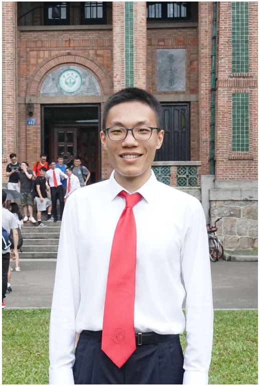
Kernel Smoothing and Mean Shift Theories with Applications to Cosmic Web Detection

Yikun Zhang[*]
(Joint work with *Yen-Chi Chen*[*]
and *Rafael S. de Souza*[†])

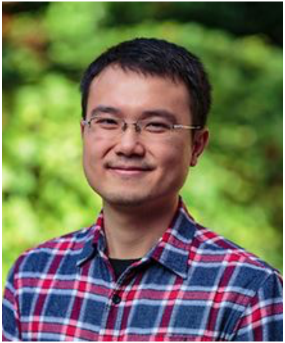
[*]Department of Statistics,
University of Washington
[†] Shanghai Astronomical Observatory

Venue: September 23, 2022 at Sun Yat-Sen University

Contributors



Me in 2018



Professor Yen-Chi Chen

Professor Rafael S. de Souza

Introduction

Background: What Are Directional Data?

In some real-world applications, observations are inherently directional (or L_2 normalized) in nature. Data of this type arise in:

Background: What Are Directional Data?

In some real-world applications, observations are inherently directional (or L_2 normalized) in nature. Data of this type arise in:

- **Astronomy**: celestial coordinates of galaxies or stars.
- **Geology**: locations of craters, lakes, and other geological features on Earth or other planets, epicenters of earthquakes (**Seismology**).
- **Biology**: yeast gene expression analysis, animal navigation.
- **Text mining**: cosine similarities between text documents.

Background: What Are Directional Data?

In some real-world applications, observations are inherently directional (or L_2 normalized) in nature. Data of this type arise in:

- **Astronomy**: celestial coordinates of galaxies or stars.
- **Geology**: locations of craters, lakes, and other geological features on Earth or other planets, epicenters of earthquakes (**Seismology**).
- **Biology**: yeast gene expression analysis, animal navigation.
- **Text mining**: cosine similarities between text documents.

Mathematically, a *directional dataset* consists of observations

$$\boldsymbol{X}_1, ..., \boldsymbol{X}_n \overset{i.i.d.}{\sim} f,$$

where f is a directional density supported on the *unit hypersphere*

$$\Omega_q := \left\{\boldsymbol{x} \in \mathbb{R}^{q+1} : ||\boldsymbol{x}||_2 = 1\right\}$$

with $\int_{\Omega_q} f(\boldsymbol{x})\, \omega_q(d\boldsymbol{x}) = 1$ and $||\cdot||_2$ is the L_2-norm in $\mathbb{R}^{q+1}$.

$*$ Notes: ω_q is the Lebesgue measure on Ω_q.

In astronomical surveys, the positions of observed objects are recorded as $\{(\alpha_1, \delta_1, z_1), ..., (\alpha_n, \delta_n, z_n)\} \subset \Omega_2 \times \mathbb{R}^+$, where, for $i = 1, ..., n$, In astronomical surveys, the positions of observed objects are recorded as $\{(\alpha_1, \delta_1, z_1), ..., (\alpha_n, \delta_n, z_n)\} \subset \Omega_2 \times \mathbb{R}^+$, where, for $i = 1, ..., n$,

- $\alpha_i \in [0, 360°)$ is the *right ascension* (RA), i.e., celestial longitude,
- $\delta_i \in [-90°, 90°]$ is the *declination* (DEC), i.e., celestial latitude.
- $z_i \in (0, \infty)$ is the *redshift* value.

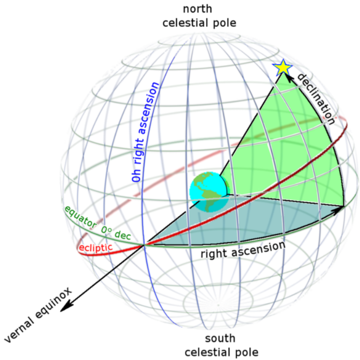


Figure 1: Illustration of RA and DEC (Image Courtesy of Wikipedia).

Background: What Is Cosmic Web?

Cosmic Web is a large-scale network structure revealing that the matter in our Universe is not uniformly distributed (Bond et al., 1996).

Background: What Is Cosmic Web?

Cosmic Web is a large-scale network structure revealing that the matter in our Universe is not uniformly distributed (Bond et al., 1996).

- **Large scale**: 1 Mpc $\approx$ 3.26 light-years.
- **Cause**: the anisotropic collapse of matter in gravitational instability scenarios at the early stage of the Universe (Zel'Dovich, 1970).

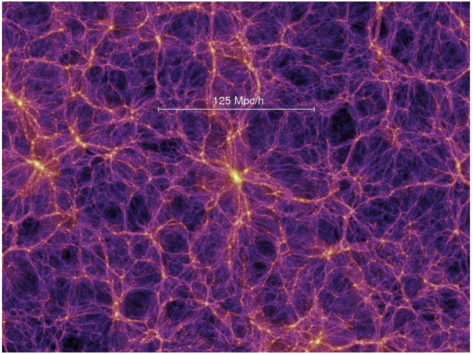


Figure 2: Visualization of *Cosmic Web* (credited to the millennium simulation project (Springel et al., 2005)).

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

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

- Massive galaxy *clusters* (*or nodes*),

Background: Characteristics of Cosmic Web

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

- Massive galaxy *clusters* (*or nodes*),
- Interconnected *filaments*,

Background: Characteristics of Cosmic Web

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

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

Background: Characteristics of Cosmic Web

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

- Massive galaxy *clusters* (*or nodes*),
- Interconnected *filaments*,
- Two-dimensional tenuous *sheets*/*walls*,

} on which matter concentrates.

around • Vast and near-empty *voids*.

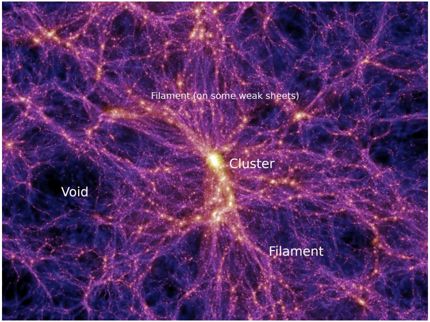


Figure 3: Characteristics of *Cosmic Web* (credited to the millennium simulation).

Motivation: Cosmic Web Detection

Objective: Detect the cosmic web structure based on the distribution of observed galaxies.

Motivation: Cosmic Web Detection

Objective: Detect the cosmic web structure based on the distribution of observed galaxies.

1. First on the 2D celestial sphere Ω_2.
2. Then generalize to the 3D (redshift) space.

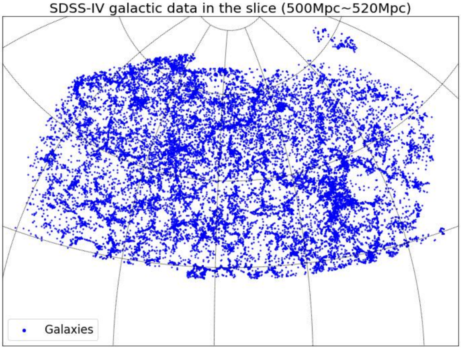


Figure 4: Distribution of galaxies on Ω_2 within a thin redshift slice.

Motivation: Cosmic Web Detection

Objective: Detect the cosmic web structure based on the distribution of observed galaxies.

1. First on the 2D celestial sphere Ω_2.
2. Then generalize to the 3D (redshift) space.

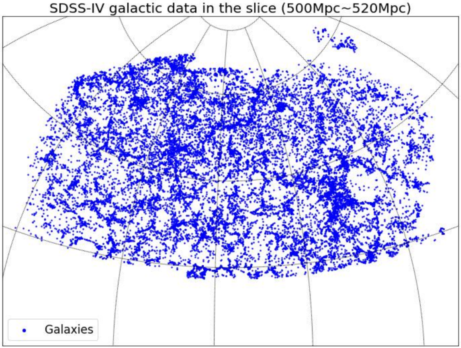


Figure 4: Distribution of galaxies on Ω_2 within a thin redshift slice.

- In particular, we focus on identifying the **cosmic filaments**.

Motivation: Significance of Cosmic Filaments

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

- 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).

Motivation: Significance of Cosmic Filaments

- 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).
- The trajectory of cosmic microwave background light can be distorted due to cosmic filaments, creating the weak lensing effect.

