlying true ring/filament structure (blue line)

Why can't we ignore the spherical geometry? (II)

Methods:

- We encode those data points with their angular coordinates on a **flat** rectangle plane $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times [0, 2\pi)$, and recover the ring/filament structure using the regular SCMS algorithm (Ozertem and Erdogmus, 2011).

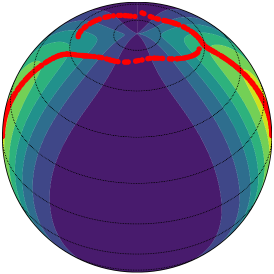
Or

- We consider those data points on the unit sphere $\Omega_2 = \left\{\boldsymbol{x} \in \mathbb{R}^3 : ||\boldsymbol{x}||_2 = 1\right\}$, and recover the ring/filament structure using DirSCMS algorithm (Zhang and Chen, 2021b).

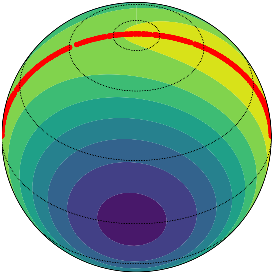
We will discuss more on the directional subspace constrained mean shift (DirSCMS) algorithm soon.

Why can't we ignore the spherical geometry? (III)

The background contour plots are kernel density estimators on the flat plane $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times [0, 2\pi)$ and unit sphere Ω_2, respectively.

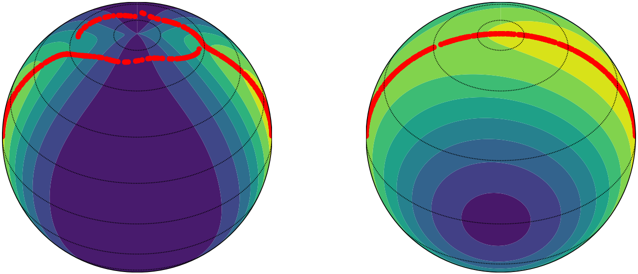


(a) Method 1: converged points after the regular SCMS algorithm



(b) Method 2: converged points after our DirSCMS algorithm

The background contour plots are kernel density estimators on the flat plane $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times [0, 2\pi)$ and unit sphere Ω_2, respectively.



(a) Method 1: converged points after the regular SCMS algorithm

(b) Method 2: converged points after our DirSCMS algorithm

Conclusion: We should not ignore the spherical geometry when detecting the filaments on each (redshift) slice of the Universe.

Methodology
(Directional Density Ridges and Subspace Constrained Mean Shift Algorithm)

Directional Density Ridges (I)

We formulate the cosmic filaments as *directional density ridges* of the underlying galaxy density function f on Ω_2.

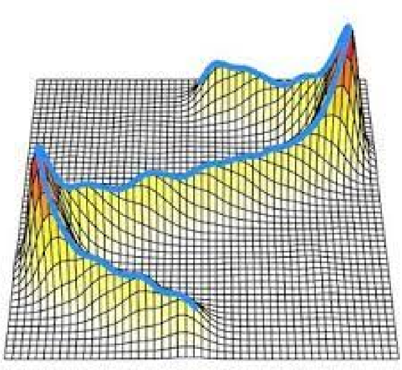


Figure 10: Density ridge (lifted onto the underlying density function) (credit to Yen-Chi Chen)

Directional Density Ridges (II)

More formally, (directional) density ridges are generalized local maxima (within some subspaces) of the underlying density function (on Ω_q).

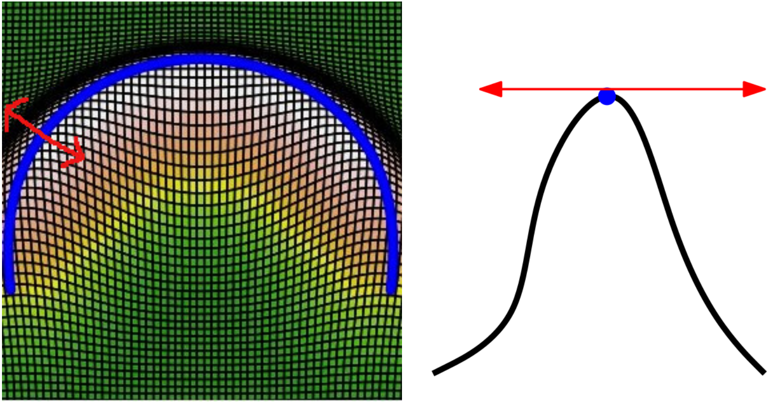


Figure 11: Density ridge (lifted onto the underlying density function; Chen et al. 2015a)

Question: How can we recover the directional density ridges (or equivalently, cosmic filaments) from some discrete galaxy observations?

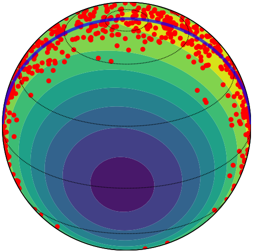


Figure 12: Counter plot of directional KDE

Question: How can we recover the directional density ridges (or equivalently, cosmic filaments) from some discrete galaxy observations?

Solution: (Step 1) We estimate the galaxy distribution via the directional kernel density estimator (KDE; Hall et al. 1987; Bai et al. 1988; García-Portugués 2013).

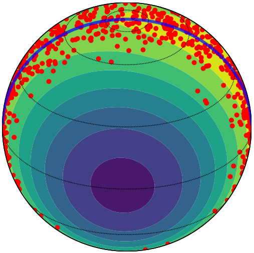


Figure 12: Counter plot of directional KDE

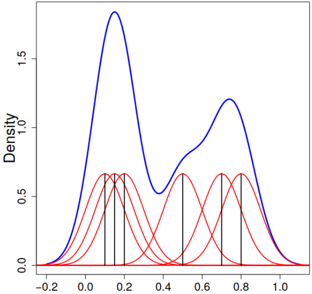
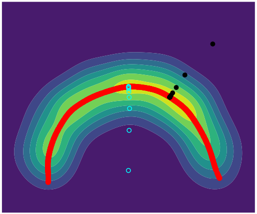


Figure 13: Illustration of one-dimensional KDE (Chen, 2017)

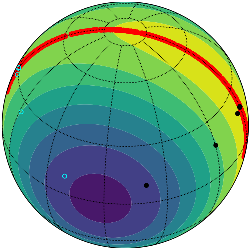
DirSCMS Algorithm

Question: How can we recover the directional density ridges (or equivalently, cosmic filaments) from some discrete galaxy observations?