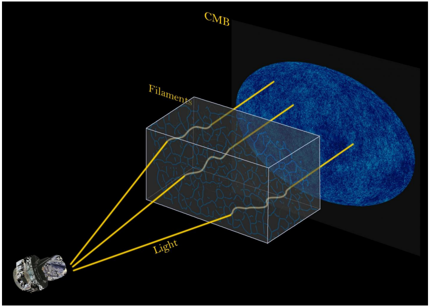


Figure 5: Illustration of the bending trajectory of CMB lights (credit to Siyu He, Shadab Alam, Wei Chen, and Planck/ESA; see He et al. (2018) for details).

Challenges in Detecting Cosmic Filaments

- The filamentary structures are overwhelmingly complex (Cautun et al., 2013, 2014):

Challenges in Detecting Cosmic Filaments

- The filamentary structures are overwhelmingly complex (Cautun et al., 2013, 2014):
  - Lack of structural symmetries,
  - Uncertainty in measuring its connectivity,
  - Intrinsic multi-scale nature, etc.

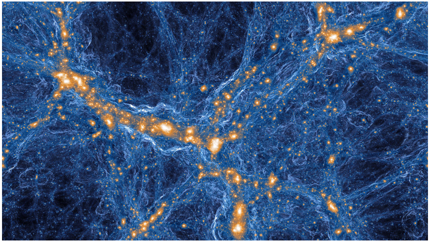


Figure 6: A view of the present-day cosmic web 300 million light-years across, as modeled by IllustrisTNG (Vogelsberger et al., 2014).

Challenges in Detecting Cosmic Filaments

- The filamentary structures are overwhelmingly complex (Cautun et al., 2013, 2014):
  - Lack of structural symmetries,
  - Uncertainty in measuring its connectivity,
  - Intrinsic multi-scale nature, etc.

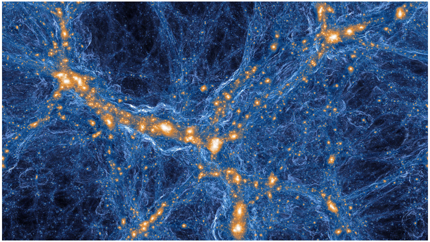


Figure 6: A view of the present-day cosmic web 300 million light-years across, as modeled by IllustrisTNG (Vogelsberger et al., 2014).

- There exist no universal and mathematically rigorous definitions about cosmic filaments!

Highlights of the Today's Talk

Highlights of the Today's Talk

1. A brief review about existing research on cosmic filament detection.

Highlights of the Today's Talk

1. A brief review about existing research on cosmic filament detection.

2. Introduction to our cosmic filament model.
   - Formulate cosmic filaments as the *directional density ridges* on Ω_2.

Highlights of the Today's Talk

1. A brief review about existing research on cosmic filament detection.
2. Introduction to our cosmic filament model.
   - Formulate cosmic filaments as the *directional density ridges* on Ω_2.
3. Discussion on *directional density ridges* from both statistical and computational perspectives.

Highlights of the Today's Talk

1. A brief review about existing research on cosmic filament detection.
2. Introduction to our cosmic filament model.
   - Formulate cosmic filaments as the *directional density ridges* on Ω_2.
3. Discussion on *directional density ridges* from both statistical and computational perspectives.
   - Prove the statistical consistency in estimating the true density ridges with directional kernel density estimator (KDE).

Highlights of the Today's Talk

1. A brief review about existing research on cosmic filament detection.
2. Introduction to our cosmic filament model.
   - Formulate cosmic filaments as the *directional density ridges* on Ω_2.
3. Discussion on *directional density ridges* from both statistical and computational perspectives.
   - Prove the statistical consistency in estimating the true density ridges with directional kernel density estimator (KDE).
   - Estimate the directional density ridges via our *Directional Subspace Constrained Mean Shift* (DirSCMS) algorithm.

Highlights of the Today's Talk

1. A brief review about existing research on cosmic filament detection.
2. Introduction to our cosmic filament model.
   - Formulate cosmic filaments as the *directional density ridges* on Ω_2.
3. Discussion on *directional density ridges* from both statistical and computational perspectives.
   - Prove the statistical consistency in estimating the true density ridges with directional kernel density estimator (KDE).
   - Estimate the directional density ridges via our *Directional Subspace Constrained Mean Shift* (DirSCMS) algorithm.
   - Establish the linear convergence properties of our DirSCMS algorithm.

Highlights of the Today's Talk

1. A brief review about existing research on cosmic filament detection.
2. Introduction to our cosmic filament model.
   - Formulate cosmic filaments as the *directional density ridges* on Ω_2.
3. Discussion on *directional density ridges* from both statistical and computational perspectives.
   - Prove the statistical consistency in estimating the true density ridges with directional kernel density estimator (KDE).
   - Estimate the directional density ridges via our *Directional Subspace Constrained Mean Shift* (DirSCMS) algorithm.
   - Establish the linear convergence properties of our DirSCMS algorithm.
4. Application on Sloan Digital Sky Survey (SDSS-IV; Ahumada et al. 2020) galactic data to construct a cosmic web catalog.

Previous Works on Filament Detection

Categorization of Cosmic Filament Finders

Recall that the observed galaxies in any astronomical survey have their coordinates as $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^n \subset \Omega_2 \times \mathbb{R}^+$.

The existing methods for detecting cosmic filaments from survey data can be classified into the following two categories:

Categorization of Cosmic Filament Finders

Recall that the observed galaxies in any astronomical survey have their coordinates as $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^n \subset \Omega_2 \times \mathbb{R}^+$.

The existing methods for detecting cosmic filaments from survey data can be classified into the following two categories:

- **3D method**: Convert redshifts into (comoving) distances (Tempel et al., 2014; Sousbie et al., 2011; Pfeifer et al., 2022).
- **2D method**: Slice the Universe into thin redshift slices (Chen et al., 2015b; Duque et al., 2022).

► Our method can easily switch between the above two categories.

3D Method for Detecting Filaments

One convert $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^{n}$ to their Cartesian coordinates as

$$X_i = d(z_i) \cos \alpha_i \cos \delta_i,$$
$$Y_i = d(z_i) \sin \alpha_i \cos \delta_i,$$
$$Z_i = d(z_i) \sin \delta_i,$$

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

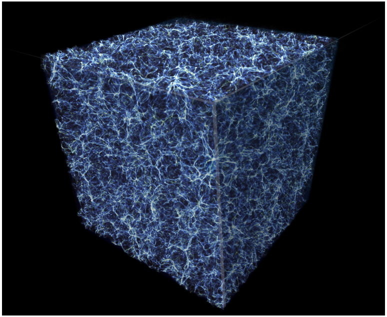


Figure 7: Matter distribution in a cubic section of the Universe (credit to NASA, ESA, and E. Hallman at University of Colorado, Boulder)

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

Drawback of 3D Methods

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

- The determination of $d(\cdot)$ relies on complex cosmological models.

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

- The determination of $d(\cdot)$ relies on complex cosmological models.
- The galaxy distribution is distorted along the line of sight due to the peculiar velocities of galaxies (i.e., the so-called *finger-of-god* (Sargent and Turner, 1977) and *Kaiser* (Kaiser, 1987) effects).

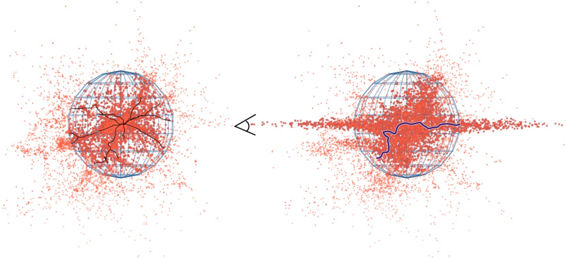


Figure 8: Redshift distortions along the line of sight (Kuchner et al., 2021).

Drawback of 3D Methods

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

- The determination of $d(\cdot)$ relies on complex cosmological models.
- The galaxy distribution is distorted along the line of sight due to the peculiar velocities of galaxies (i.e., the so-called *finger-of-god* (Sargent and Turner, 1977) and *Kaiser* (Kaiser, 1987) effects).

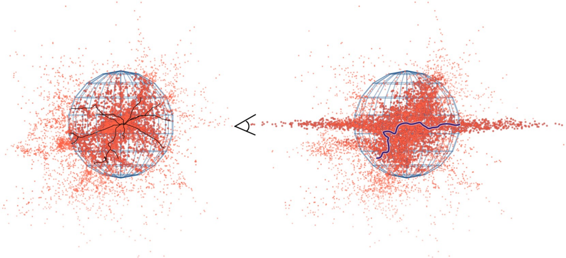


Figure 8: Redshift distortions along the line of sight (Kuchner et al., 2021).

- The number of galaxies varies across different redshift values, so applying 3D approaches will be computationally intensive.

We partition the redshift range of observed galaxies into several non-overlapping thin slices.

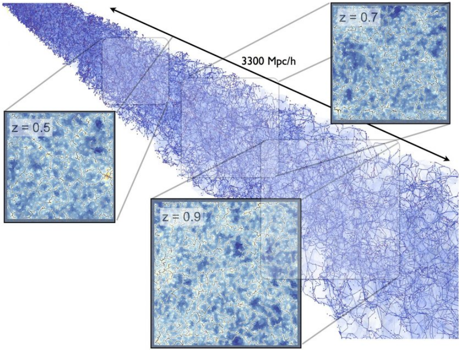


Figure 9: Illustration of slicing the Universe (credit to Laigle et al. 2018)

We partition the redshift range of observed galaxies into several non-overlapping thin slices.

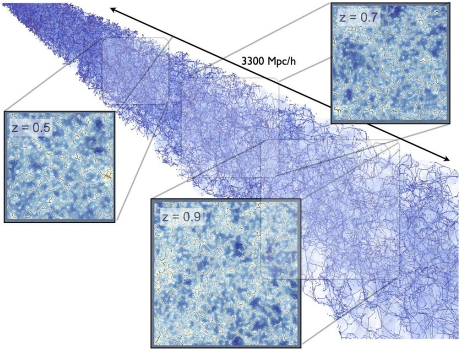


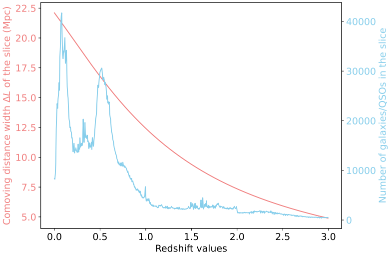
Figure 9: Illustration of slicing the Universe (credit to Laigle et al. 2018)

This tomographic approach has its own advantages over 3D methods:

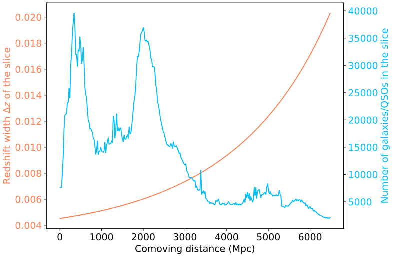
- It controls the redshift distortions along the line-of-sight direction.
- The measurement error in one slice won't propagate to other slices.
- It helps reduce computational cost...

Caveats in Slicing the Universe (I)

1. We slice the Universe via a cosmological model, such as Planck15 (Ade et al., 2016) or WMAP9 (Hinshaw et al., 2013) ΛCDM cosmology, but not in the original redshift space.

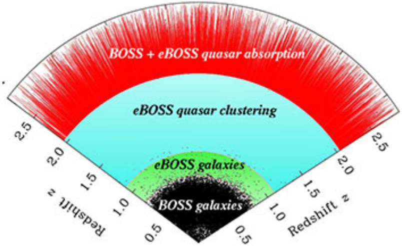


(a) By redshift $\Delta z = 0.005$.

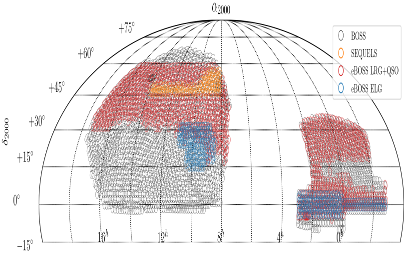


(b) By comoving distance $\Delta L = 20\,\mathrm{Mpc}$.

2. The resulting (redshift) slices are not flat 2D planes, but some **spherical shell**, which have a *nonlinear* curvature!
   - Recall that the locations of galaxies in a slice are recorded by $\{(\alpha_i, \delta_i)\}_{i=1}^n \subset \Omega_2$ on a celestial sphere.



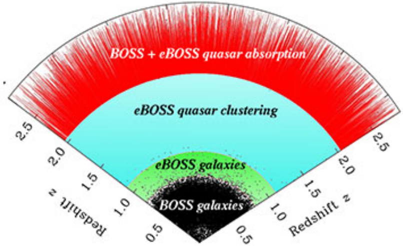
(a) Planned eBOSS coverage of the Universe (credit to M. Blanton and SDSS)



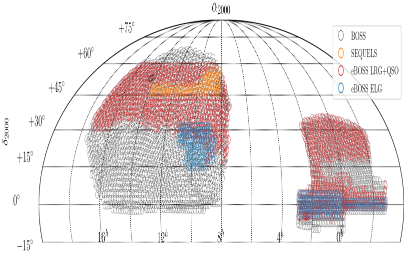
(b) BOSS/eBOSS Spectroscopic Footprint as of DR16 (credit to SDSS)

Caveats in Slicing the Universe (II)

2. The resulting (redshift) slices are not flat 2D planes, but some **spherical shell**, which have a *nonlinear* curvature!
   - Recall that the locations of galaxies in a slice are recorded by $\{(\alpha_i, \delta_i)\}_{i=1}^{n} \subset \Omega_2$ on a celestial sphere.



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



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

Problem: How do we model and estimate the cosmic filaments based on the observed galaxies in each (redshift) slice?

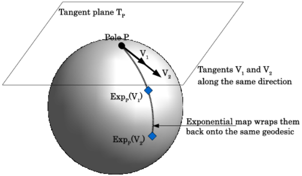
Cosmic Filament Model: Directional Density Ridges

Definition (Tangent space of Ω_q)

The *tangent space* of the sphere Ω_q at $\boldsymbol{x} \in \Omega_q$ is given by

$$T_{\boldsymbol{x}} \equiv T_{\boldsymbol{x}}(\Omega_q) = \left\{\boldsymbol{u} - \boldsymbol{x} \in \mathbb{R}^{q+1} : \boldsymbol{x}^T(\boldsymbol{u} - \boldsymbol{x}) = 0\right\} \simeq \left\{\boldsymbol{v} \in \mathbb{R}^{q+1} : \boldsymbol{x}^T\boldsymbol{v} = 0\right\},$$

where $V_1 \simeq V_2$ signifies that the two vector spaces are isomorphic. In what follows, $\boldsymbol{v} \in T_{\boldsymbol{x}}$ indicates that $\boldsymbol{v}$ is a vector tangent to Ω_q at $\boldsymbol{x}$.



Definition (Exponential Map)

An *exponential map* at $\boldsymbol{x} \in \Omega_q$ is a mapping $\mathtt{Exp}_{\boldsymbol{x}} : T_{\boldsymbol{x}} \to \Omega_q$ such that the vector $\boldsymbol{v} \in T_{\boldsymbol{x}}$ is mapped to point $\boldsymbol{y} := \mathtt{Exp}_{\boldsymbol{x}}(\boldsymbol{v}) \in \Omega_q$ with $\gamma(0) = \boldsymbol{x}, \gamma(1) = \boldsymbol{y}$ and $\gamma'(0) = \boldsymbol{v}$, where $\gamma : [0, 1] \to \Omega_q$ is a geodesic.

Riemannian Gradient on Ω_q

Given a smooth function $f : \Omega_q \to \mathbb{R}$, we extend its domain from Ω_q to $\mathbb{R}^{q+1} \setminus \{\mathbf{0}\}$ as:

$$f(\boldsymbol{x}) \equiv f\left(\frac{\boldsymbol{x}}{||\boldsymbol{x}||_2}\right) \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^{q+1} \setminus \{\mathbf{0}\}.$$

Given a smooth curve $\gamma : (-\epsilon, \epsilon) \to \Omega_q$ with $\gamma(0) = \boldsymbol{x}$ and $\gamma'(0) = \boldsymbol{v} \in T_{\boldsymbol{x}}$, the *differential* of f at point $\boldsymbol{x} \in \Omega_q$ is a linear map $df_{\boldsymbol{x}} : T_{\boldsymbol{x}} \to T_{f(\boldsymbol{x})}(\mathbb{R}) \simeq \mathbb{R}$ given by

$$df_{\boldsymbol{x}}(\boldsymbol{v}) = \frac{d}{dt} f(\gamma(t))\Big|_{t=0} = (f \circ \gamma)'(0). \tag{1}$$

Riemannian Gradient on Ω_q

Given a smooth function $f : \Omega_q \to \mathbb{R}$, we extend its domain from Ω_q to $\mathbb{R}^{q+1} \setminus \{\mathbf{0}\}$ as:

$$f(\boldsymbol{x}) \equiv f\left(\frac{\boldsymbol{x}}{||\boldsymbol{x}||_2}\right) \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^{q+1} \setminus \{\mathbf{0}\}.$$

Given a smooth curve $\gamma : (-\epsilon, \epsilon) \to \Omega_q$ with $\gamma(0) = \boldsymbol{x}$ and $\gamma'(0) = \boldsymbol{v} \in T_{\boldsymbol{x}}$, the *differential* of f at point $\boldsymbol{x} \in \Omega_q$ is a linear map $df_{\boldsymbol{x}} : T_{\boldsymbol{x}} \to T_{f(\boldsymbol{x})}(\mathbb{R}) \simeq \mathbb{R}$ given by

$$df_{\boldsymbol{x}}(\boldsymbol{v}) = \frac{d}{dt} f(\gamma(t))\Big|_{t=0} = (f \circ \gamma)'(0). \tag{1}$$

Definition (Riemannian Gradient)

The *Riemannian gradient* $\texttt{grad}\, f(\boldsymbol{x}) \in T_{\boldsymbol{x}} \subset \mathbb{R}^{q+1}$ is defined by

$$\langle \texttt{grad}\, f(\boldsymbol{x}), \boldsymbol{v} \rangle_{\boldsymbol{x}} = df_{\boldsymbol{x}}(\boldsymbol{v}) \tag{2}$$

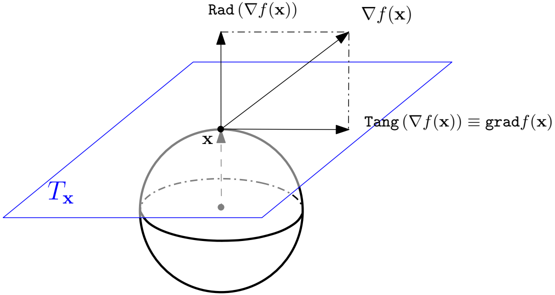
for any $\boldsymbol{v} \in T_{\boldsymbol{x}}$ and the predefined Riemannian metric $\langle \cdot, \cdot \rangle_{\boldsymbol{x}}$.