Solution: (Step 2) We propose the directional subspace constrained mean shift (DirSCMS) algorithm (Zhang and Chen, 2021b), which iterates a point on Ω_2 along the (subspace constrained) *gradient* of directional KDE.



(a) Two regular SCMS iterative paths



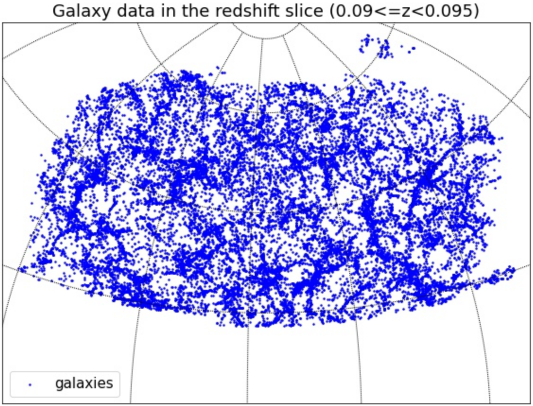
(b) Two DirSCMS iterative paths

Applications on SDSS-IV Galaxy Data

Step 1 (Slicing the Universe): Partition the redshift range $0.05 \leq z < 0.7$ into 130 spherical slices, each of which has the width $\Delta z = 0.005$.

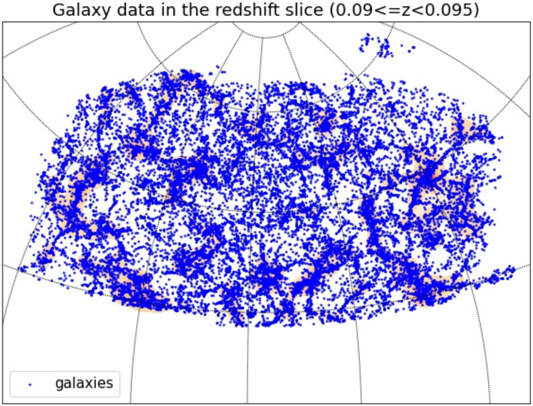
Step 1 (Slicing the Universe): Partition the redshift range $0.05 \leq z < 0.7$ into 130 spherical slices, each of which has the width $\Delta z = 0.005$.

- Within each slice, we consider the redshifts of galaxies to be the same so that the galaxies are located on a sphere.



Step 2 (Density Estimation): Estimate the galaxy distribution via directional KDE.

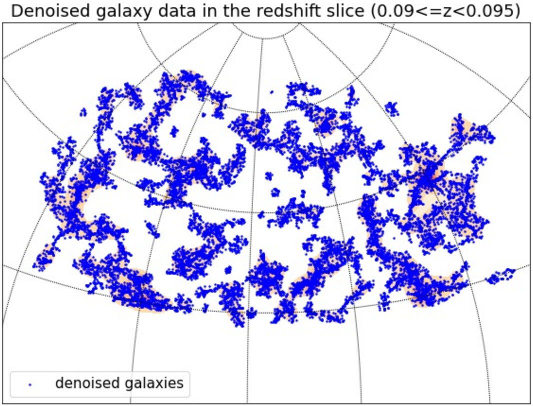
- The bandwidth parameter is selected in a data-adaptive approach.



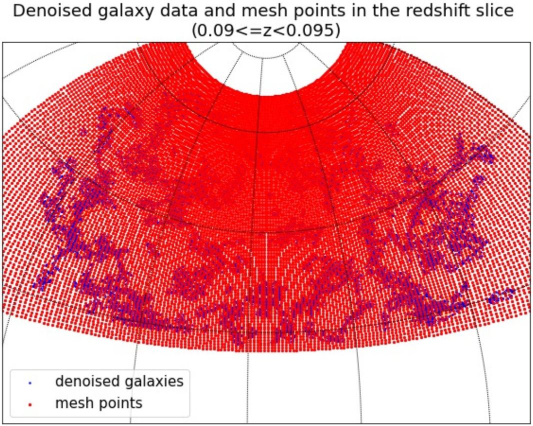
Step 3 (Denoising): Remove the galaxies with low-density values.

Step 3 (Denoising): Remove the galaxies with low-density values.

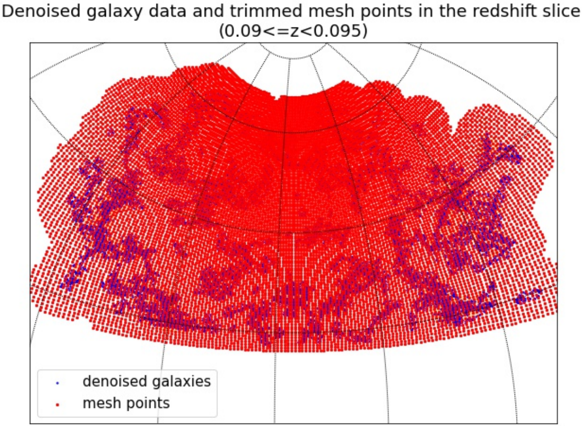
- We keep at least 80% of the original galaxy data in the slice.



Step 4 (Laying Down the Mesh Points): We place a set of dense mesh points on the interested region, which are the initial points of our DirSCMS iterations.



Step 5 (Thresholding the Mesh Points): We discard those mesh points with low-density values and keep 85% of the original mesh points.



Step 6 (DirSCMS Iterations): We iterate our DirSCMS algorithm on each remaining mesh point until convergence.

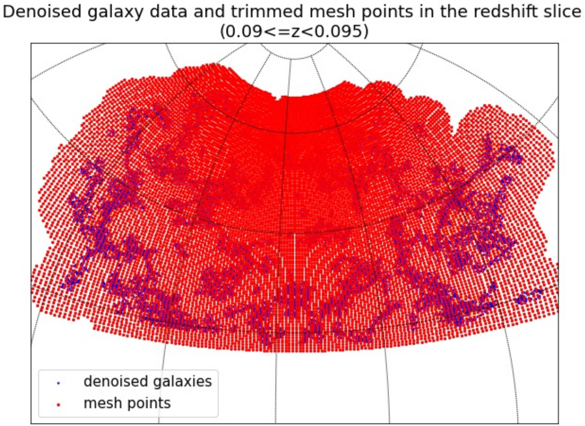


Figure 15: DirSCMS Iterations (Step 0)

Step 6 (DirSCMS Iterations): We iterate our DirSCMS algorithm on each remaining mesh point until convergence.

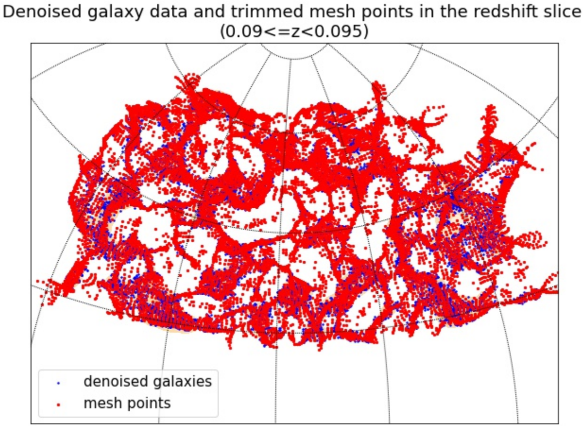


Figure 15: DirSCMS Iterations (Step 1)

Step 6 (DirSCMS Iterations): We iterate our DirSCMS algorithm on each remaining mesh point until convergence.

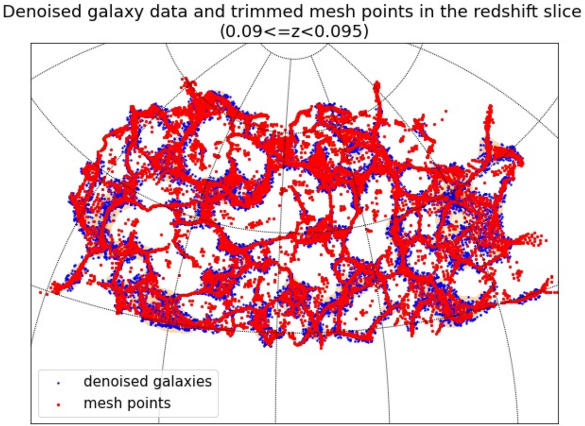


Figure 15: DirSCMS Iterations (Step 2)

Step 6 (DirSCMS Iterations): We iterate our DirSCMS algorithm on each remaining mesh point until convergence.

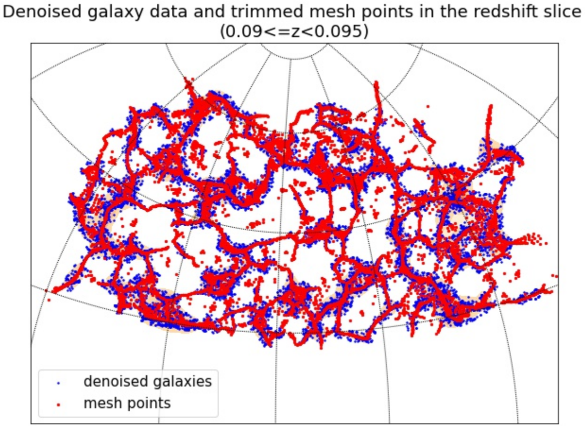


Figure 15: DirSCMS Iterations (Step 3)

Step 6 (DirSCMS Iterations): We iterate our DirSCMS algorithm on each remaining mesh point until convergence.

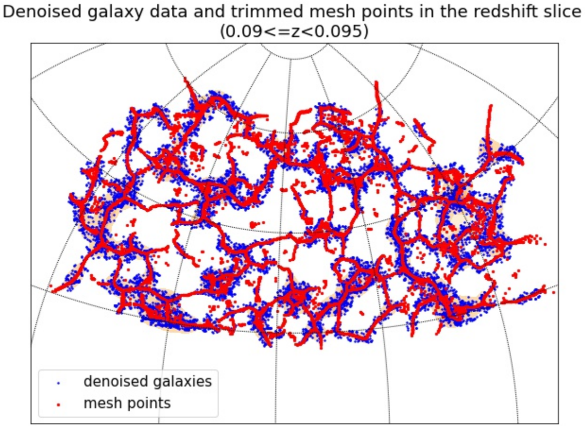


Figure 15: DirSCMS Iterations (Step 5)

Step 6 (DirSCMS Iterations): We iterate our DirSCMS algorithm on each remaining mesh point until convergence.

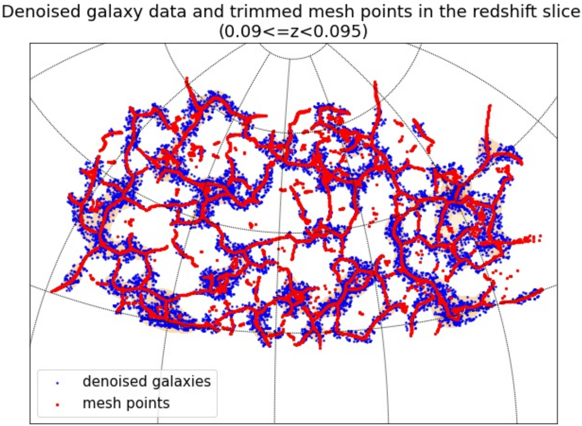


Figure 15: DirSCMS Iterations (Step 8)

Step 6 (DirSCMS Iterations): We iterate our DirSCMS algorithm on each remaining mesh point until convergence.

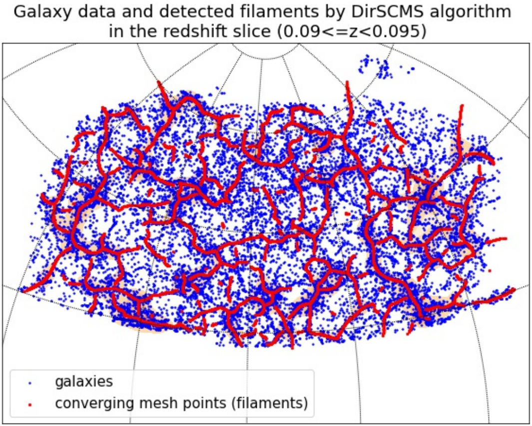


Figure 15: DirSCMS Iterations (Final)

Filament Effects on Galaxy Properties

1. We compute the angular distance (or equivalently, *geodesic distance*) of each observed galaxy in the redshift range $0.05 \leq z < 0.7$ to our detected filaments in the corresponding slice.

Filament Effects on Galaxy Properties

1. We compute the angular distance (or equivalently, *geodesic distance*) of each observed galaxy in the redshift range $0.05 \leq z < 0.7$ to our detected filaments in the corresponding slice.

2. We obtain the galaxy properties, such as stellar mass and metallicity, from the FIREFLY value-added catalog (Wilkinson et al., 2017; Maraston and Strömbäck, 2011).

Filament Effects on Galaxy Properties

1. We compute the angular distance (or equivalently, *geodesic distance*) of each observed galaxy in the redshift range $0.05 \leq z < 0.7$ to our detected filaments in the corresponding slice.