Riemannian Gradient on Ω_q

Given that Ω_q is a submanifold in $\mathbb{R}^{q+1}$, we relate the Riemannian gradient $\texttt{grad} f(\boldsymbol{x})$ on Ω_q with the total gradient $\nabla f(\boldsymbol{x})$ in $\mathbb{R}^{q+1}$ as:

$$\texttt{grad} f(\boldsymbol{x}) = \left(I_{q+1} - \boldsymbol{x}\boldsymbol{x}^T\right) \nabla f(\boldsymbol{x}), \tag{3}$$

which is the projection of $\nabla f(\boldsymbol{x})$ onto the tangent space $T_{\boldsymbol{x}}$ at $\boldsymbol{x} \in \Omega_q$ (Absil et al., 2009). Here, I_{q+1} is the identity matrix in $\mathbb{R}^{(q+1)\times(q+1)}$.



Riemannian Hessian on Ω_q

Definition (Riemannian Hessian)

The *Riemannian Hessian* of f at $\boldsymbol{x} \in \Omega_q$ is a linear mapping $\mathcal{H}f(\boldsymbol{x}) : T_{\boldsymbol{x}} \to T_{\boldsymbol{x}}$ defined by

$$\mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}] = \bar{\nabla}_{\boldsymbol{v}} \texttt{grad} f(\boldsymbol{x}) \tag{4}$$

for any $\boldsymbol{v} \in T_{\boldsymbol{x}}$, where $\bar{\nabla}_{\boldsymbol{v}}$ is the Riemannian connection on Ω_q.

Riemannian Hessian on Ω_q

Definition (Riemannian Hessian)

The *Riemannian Hessian* of f at $\boldsymbol{x} \in \Omega_q$ is a linear mapping $\mathcal{H}f(\boldsymbol{x}) : T_{\boldsymbol{x}} \to T_{\boldsymbol{x}}$ defined by

$$\mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}] = \bar{\nabla}_{\boldsymbol{v}} \texttt{grad} f(\boldsymbol{x}) \tag{4}$$

for any $\boldsymbol{v} \in T_{\boldsymbol{x}}$, where $\bar{\nabla}_{\boldsymbol{v}}$ is the Riemannian connection on Ω_q.

1. It is self-adjoint with respect to the Riemannian metric as:

$$\langle \mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}], \boldsymbol{u} \rangle_{\boldsymbol{x}} = \langle \boldsymbol{v}, \mathcal{H}f(\boldsymbol{x})[\boldsymbol{u}] \rangle_{\boldsymbol{x}}.$$

Riemannian Hessian on Ω_q

Definition (Riemannian Hessian)

The *Riemannian Hessian* of f at $\boldsymbol{x} \in \Omega_q$ is a linear mapping $\mathcal{H}f(\boldsymbol{x}) : T_{\boldsymbol{x}} \to T_{\boldsymbol{x}}$ defined by

$$\mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}] = \bar{\nabla}_{\boldsymbol{v}} \texttt{grad} f(\boldsymbol{x}) \tag{4}$$

for any $\boldsymbol{v} \in T_{\boldsymbol{x}}$, where $\bar{\nabla}_{\boldsymbol{v}}$ is the Riemannian connection on Ω_q.

1. It is self-adjoint with respect to the Riemannian metric as:

$$\langle \mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}], \boldsymbol{u} \rangle_{\boldsymbol{x}} = \langle \boldsymbol{v}, \mathcal{H}f(\boldsymbol{x})[\boldsymbol{u}] \rangle_{\boldsymbol{x}}.$$

Riemannian Hessian on Ω_q

Definition (Riemannian Hessian)

The *Riemannian Hessian* of f at $\boldsymbol{x} \in \Omega_q$ is a linear mapping $\mathcal{H}f(\boldsymbol{x}) : T_{\boldsymbol{x}} \to T_{\boldsymbol{x}}$ defined by

$$\mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}] = \bar{\nabla}_{\boldsymbol{v}} \texttt{grad} f(\boldsymbol{x}) \tag{4}$$

for any $\boldsymbol{v} \in T_{\boldsymbol{x}}$, where $\bar{\nabla}_{\boldsymbol{v}}$ is the Riemannian connection on Ω_q.

1. It is self-adjoint with respect to the Riemannian metric as:

$$\langle \mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}], \boldsymbol{u} \rangle_{\boldsymbol{x}} = \langle \boldsymbol{v}, \mathcal{H}f(\boldsymbol{x})[\boldsymbol{u}] \rangle_{\boldsymbol{x}}.$$

2. It is related to the total gradient $\nabla f(\boldsymbol{x})$ and total Hessian $\nabla\nabla f(\boldsymbol{x})$ as (Zhang and Chen, 2021c):

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

Definition (Riemannian Hessian)

The *Riemannian Hessian* of f at $\boldsymbol{x} \in \Omega_q$ is a linear mapping $\mathcal{H}f(\boldsymbol{x}) : T_{\boldsymbol{x}} \to T_{\boldsymbol{x}}$ defined by

$$\mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}] = \bar{\nabla}_{\boldsymbol{v}} \texttt{grad} f(\boldsymbol{x}) \tag{4}$$

for any $\boldsymbol{v} \in T_{\boldsymbol{x}}$, where $\bar{\nabla}_{\boldsymbol{v}}$ is the Riemannian connection on Ω_q.

1. It is self-adjoint with respect to the Riemannian metric as:
$$\langle \mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}], \boldsymbol{u} \rangle_{\boldsymbol{x}} = \langle \boldsymbol{v}, \mathcal{H}f(\boldsymbol{x})[\boldsymbol{u}] \rangle_{\boldsymbol{x}}.$$
2. It is related to the total gradient $\nabla f(\boldsymbol{x})$ and total Hessian $\nabla\nabla f(\boldsymbol{x})$ as (Zhang and Chen, 2021c):
$$\mathcal{H}f(\boldsymbol{x}) = (\boldsymbol{I}_{q+1} - \boldsymbol{x}\boldsymbol{x}^T)\left[\nabla\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{x})^T \boldsymbol{x} \cdot \boldsymbol{I}_{q+1}\right](\boldsymbol{I}_{q+1} - \boldsymbol{x}\boldsymbol{x}^T).$$
3. Taylor's expansion (Pennec, 2006):
$$(f \circ \texttt{Exp}_{\boldsymbol{x}})(\boldsymbol{v}) = f(\boldsymbol{x}) + \langle \texttt{grad} f(\boldsymbol{x}), \boldsymbol{v} \rangle_{\boldsymbol{x}} + \frac{1}{2}\langle \mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}], \boldsymbol{v} \rangle_{\boldsymbol{x}} + O\left(||\boldsymbol{v}||^3\right).$$

Characterization of High-Density Regions

We perform the spectral decomposition (Horn and Johnson, 2012) on the Riemannian Hessian $\mathcal{H}f(\boldsymbol{x})$ as:

$$\mathcal{H}f(\boldsymbol{x}) = V(\boldsymbol{x}) \begin{pmatrix} 0 & & & \\ & \lambda_1(\boldsymbol{x}) & & \\ & & \ddots & \\ & & & \lambda_q(\boldsymbol{x}) \end{pmatrix} V(\boldsymbol{x})^T,$$

where $V(\boldsymbol{x}) = \left(\boldsymbol{x}, \boldsymbol{v}_1(\boldsymbol{x}), ..., \boldsymbol{v}_q(\boldsymbol{x})\right) \in \mathbb{R}^{(q+1)\times(q+1)}$ has its columns as the unit eigenvectors of $\mathcal{H}f(\boldsymbol{x})$. Here,

- Eigenvectors $\boldsymbol{v}_1(\boldsymbol{x}), ..., \boldsymbol{v}_q(\boldsymbol{x})$ lie within the tangent space $T_{\boldsymbol{x}}$.
- Descending eigenvalues: $\lambda_1(\boldsymbol{x}) \geq \cdots \geq \lambda_q(\boldsymbol{x})$.
- It has an eigenvector $\boldsymbol{x}$ that is normal to $T_{\boldsymbol{x}}$ and with eigenvalue 0.

We perform the spectral decomposition (Horn and Johnson, 2012) on the Riemannian Hessian $\mathcal{H}f(\boldsymbol{x})$ as:

$$\mathcal{H}f(\boldsymbol{x}) = V(\boldsymbol{x}) \begin{pmatrix} 0 & & & \\ & \lambda_1(\boldsymbol{x}) & & \\ & & \ddots & \\ & & & \lambda_q(\boldsymbol{x}) \end{pmatrix} V(\boldsymbol{x})^T,$$

where $V(\boldsymbol{x}) = \left(\boldsymbol{x}, \boldsymbol{v}_1(\boldsymbol{x}), ..., \boldsymbol{v}_q(\boldsymbol{x})\right) \in \mathbb{R}^{(q+1)\times(q+1)}$ has its columns as the unit eigenvectors of $\mathcal{H}f(\boldsymbol{x})$. Here,

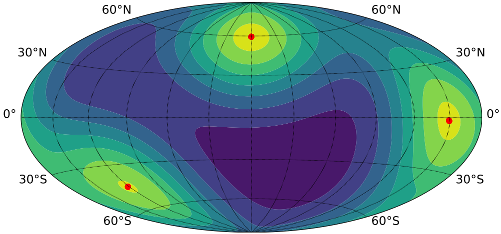
- Eigenvectors $\boldsymbol{v}_1(\boldsymbol{x}), ..., \boldsymbol{v}_q(\boldsymbol{x})$ lie within the tangent space $T_{\boldsymbol{x}}$.
- Descending eigenvalues: $\lambda_1(\boldsymbol{x}) \geq \cdots \geq \lambda_q(\boldsymbol{x})$.
- It has an eigenvector $\boldsymbol{x}$ that is normal to $T_{\boldsymbol{x}}$ and with eigenvalue 0.

Local Modes/Maxima of f on Ω_q:

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

The set of local modes $\mathcal{M}$ signifies the **zero-dimensional** high-density regions of f.

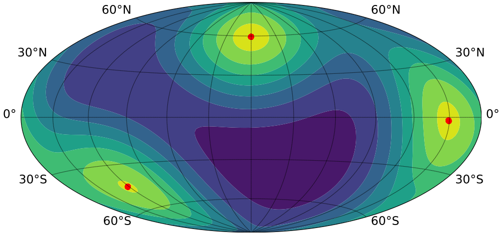
- When f is the underlying galaxy density function, $\mathcal{M}$ points to some good candidates of *galaxy clusters*.



Local modes of the Density Function on Ω_q

The set of local modes $\mathcal{M}$ signifies the **zero-dimensional** high-density regions of f.

- When f is the underlying galaxy density function, $\mathcal{M}$ points to some good candidates of *galaxy clusters*.



► However, cosmic filaments are some one-dimensional curves!

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

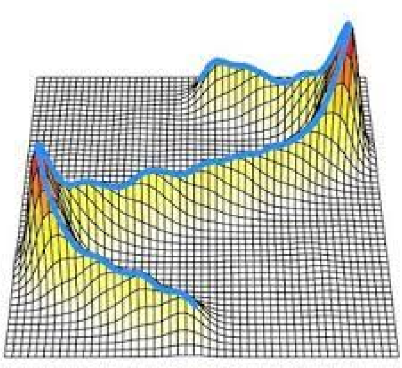


Figure 12: Density ridge (lifted onto the density function f) (credit to Yen-Chi Chen)

Directional Density Ridges (II)

The order-d density ridge on Ω_q (or *directional density ridge*) of f is defined as:

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

where $V_d(\boldsymbol{x}) = \left[\boldsymbol{v}_{d+1}(\boldsymbol{x}), ..., \boldsymbol{v}_q(\boldsymbol{x})\right] \in \mathbb{R}^{(q+1)\times(q-d)}$ consists of the last $q-d$ eigenvectors of $\mathcal{H}f(\boldsymbol{x})$ within $T_{\boldsymbol{x}}$.

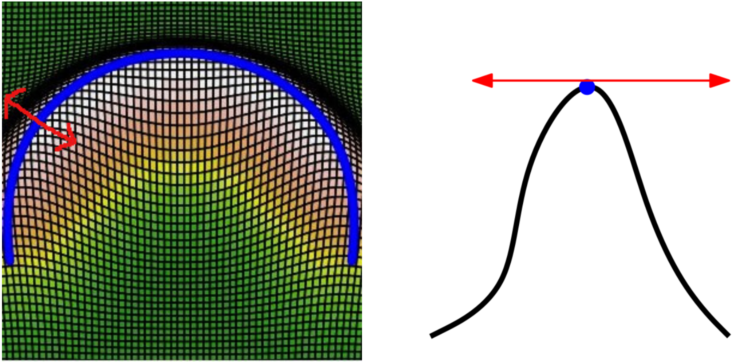


Figure 13: Density ridge (lifted onto the density function f; Chen et al. 2015a)

Statistical and Algorithmic Estimation of Directional Density Ridges

Directional Density Estimation

How do we estimate the directional density ridge $\mathcal{R}_d$ and the set of local mode $\mathcal{M}$ on Ω_q from directional data $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\} \subset \Omega_q$?

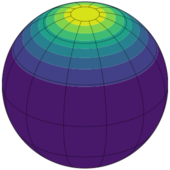
Directional Density Estimation

How do we estimate the directional density ridge $\mathcal{R}_d$ and the set of local mode $\mathcal{M}$ on Ω_q from directional data $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\} \subset \Omega_q$?

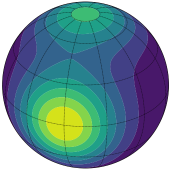
We first estimate the density function f on Ω_q via the directional KDE (Hall et al., 1987; Bai et al., 1988; García-Portugués, 2013) as:

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

- $L : [0, \infty) \to [0, \infty)$ is a directional kernel, *i.e.*, a rapidly decaying nonnegative function. (Example: von Mises kernel $L(r) = e^{-r}$.)
- $h > 0$ is the bandwidth parameter, and $C_{L,q}(h)$ is a normalizing term.



(a) $f_{\text{vMF},2}(\boldsymbol{x}; \boldsymbol{\mu}, \nu)$ with $\boldsymbol{\mu} = (0, 0, 1)$ and $\nu = 4.0$.



(b) $\frac{2}{5} \cdot f_{\text{vMF},2}(\boldsymbol{x}; \boldsymbol{\mu}_1, 5) + \frac{3}{5} \cdot f_{\text{vMF},2}(\boldsymbol{x}; \boldsymbol{\mu}_2, 5)$ with $\boldsymbol{\mu}_1 = (0, 0, 1)$, $\boldsymbol{\mu}_2 = (1, 0, 0)$.

The directional KDE $\widehat{f}_h$ is useful because its plug-in estimators

$$\widehat{\mathcal{M}} = \left\{ \boldsymbol{x} \in \Omega_q : \texttt{grad} \widehat{f}_h(\boldsymbol{x}) = \mathbf{0}, \widehat{\lambda}_1(\boldsymbol{x}) < 0 \right\}$$

and

$$\widehat{\mathcal{R}}_d = \left\{ \boldsymbol{x} \in \Omega_q : \widehat{V}_d(\boldsymbol{x}) \widehat{V}_d(\boldsymbol{x})^T \texttt{grad} \widehat{f}_h(\boldsymbol{x}) = \mathbf{0}, \widehat{\lambda}_{d+1}(\boldsymbol{x}) < 0 \right\}$$

The directional KDE $\widehat{f}_h$ is useful because its plug-in estimators

$$\widehat{\mathcal{M}} = \left\{ \boldsymbol{x} \in \Omega_q : \texttt{grad} \widehat{f}_h(\boldsymbol{x}) = \mathbf{0}, \widehat{\lambda}_1(\boldsymbol{x}) < 0 \right\}$$

and

$$\widehat{\mathcal{R}}_d = \left\{ \boldsymbol{x} \in \Omega_q : \widehat{V}_d(\boldsymbol{x}) \widehat{V}_d(\boldsymbol{x})^T \texttt{grad} \widehat{f}_h(\boldsymbol{x}) = \mathbf{0}, \widehat{\lambda}_{d+1}(\boldsymbol{x}) < 0 \right\}$$

approach $\mathcal{M}$ and $\mathcal{R}_d$ in a statistically consistent way (Theorem 6 in Zhang and Chen 2021c and Theorem 4.1 in Zhang and Chen 2022):