2. We obtain the galaxy properties, such as stellar mass and metallicity, from the FIREFLY value-added catalog (Wilkinson et al., 2017; Maraston and Strömbäck, 2011).

3. Our subsequent analyses focus on the following three regions:
   - **Low redshift region**: $0.05 \leq z < 0.07$.
   - **Medium redshift region**: $0.25 \leq z < 0.27$.
   - **High redshift region**: $0.55 \leq z < 0.57$.

Filament Effects on Galaxy Properties

1. We compute the angular distance (or equivalently, *geodesic distance*) of each observed galaxy in the redshift range $0.05 \leq z < 0.7$ to our detected filaments in the corresponding slice.

2. We obtain the galaxy properties, such as stellar mass and metallicity, from the FIREFLY value-added catalog (Wilkinson et al., 2017; Maraston and Strömbäck, 2011).

3. Our subsequent analyses focus on the following three regions:
   - **Low redshift region**: $0.05 \leq z < 0.07$.
   - **Medium redshift region**: $0.25 \leq z < 0.27$.
   - **High redshift region**: $0.55 \leq z < 0.57$.

4. We partition the galaxies within each region into several bins according to their distances to our detected filaments.

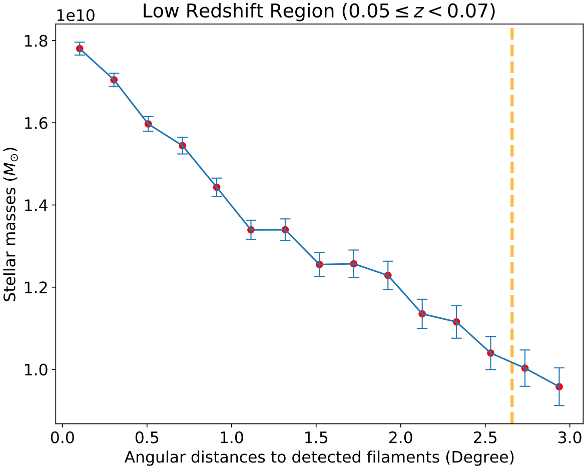


Figure 16: Comparison between stellar masses of galaxies and their distances to filaments (**Low redshift region**)

Analysis: Stellar Mass and Distance to Filaments

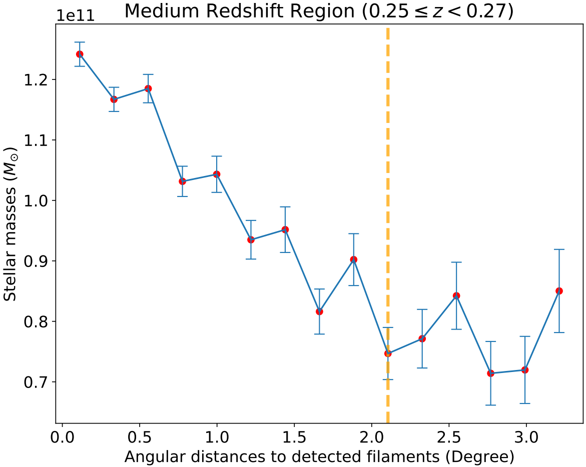


Figure 16: Comparison between stellar masses of galaxies and their distances to filaments (**Medium redshift region**)

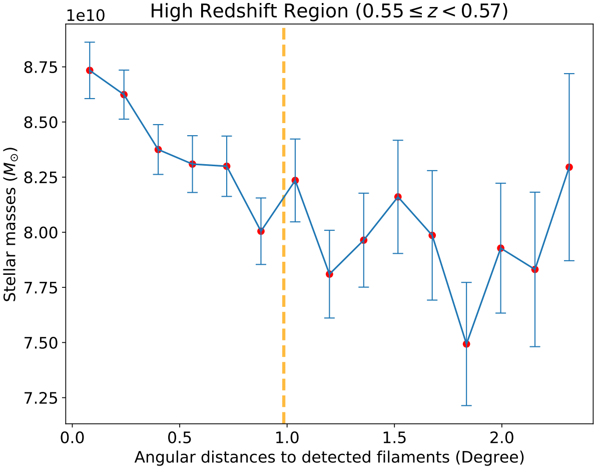


Figure 16: Comparison between stellar masses of galaxies and their distances to filaments (**High redshift region**)

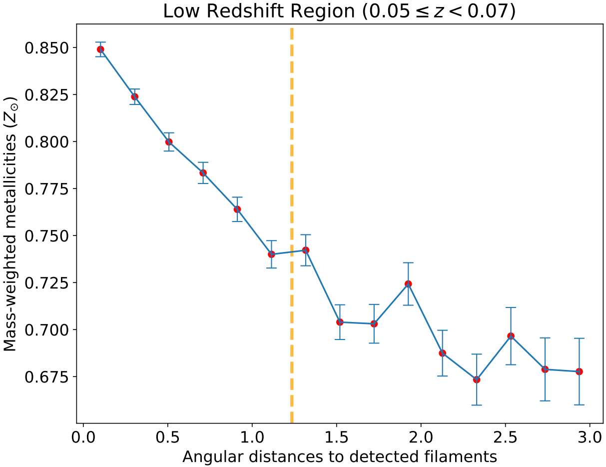


Figure 17: Comparison between (mass-weighted) metallicities of galaxies and their distances to filaments (**Low redshift region**)

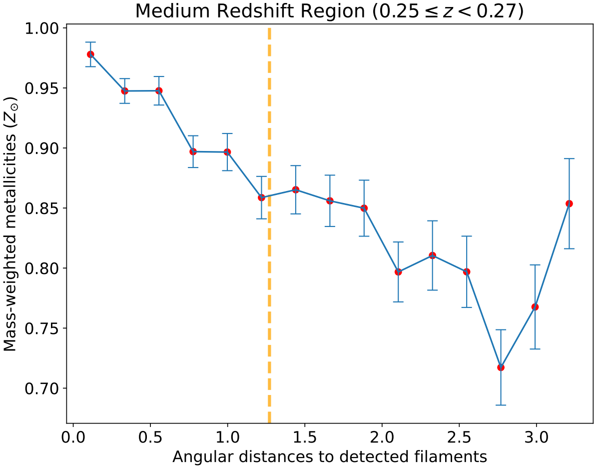


Figure 17: Comparison between (mass-weighted) metallicities of galaxies and their distances to filaments (**Medium redshift region**)