- $\texttt{Haus}\left(\mathcal{M}, \widehat{\mathcal{M}}\right) = O(h^2) + O_P\left(\sqrt{\frac{1}{nh^{q+2}}}\right)$, as $h \to 0$ and $nh^{q+2} \to \infty$,
- $\texttt{Haus}\left(\mathcal{R}_d, \widehat{\mathcal{R}}_d\right) = O(h^2) + O_P\left(\sqrt{\frac{|\log h|}{nh^{q+4}}}\right)$, as $h \to 0$ and $\frac{nh^{q+6}}{|\log h|} \to \infty$,

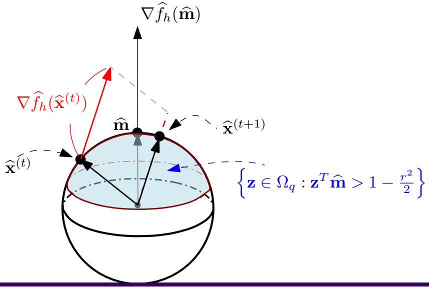
where $\texttt{Haus}(A, B) = \max\left\{ r > 0 : \sup_{\boldsymbol{x} \in A} d(\boldsymbol{x}, B), \sup_{\boldsymbol{y} \in B} d(\boldsymbol{y}, A) \right\}$.

How do we identify the sets of directional local modes $\widehat{\mathcal{M}}$ in practice?

How do we identify the sets of directional local modes $\widehat{\mathcal{M}}$ in practice?

▶ We develop the directional mean shift procedure to estimate $\widehat{\mathcal{M}}$ as (Section 3 in Zhang and Chen 2021c):

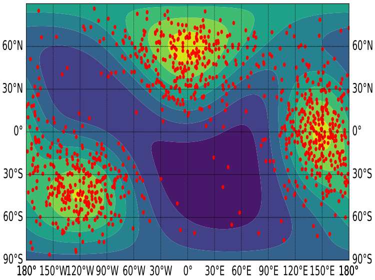
$$\widehat{\boldsymbol{x}}^{(t+1)} = -\frac{\sum_{i=1}^{n} \boldsymbol{X}_i L'\left(\frac{1-\boldsymbol{X}_i^T \widehat{\boldsymbol{x}}^{(t)}}{h^2}\right)}{\left\|\sum_{i=1}^{n} \boldsymbol{X}_i L'\left(\frac{1-\boldsymbol{X}_i^T \widehat{\boldsymbol{x}}^{(t)}}{h^2}\right)\right\|_2} = \frac{\nabla \widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})}{\left\|\nabla \widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})\right\|_2} \quad \text{for } t = 0, 1, \ldots.$$



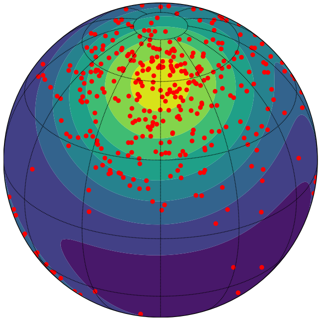
We simulate 1000 data points from the following density

$$f_3(\boldsymbol{x}) = 0.3 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_1, \nu_1) + 0.3 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_2, \nu_2) + 0.4 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_3, \nu_3)$$

with $\boldsymbol{\mu}_1 = [-120°, -45°]$, $\boldsymbol{\mu}_2 = [0°, 60°]$, $\boldsymbol{\mu}_3 = [150°, 0°]$, and $\nu_1 = \nu_2 = 8$, $\nu_3 = 5$.



(a) Step 0

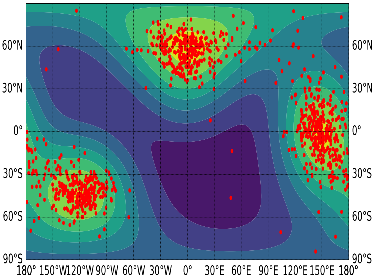


(b) Step 0

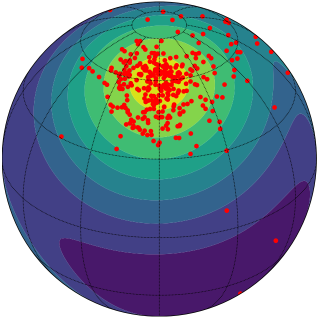
We simulate 1000 data points from the following density

$$f_3(\boldsymbol{x}) = 0.3 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_1, \nu_1) + 0.3 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_2, \nu_2) + 0.4 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_3, \nu_3)$$

with $\boldsymbol{\mu}_1 = [-120°, -45°]$, $\boldsymbol{\mu}_2 = [0°, 60°]$, $\boldsymbol{\mu}_3 = [150°, 0°]$, and $\nu_1 = \nu_2 = 8$, $\nu_3 = 5$.



(a) Step 1

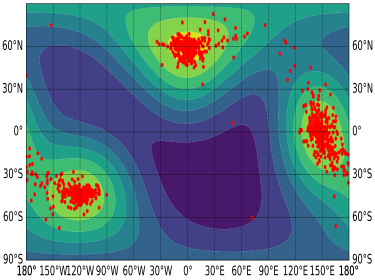


(b) Step 1

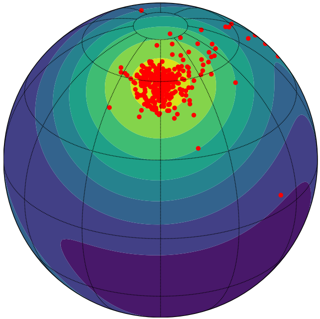
We simulate 1000 data points from the following density

$$f_3(\boldsymbol{x}) = 0.3 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_1, \nu_1) + 0.3 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_2, \nu_2) + 0.4 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_3, \nu_3)$$

with $\boldsymbol{\mu}_1 = [-120°, -45°]$, $\boldsymbol{\mu}_2 = [0°, 60°]$, $\boldsymbol{\mu}_3 = [150°, 0°]$, and $\nu_1 = \nu_2 = 8$, $\nu_3 = 5$.



(a) Step 2

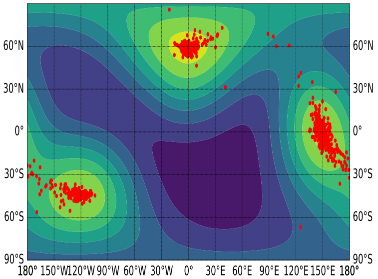


(b) Step 2

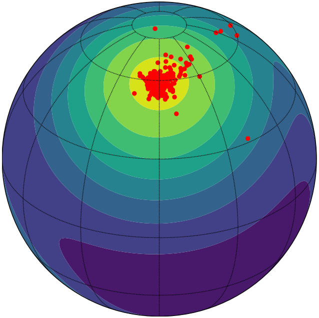
We simulate 1000 data points from the following density

$$f_3(\boldsymbol{x}) = 0.3 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_1, \nu_1) + 0.3 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_2, \nu_2) + 0.4 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_3, \nu_3)$$

with $\boldsymbol{\mu}_1 = [-120^\circ, -45^\circ]$, $\boldsymbol{\mu}_2 = [0^\circ, 60^\circ]$, $\boldsymbol{\mu}_3 = [150^\circ, 0^\circ]$, and $\nu_1 = \nu_2 = 8$, $\nu_3 = 5$.



(a) Step 3



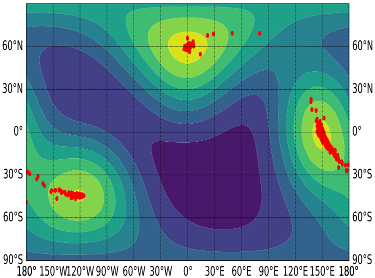
(b) Step 3

Directional Mean Shift: Simulation Study

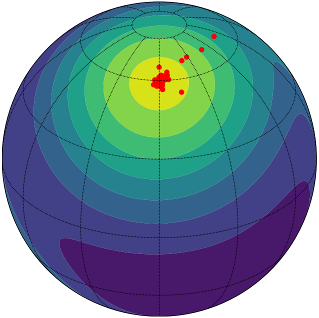
We simulate 1000 data points from the following density

$$f_3(\boldsymbol{x}) = 0.3 \cdot f_{\mathrm{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_1, \nu_1) + 0.3 \cdot f_{\mathrm{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_2, \nu_2) + 0.4 \cdot f_{\mathrm{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_3, \nu_3)$$

with $\boldsymbol{\mu}_1 = [-120°, -45°]$, $\boldsymbol{\mu}_2 = [0°, 60°]$, $\boldsymbol{\mu}_3 = [150°, 0°]$, and $\nu_1 = \nu_2 = 8$, $\nu_3 = 5$.