Analysis: Metallicity and Distance to Filaments

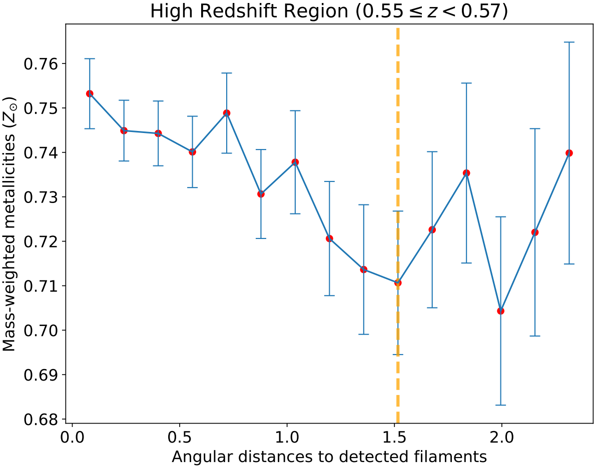


Figure 17: Comparison between (mass-weighted) metallicities of galaxies and their distances to filaments (**High redshift region**)

Conclusions and Ongoing Works

Summary of Today's Talk

In this talk, we discussed our methodology of recovering filament structures from some discrete observations (on the sphere).

Summary of Today's Talk

In this talk, we discussed our methodology of recovering filament structures from some discrete observations (on the sphere).

1. Our DirSCMS algorithm took into account the spherical geometry of the slice on which the galaxies/particles are located.

Summary of Today's Talk

In this talk, we discussed our methodology of recovering filament structures from some discrete observations (on the sphere).

1. Our DirSCMS algorithm took into account the spherical geometry of the slice on which the galaxies/particles are located.

2. We applied our method to the latest galaxy survey data (SDSS-IV, Data Release 16).

Summary of Today's Talk

In this talk, we discussed our methodology of recovering filament structures from some discrete observations (on the sphere).

1. Our DirSCMS algorithm took into account the spherical geometry of the slice on which the galaxies/particles are located.
2. We applied our method to the latest galaxy survey data (SDSS-IV, Data Release 16).
3. Our analyses reveal some signals that galaxies near the filaments are heavier in their stellar masses and richer in their metallicities.

The application of our filament detection method is not limited to galaxy data. Other potential playgrounds include

- SDSS Quasar survey data,
- Dark Energy Survey (cosmic trough identification; Moews et al. 2021),
- hydrodynamical cosmological simulation data (Chen et al., 2015c),
- ...

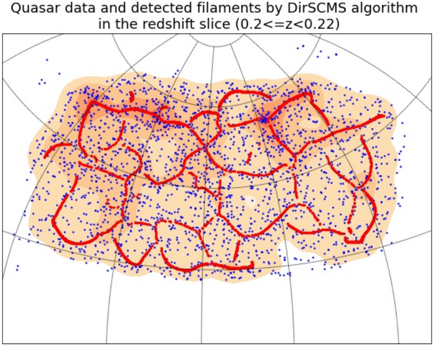
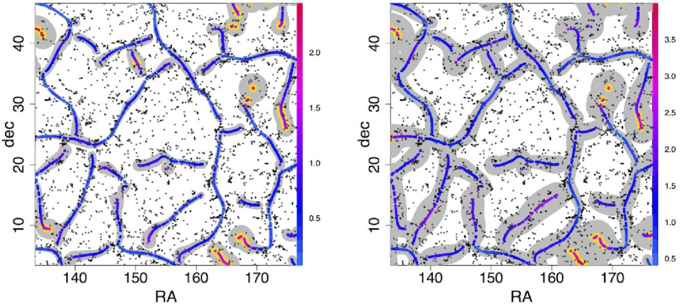


Figure 18: An example of the Quasar filaments

One merit of recovering cosmic filaments via SCMS/DirSCMS algorithm is that such procedure naturally produces uncertain measures for the yielded filaments (Chen et al., 2015a,b).

One merit of recovering cosmic filaments via SCMS/DirSCMS algorithm is that such procedure naturally produces uncertain measures for the yielded filaments (Chen et al., 2015a,b).

- Notes: Smooth bootstrap on Ω_2 can be done in an efficient way! (Ulrich, 1984; Wood, 1994)



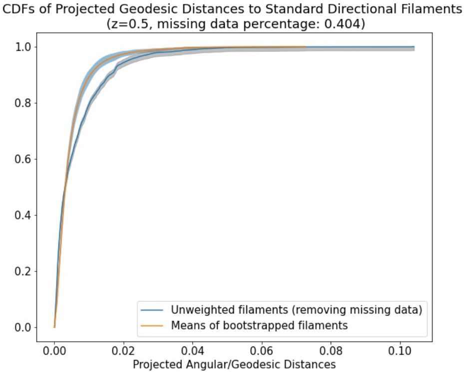
(a) The bootstrap estimate. (b) The smooth bootstrap estimate.

Figure 19: Local uncertainty estimates for detected filaments (Chen et al., 2015b)

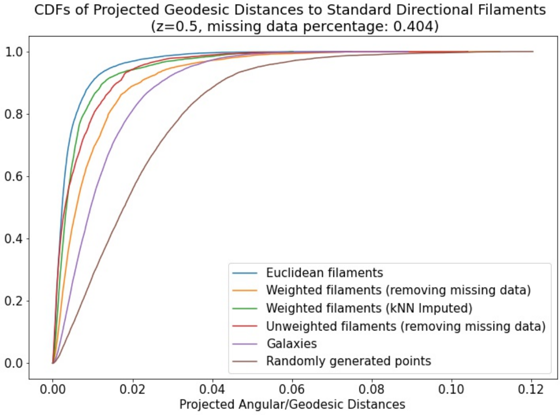
- **Current**: The standard DirSCMS algorithm assumes that each galaxy/particle contributes *equally* to the density estimator and subsequent filament detection in each slice.

- **Current**: The standard DirSCMS algorithm assumes that each galaxy/particle contributes *equally* to the density estimator and subsequent filament detection in each slice.
- **Extension**: It may be interesting to incorporate some extra properties (like stellar masses) to reweight the density estimator.

- **Current**: The standard DirSCMS algorithm assumes that each galaxy/particle contributes *equally* to the density estimator and subsequent filament detection in each slice.
- **Extension**: It may be interesting to incorporate some extra properties (like stellar masses) to reweight the density estimator.
- **Challenge**: Some (galaxy) properties might be missing in our dataset.



One simple method to address the preceding missing data problem is to conduct *data imputation* (through, for instance, K-Nearest Neighbors).

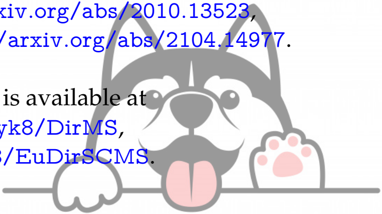


Thank you!

An upcoming talk by my advisor Yen-Chi Chen at JSM 2021 (Virtual):
Talk title: Finding cosmic filament by the directional ridge finding algorithm.
Date & Time: Monday, August 9, 2021 : 6:30 AM to 8:20 AM (Beijing Time)
Session Title: Statistical Answers to Astrophysical Questions: A Vital Chapter in the Chase for New Discoveries – Invited Papers.

More details can be found in https://arxiv.org/abs/2010.13523, https://arxiv.org/abs/2101.10058, https://arxiv.org/abs/2104.14977.

The code for our experiments is available at https://github.com/zhangyk8/DirMS, https://github.com/zhangyk8/EuDirSCMS.



R. Ahumada, C. A. Prieto, A. Almeida, F. Anders, S. F. Anderson, B. H. Andrews, B. Anguiano, R. Arcodia, E. Armengaud, M. Aubert, et al. The 16th data release of the sloan digital sky surveys: first release from the apogee-2 southern survey and full release of eboss spectra. *The Astrophysical Journal Supplement Series*, 249(1):3, 2020.

Z. Bai, C. Rao, and L. Zhao. Kernel estimators of density function of directional data. *Journal of Multivariate Analysis*, 27(1):24 – 39, 1988.

J. R. Bond, L. Kofman, and D. Pogosyan. How filaments of galaxies are woven into the cosmic web. *Nature*, 380(6575):603–606, 1996.

V. Bonjean, N. Aghanim, P. Salomé, M. Douspis, and A. Beelen. Gas and galaxies in filaments between clusters of galaxies-the study of a399-a401. *Astronomy & Astrophysics*, 609:A49, 2018.

M. Cautun, R. van de Weygaert, and B. J. Jones. Nexus: tracing the cosmic web connection. *Monthly Notices of the Royal Astronomical Society*, 429(2):1286–1308, 2013.

M. Cautun, R. Van De Weygaert, B. J. Jones, and C. S. Frenk. Evolution of the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 441(4):2923–2973, 2014.

Y.-C. Chen. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1 (1):161–187, 2017.

Y.-C. Chen, C. R. Genovese, and L. Wasserman. Asymptotic theory for density ridges. *The Annals of Statistics*, 43(5):1896–1928, 2015a.

Y.-C. Chen, S. Ho, P. E. Freeman, C. R. Genovese, and L. Wasserman. Cosmic web reconstruction through density ridges: method and algorithm. *Monthly Notices of the Royal Astronomical Society*, 454 (1):1140–1156, 2015b.

Y.-C. Chen, S. Ho, A. Tenneti, R. Mandelbaum, R. Croft, T. DiMatteo, P. E. Freeman, C. R. Genovese, and L. Wasserman. Investigating galaxy-filament alignments in hydrodynamic simulations using density ridges. *Monthly Notices of the Royal Astronomical Society*, 454(3):3341–3350, 2015c.

Reference II

Y.-C. Chen, S. Ho, J. Brinkmann, P. E. Freeman, C. R. Genovese, D. P. Schneider, and L. Wasserman. Cosmic web reconstruction through density ridges: catalogue. *Monthly Notices of the Royal Astronomical Society*, 461(4):3896–3909, 2016.

Y.-C. Chen, S. Ho, R. Mandelbaum, N. A. Bahcall, J. R. Brownstein, P. E. Freeman, C. R. Genovese, D. P. Schneider, and L. Wasserman. Detecting effects of filaments on galaxy properties in the sloan digital sky survey iii. *Monthly Notices of the Royal Astronomical Society*, 466(2):1880–1893, 2017.

J. Clampitt, H. Miyatake, B. Jain, and M. Takada. Detection of stacked filament lensing between sdss luminous red galaxies. *Monthly Notices of the Royal Astronomical Society*, 457(3):2391–2400, 2016.

E. García-Portugués. Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electronic Journal of Statistics*, 7:1655–1685, 2013.

F. Govoni, E. Orrù, A. Bonafede, M. Iacobelli, R. Paladino, F. Vazza, M. Murgia, V. Vacca, G. Giovannini, L. Feretti, et al. A radio ridge connecting two galaxy clusters in a filament of the cosmic web. *Science*, 364(6444):981–984, 2019.

P. Hall, G. S. Watson, and J. Cabrara. Kernel density estimation with spherical data. *Biometrika*, 74(4): 751–762, 12 1987. ISSN 0006-3444. URL `https://doi.org/10.1093/biomet/74.4.751`.

S. He, S. Alam, S. Ferraro, Y.-C. Chen, and S. Ho. The detection of the imprint of filaments on cosmic microwave background lensing. *Nature Astronomy*, 2(5):401–406, 2018.

C. Laigle, C. Pichon, S. Arnouts, H. J. Mccracken, Y. Dubois, J. Devriendt, A. Slyz, D. Le Borgne, A. Benoit-Levy, H. S. Hwang, et al. Cosmos2015 photometric redshifts probe the impact of filaments on galaxy properties. *Monthly Notices of the Royal Astronomical Society*, 474(4):5437–5458, 2018.

N. I. Libeskind, R. Van De Weygaert, M. Cautun, B. Falck, E. Tempel, T. Abel, M. Alpaslan, M. A. Aragón-Calvo, J. E. Forero-Romero, R. Gonzalez, et al. Tracing the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 473(1):1195–1217, 2018.

Reference III

D. Lynden-Bell, S. Faber, D. Burstein, R. L. Davies, A. Dressler, R. Terlevich, and G. Wegner. Spectroscopy and photometry of elliptical galaxies. v-galaxy streaming toward the new supergalactic center. *The Astrophysical Journal*, 326:19–49, 1988.

C. Maraston and G. Strömbäck. Stellar population models at high spectral resolution. *Monthly Notices of the Royal Astronomical Society*, 418(4):2785–2811, 2011.

B. Moews, M. A. Schmitz, A. J. Lawler, J. Zuntz, A. I. Malz, R. S. de Souza, R. Vilalta, A. Krone-Martins, E. E. Ishida, and C. Collaboration. Ridges in the dark energy survey for cosmic trough identification. *Monthly Notices of the Royal Astronomical Society*, 500(1):859–870, 2021.

U. Ozertem and D. Erdogmus. Locally defined principal curves and surfaces. *Journal of Machine Learning Research*, 12(34):1249–1286, 2011.