(a) Step 5

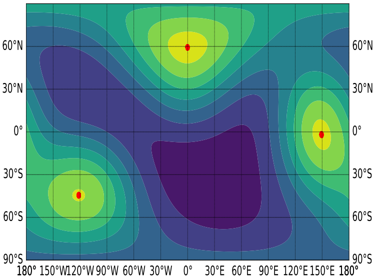


(b) Step 5

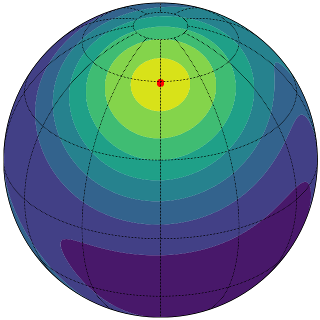
We simulate 1000 data points from the following density

$$f_3(\boldsymbol{x}) = 0.3 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_1, \nu_1) + 0.3 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_2, \nu_2) + 0.4 \cdot f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}_3, \nu_3)$$

with $\boldsymbol{\mu}_1 = [-120^\circ, -45^\circ]$, $\boldsymbol{\mu}_2 = [0^\circ, 60^\circ]$, $\boldsymbol{\mu}_3 = [150^\circ, 0^\circ]$, and $\nu_1 = \nu_2 = 8$, $\nu_3 = 5$.



(a) Step 22 (converged)



(b) Step 22 (converged)

Directional Mean Shift Clustering

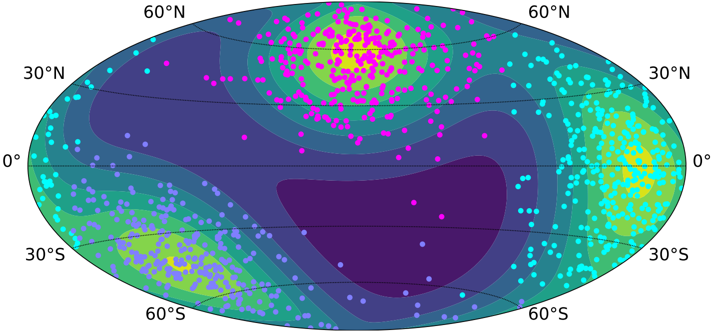
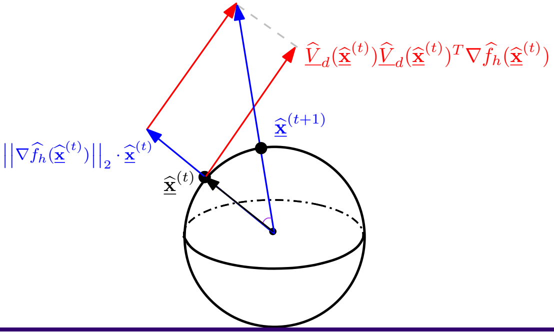


Figure 16: Mode clustering (Hammer projection view)

We also generalize the preceding directional mean shift procedure to estimate $\widehat{\mathcal{R}}_d$ in practice as the directional subspace constrained mean shift (DirSCMS) algorithm (Section 4.2 in Zhang and Chen 2022):

$$\widehat{\boldsymbol{x}}^{(t+1)} \leftarrow \widehat{\boldsymbol{x}}^{(t)} + \widehat{V}_d(\widehat{\boldsymbol{x}}^{(t)})\widehat{V}_d(\widehat{\boldsymbol{x}}^{(t)})^T \cdot \frac{\nabla \widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})}{\left|\left|\nabla \widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})\right|\right|_2} \quad \text{and} \quad \widehat{\boldsymbol{x}}^{(t+1)} \leftarrow \frac{\widehat{\boldsymbol{x}}^{(t+1)}}{\left|\left|\widehat{\boldsymbol{x}}^{(t+1)}\right|\right|_2},$$

for $t = 0, 1, \ldots$.



DirSCMS Algorithm: Simulation Study

We simulate 2000 data points from a circle on Ω_2 with additive Gaussian noises $\mathcal{N}(0, 0.1^2)$ on their Cartesian coordinates and L_2 normalization.

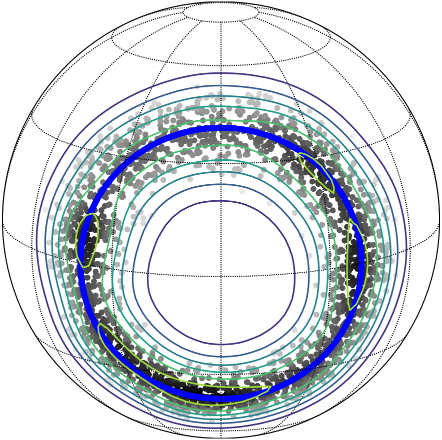


Figure 17: The underlying circle (blue curve) and sampled points (gray dots) on Ω_2.

DirSCMS Algorithm: Simulation Study

We simulate 2000 data points from a circle on Ω_2 with additive Gaussian noises $\mathcal{N}(0, 0.1^2)$ on their Cartesian coordinates and L_2 normalization.

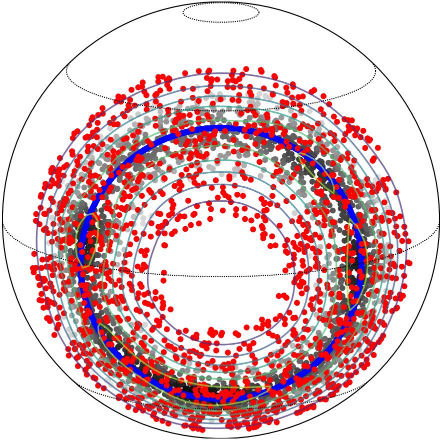


Figure 17: Directional SCMS at Step 0

DirSCMS Algorithm: Simulation Study

We simulate 2000 data points from a circle on Ω_2 with additive Gaussian noises $\mathcal{N}(0, 0.1^2)$ on their Cartesian coordinates and L_2 normalization.

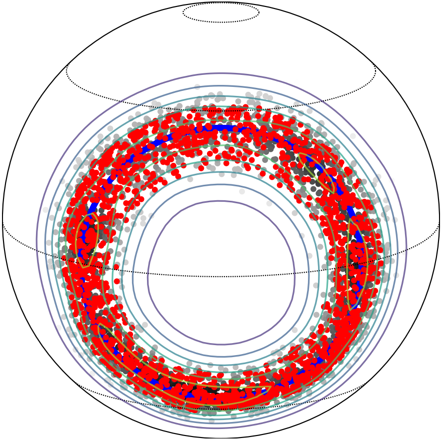


Figure 17: Directional SCMS at Step 1

We simulate 2000 data points from a circle on Ω_2 with additive Gaussian noises $\mathcal{N}(0, 0.1^2)$ on their Cartesian coordinates and L_2 normalization.

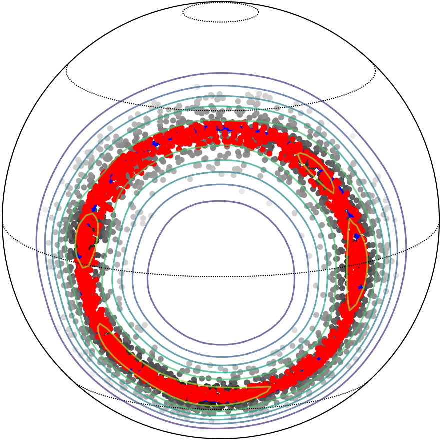


Figure 17: Directional SCMS at Step 2

DirSCMS Algorithm: Simulation Study

We simulate 2000 data points from a circle on Ω_2 with additive Gaussian noises $\mathcal{N}(0, 0.1^2)$ on their Cartesian coordinates and L_2 normalization.

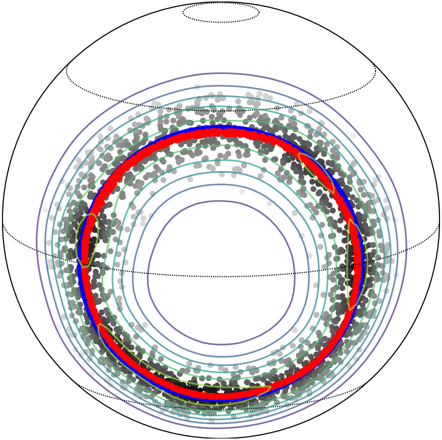


Figure 17: Directional SCMS at Step 4

DirSCMS Algorithm: Simulation Study

We simulate 2000 data points from a circle on Ω_2 with additive Gaussian noises $\mathcal{N}(0, 0.1^2)$ on their Cartesian coordinates and L_2 normalization.