A. Poudel, P. Heinämäki, E. Tempel, M. Einasto, H. Lietzen, and P. Nurmi. The effect of cosmic web filaments on the properties of groups and their central galaxies. *Astronomy & Astrophysics*, 597:A86, 2017.

S. F. Shandarin and Y. B. Zeldovich. The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium. *Reviews of Modern Physics*, 61(2):185, 1989.

V. Springel, S. D. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, et al. Simulations of the formation, evolution and clustering of galaxies and quasars. *nature*, 435(7042):629–636, 2005.

E. Tempel, R. Stoica, V. J. Martinez, L. Liivamägi, G. Castellan, and E. Saar. Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the sdss. *Monthly Notices of the Royal Astronomical Society*, 438(4):3465–3482, 2014.

G. Ulrich. Computer generation of distributions on the m-sphere. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(2):158–163, 1984.

Reference IV

D. M. Wilkinson, C. Maraston, D. Goddard, D. Thomas, and T. Parikh. Firefly (fitting iteratively for likelihood analysis): a full spectral fitting code. *Monthly Notices of the Royal Astronomical Society*, 472 (4):4297–4326, 2017.

A. T. Wood. Simulation of the von mises fisher distribution. *Communications in Statistics - Simulation and Computation*, 23(1):157–164, 1994.

Y. B. Zel'Dovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy and astrophysics*, 5:84–89, 1970.

Y. Zhang and Y.-C. Chen. Kernel smoothing, mean shift, and their learning theory with directional data. *arXiv preprint arXiv:2010.13523*, 2020. URL https://arxiv.org/abs/2010.13523.

Y. Zhang and Y.-C. Chen. The em perspective of directional mean shift algorithm. *arXiv preprint arXiv:2101.10058*, 2021a. URL https://arxiv.org/abs/2101.10058.

Y. Zhang and Y.-C. Chen. Linear convergence of the subspace constrained mean shift algorithm: From euclidean to directional data. *arXiv preprint arXiv:2104.14977*, 2021b. URL https://arxiv.org/abs/2104.14977.

Y. Zhang, X. Yang, A. Faltenbacher, V. Springel, W. Lin, and H. Wang. The spin and orientation of dark matter halos within cosmic filaments. *The Astrophysical Journal*, 706(1):747, 2009.

Y. Zhang, X. Yang, H. Wang, L. Wang, H. Mo, and F. C. Van den Bosch. Alignments of galaxies within cosmic filaments from sdss dr7. *The Astrophysical Journal*, 779(2):160, 2013.

Riemannian Gradient and Hessian on Ω_q

Assume tentatively that the directional function f is well-defined and smooth in $\mathbb{R}^{q+1} \setminus \{\mathbf{0}\}$ (or at least in an open neighborhood $U \supset \Omega_q$).

- *Riemannian gradient* $\mathtt{grad} f(\boldsymbol{x})$ on Ω_q:

$$\mathtt{grad} f(\boldsymbol{x}) = \left(\boldsymbol{I}_{q+1} - \boldsymbol{x}\boldsymbol{x}^T\right) \nabla f(\boldsymbol{x}),$$

  where $\boldsymbol{I}_{q+1}$ is the identity matrix in $\mathbb{R}^{(q+1)\times(q+1)}$.

- *Riemannian Hessian* $\mathcal{H}f(\boldsymbol{x})$ on Ω_q (**?**):

$$\mathcal{H}f(\boldsymbol{x}) = (\boldsymbol{I}_{q+1} - \boldsymbol{x}\boldsymbol{x}^T) \left[\nabla\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{x})^T \boldsymbol{x} \cdot \boldsymbol{I}_{q+1}\right] (\boldsymbol{I}_{q+1} - \boldsymbol{x}\boldsymbol{x}^T).$$

Here, $\boldsymbol{I}_{q+1}$ is the identity matrix in $\mathbb{R}^{(q+1)\times(q+1)}$, while $\nabla f(\boldsymbol{x})$ and $\nabla\nabla f(\boldsymbol{x})$ are total gradient and Hessian in $\mathbb{R}^{q+1}$.

Formal Definitions of Directional Density Ridges

- A smooth density function $f : \Omega_q \to \mathbb{R}$.
- Riemannian gradient $\mathtt{grad} f(\boldsymbol{x})$ and Riemannian Hessian $\mathcal{H}f(\boldsymbol{x})$.
- Eigenvalues of Riemannian Hessian $\mathcal{H}f(\boldsymbol{x})$:

$$\lambda_1(\boldsymbol{x}) \geq \cdots \geq \lambda_q(\boldsymbol{x})$$

associated with its unit eigenvectors $\boldsymbol{v}_1(\boldsymbol{x}), ..., \boldsymbol{v}_q(\boldsymbol{x})$ lying within the tangent space $T_{\boldsymbol{x}}$ at $\boldsymbol{x} \in \Omega_q$. (Note that the Riemannian Hessian $\mathcal{H}f(\boldsymbol{x})$ has a unit eigenvector $\boldsymbol{x}$ that is orthogonal to $T_{\boldsymbol{x}}$ and corresponds to eigenvalue 0.)

- Denote $V_d(\boldsymbol{x}) = \left[\boldsymbol{v}_{d+1}(\boldsymbol{x}), ..., \boldsymbol{v}_q(\boldsymbol{x})\right] \in \mathbb{R}^{(q+1)\times(q-d)}$.

$\Longrightarrow$

Local modes of f on Ω_q:

$$\mathcal{M} \equiv \mathtt{Mode}(f) = \left\{\boldsymbol{x} \in \Omega_q : \mathtt{grad} f(\boldsymbol{x}) = \mathbf{0}, \lambda_1(\boldsymbol{x}) < 0\right\}$$

Order-d density ridge on Ω_q (or directional density ridge) of f:

$$\mathcal{R}_d \equiv \mathtt{Ridge}(f) = \left\{\boldsymbol{x} \in \Omega_q : V_d(\boldsymbol{x})V_d(\boldsymbol{x})^T \mathtt{grad} f(\boldsymbol{x}) = \mathbf{0}, \lambda_{d+1}(\boldsymbol{x}) < 0\right\}.$$

Under our scenario of detecting cosmic filaments within a redshift slice, $q = 2$ and $d = 1$.

Formal Definition of Directional KDE

Directional kernel density estimator (KDE; Hall et al. 1987; Bai et al. 1988; García-Portugués 2013):

$$\widehat{f}_h(\boldsymbol{x}) = \frac{c_{L,q}(h)}{n} \sum_{i=1}^{n} L\left(\frac{1-\boldsymbol{x}^T\boldsymbol{X}_i}{h^2}\right). \tag{1}$$

- $\boldsymbol{X}_1, ..., \boldsymbol{X}_n \in \Omega_q \subset \mathbb{R}^{q+1}$ are directional random observations.
- L is a directional kernel, *i.e.*, a rapidly decaying function with nonnegative values on $[0, \infty)$.
- $h > 0$ is the bandwidth parameter.
- $c_{L,q}(h)$ is a normalizing constant satisfying

$$c_{L,q}(h)^{-1} = \int_{\Omega_q} L\left(\frac{1-\boldsymbol{x}^T\boldsymbol{y}}{h^2}\right)\omega_q(d\boldsymbol{y}) = h^q\lambda_{h,q}(L) \asymp h^q\lambda_q(L) \tag{2}$$

with $\lambda_q(L) = 2^{\frac{q}{2}-1}\omega_{q-1}\int_0^\infty L(r)r^{\frac{q}{2}-1}dr$.

An Example of the Directional Kernel

Under the von Mises kernel $L(r) = e^{-r}$,

- directional KDE $\widehat{f}_h(\boldsymbol{x}) = \frac{c_{L,q}(h)}{n} \sum_{i=1}^{n} L\left(\frac{1-\boldsymbol{x}^T\boldsymbol{X}_i}{h^2}\right)$

  becomes

- a mixture of von Mises-Fisher densities:

$$\begin{aligned}\widehat{f}_h(\boldsymbol{x}) &= \frac{1}{n}\sum_{i=1}^{n} f_{\text{vMF}}\left(\boldsymbol{x}; \boldsymbol{X}_i, \frac{1}{h^2}\right) \\ &= \frac{1}{n(2\pi)^{\frac{q+1}{2}}\mathcal{I}_{\frac{q-1}{2}}(1/h^2)h^{q-1}} \sum_{i=1}^{n} \exp\left(\frac{\boldsymbol{x}^T\boldsymbol{X}_i}{h^2}\right).\end{aligned}$$

Input:

- A directional data sample $\boldsymbol{X}_1, ..., \boldsymbol{X}_n \sim f(\boldsymbol{x})$ on Ω_q
- The order d of the directional ridge, smoothing bandwidth $h > 0$, and tolerance level $\epsilon > 0$.
- A suitable mesh $\mathcal{M}_D \subset \Omega_q$ of initial points.

Step 1: Compute the directional KDE $\widehat{f}_h(\boldsymbol{x}) = \frac{c_{L,q}(h)}{n} \sum_{i=1}^{n} L\left(\frac{1-\boldsymbol{x}^T\boldsymbol{X}_i}{h^2}\right)$ on the mesh $\mathcal{M}_D$.

Step 2: For each $\widehat{\boldsymbol{x}}^{(0)} \in \mathcal{M}_D$, iterate the following DirSCMS update until convergence:

while $\left|\left|\sum_{i=1}^{n} \widehat{V}_d(\widehat{\boldsymbol{x}}^{(0)})\widehat{V}_d(\widehat{\boldsymbol{x}}^{(0)})^T\boldsymbol{X}_i \cdot L'\left(\frac{1-\boldsymbol{X}_i^T\widehat{\boldsymbol{x}}^{(0)}}{h^2}\right)\right|\right|_2 > \epsilon$ **do**:

- **Step 2-1**: Compute the scaled version of the estimated Hessian matrix as:

$$\frac{nh^2}{c_{L,q}(h)} \mathcal{H}\widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)}) = \left[\boldsymbol{I}_{q+1} - \widehat{\boldsymbol{x}}^{(t)} \left(\widehat{\boldsymbol{x}}^{(t)}\right)^T\right] \left[\frac{1}{h^2} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T \cdot L'' \left(\frac{1 - \boldsymbol{X}_i^T \widehat{\boldsymbol{x}}^{(t)}}{h^2}\right) \right.$$
$$\left. + \sum_{i=1}^{n} \boldsymbol{X}_i^T \widehat{\boldsymbol{x}}^{(t)} \boldsymbol{I}_{q+1} \cdot L' \left(\frac{1 - \boldsymbol{X}_i^T \widehat{\boldsymbol{x}}^{(t)}}{h^2}\right)\right] \left[\boldsymbol{I}_{q+1} - \widehat{\boldsymbol{x}}^{(t)} \left(\widehat{\boldsymbol{x}}^{(t)}\right)^T\right].$$

- **Step 2-2**: Perform the spectral decomposition on $\frac{nh^2}{c_{L,q}(h)} \mathcal{H}\widehat{f}_h\left(\widehat{\boldsymbol{x}}^{(t)}\right)$ and compute $\widehat{V}_d(\widehat{\boldsymbol{x}}^{(t)}) = \left[\boldsymbol{v}_{d+1}(\widehat{\boldsymbol{x}}^{(t)}), ..., \boldsymbol{v}_q(\widehat{\boldsymbol{x}}^{(t)})\right]$, whose columns are orthonormal eigenvectors corresponding to the smallest $q-d$ eigenvalues inside the tangent space $T_{\widehat{\boldsymbol{x}}^{(t)}}$.

Detailed Procedures of DirSCMS Algorithm III

- **Step 2-3**: Update

$$\widehat{\boldsymbol{x}}^{(t+1)} \leftarrow \widehat{\boldsymbol{x}}^{(t)} - \widehat{V}_d(\widehat{\boldsymbol{x}}^{(t)})\widehat{V}_d(\widehat{\boldsymbol{x}}^{(t)})^T \left[\frac{\sum_{i=1}^n \boldsymbol{X}_i L'\left(\frac{1-\boldsymbol{X}_i^T\widehat{\boldsymbol{x}}^{(t)}}{h^2}\right)}{\sum_{i=1}^n \boldsymbol{X}_i L'\left(\frac{1-\boldsymbol{X}_i^T\widehat{\boldsymbol{x}}^{(t)}}{h^2}\right)}\right].$$

- **Step 2-4**: Standardize $\widehat{\boldsymbol{x}}^{(t+1)}$ as $\widehat{\boldsymbol{x}}^{(t+1)} \leftarrow \frac{\widehat{\boldsymbol{x}}^{(t+1)}}{\left|\left|\widehat{\boldsymbol{x}}^{(t+1)}\right|\right|_2}$.

Output: An estimated directional d-ridge $\widehat{\mathcal{R}}_d$ represented by the collection of resulting points.