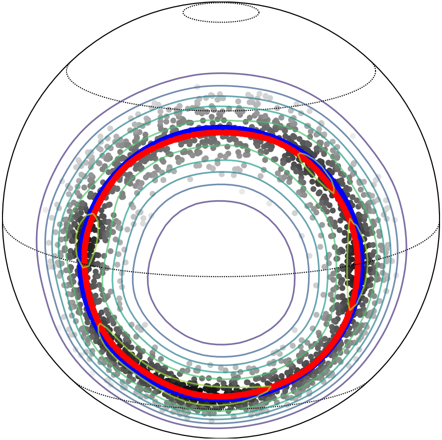


Figure 17: Directional SCMS at Step 8

We simulate 2000 data points from a circle on Ω_2 with additive Gaussian noises $\mathcal{N}(0, 0.1^2)$ on their Cartesian coordinates and L_2 normalization.

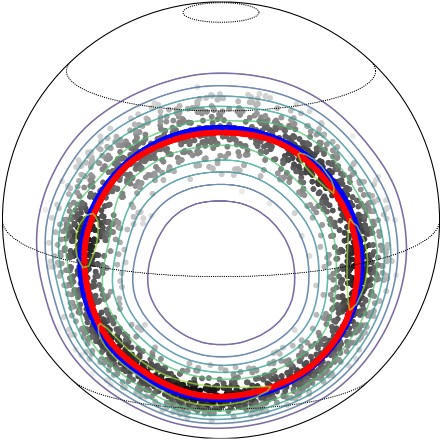


Figure 17: Directional SCMS at Step 24 (converged)

An Extension to Directional-Linear Product Spaces

Recall that the observed galactic data $\{(\phi_i, \eta_i, z_i)\}_{i=1}^n \subset \Omega_2 \times \mathbb{R}^+$ are directional-linear. We consider estimating the density ridges (and local modes) in a directional-linear product space (Zhang and Chen, 2021a).

An Extension to Directional-Linear Product Spaces

Recall that the observed galactic data $\{(\phi_i, \eta_i, z_i)\}_{i=1}^n \subset \Omega_2 \times \mathbb{R}^+$ are directional-linear. We consider estimating the density ridges (and local modes) in a directional-linear product space (Zhang and Chen, 2021a).

- Density estimation at $(\boldsymbol{x}, z) \in \Omega_q \times \mathbb{R}$ (García-Portugués et al., 2015):

$$\widehat{f}_{\boldsymbol{h}}(\boldsymbol{x}, z) = \frac{C_L(h_1)}{nh_2} \sum_{i=1}^n L\left(\frac{1 - \boldsymbol{x}^T \boldsymbol{X}_i}{h_1^2}\right) K\left(\frac{z - z_i}{h_2}\right).$$

An Extension to Directional-Linear Product Spaces

Recall that the observed galactic data $\{(\phi_i, \eta_i, z_i)\}_{i=1}^n \subset \Omega_2 \times \mathbb{R}^+$ are directional-linear. We consider estimating the density ridges (and local modes) in a directional-linear product space (Zhang and Chen, 2021a).

- Density estimation at $(\boldsymbol{x}, z) \in \Omega_q \times \mathbb{R}$ (García-Portugués et al., 2015):

$$\widehat{f}_{\boldsymbol{h}}(\boldsymbol{x}, z) = \frac{C_L(h_1)}{nh_2} \sum_{i=1}^n L\left(\frac{1-\boldsymbol{x}^T\boldsymbol{X}_i}{h_1^2}\right) K\left(\frac{z-z_i}{h_2}\right).$$

- Mode-seeking via mean shift on $\boldsymbol{y}^{(t)} = (\boldsymbol{x}^{(t)}, z^{(t)})$:

$$\boldsymbol{y}^{(t+1)} \leftarrow \Xi(\boldsymbol{y}^{(t)}) + \boldsymbol{y}^{(t)} = \begin{pmatrix} \frac{\sum_{i=1}^n \boldsymbol{X}_i L'\left(\frac{1-\boldsymbol{X}_i^T\boldsymbol{x}^{(t)}}{h_1^2}\right) K\left(\frac{z^{(t)}-z_i}{h_2}\right)}{\sum_{i=1}^n L'\left(\frac{1-\boldsymbol{X}_i^T\boldsymbol{x}^{(t)}}{h_1^2}\right) K\left(\frac{z^{(t)}-z_i}{h_2}\right)} \\ \frac{\sum_{i=1}^n z_i L\left(\frac{1-\boldsymbol{X}_i^T\boldsymbol{x}^{(t)}}{h_1^2}\right) K'\left(\frac{z^{(t)}-z_i}{h_2}\right)}{\sum_{i=1}^n L\left(\frac{1-\boldsymbol{X}_i^T\boldsymbol{x}^{(t)}}{h_1^2}\right) K'\left(\frac{z^{(t)}-z_i}{h_2}\right)} \end{pmatrix}$$

with an extra standardization $\boldsymbol{x}^{(t+1)} \leftarrow \frac{\boldsymbol{x}^{(t+1)}}{\left|\left|\boldsymbol{x}^{(t+1)}\right|\right|_2}$.

An Extension to Directional-Linear Product Spaces

- Ridge-Finding via SCMS algorithm on $\boldsymbol{y}^{(t)} = (\boldsymbol{x}^{(t)}, z^{(t)})$ as:

$$\boldsymbol{y}^{(t+1)} \leftarrow \boldsymbol{y}^{(t)} + \eta \cdot \widehat{V}_d(\boldsymbol{y}^{(t)})\widehat{V}_d(\boldsymbol{y}^{(t)})^T \boldsymbol{H}^{-1}\Xi(\boldsymbol{y}^{(t)}),$$

where

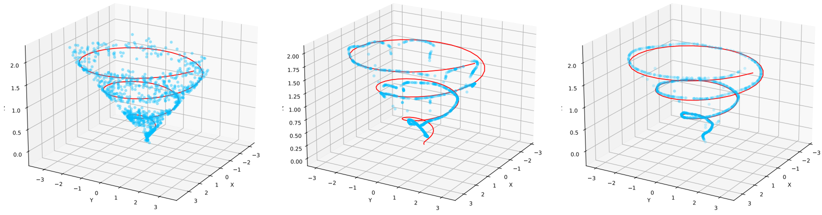
$$\Xi(\boldsymbol{y}) = \begin{pmatrix} \Xi_{\boldsymbol{x}}(\boldsymbol{x}, z) \\ \Xi_z(\boldsymbol{x}, z) \end{pmatrix} = \begin{pmatrix} \frac{\sum_{i=1}^n \boldsymbol{X}_i L'\left(\frac{1-\boldsymbol{X}_i^T \boldsymbol{x}^{(t)}}{h_1^2}\right) K\left(\frac{z^{(t)}-z_i}{h_2}\right)}{\sum_{i=1}^n L'\left(\frac{1-\boldsymbol{X}_i^T \boldsymbol{x}^{(t)}}{h_1^2}\right) K\left(\frac{z^{(t)}-z_i}{h_2}\right)} - \boldsymbol{x} \\ \frac{\sum_{i=1}^n z_i L\left(\frac{1-\boldsymbol{X}_i^T \boldsymbol{x}^{(t)}}{h_1^2}\right) K'\left(\frac{z^{(t)}-z_i}{h_2}\right)}{\sum_{i=1}^n L\left(\frac{1-\boldsymbol{X}_i^T \boldsymbol{x}^{(t)}}{h_1^2}\right) K'\left(\frac{z^{(t)}-z_i}{h_2}\right)} - z \end{pmatrix}.$$

Here, we design a theoretically motivated and empirically effective step size as $\eta = \min\{\max(\boldsymbol{h}) \cdot \min(\boldsymbol{h}), 1\} = \min\{h_1 h_2, 1\}$.

Notes: A naive generalization of SCMS algorithm $\boldsymbol{z}^{(t+1)} \leftarrow \boldsymbol{z}^{(t)} + \widehat{V}_d(\boldsymbol{z}^{(t)})\widehat{V}_d(\boldsymbol{z}^{(t)})^T\Xi(\boldsymbol{z}^{(t)})$ plus standardization as with pure Euclidean/directional data does not work (Zhang and Chen, 2021a)!

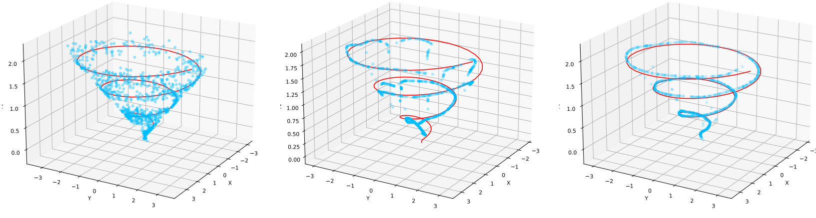
Filament Detection in the Directional-Linear Space

We sample 1000 observations on a spiral curve with additive Gaussian noises $\mathcal{N}(0, 0.2^2)$ to their angular-linear coordinates.



(a) Simulated data points. (b) Euclidean SCMS. (c) Directional-linear SCMS.

We sample 1000 observations on a spiral curve with additive Gaussian noises $\mathcal{N}(0, 0.2^2)$ to their angular-linear coordinates.



(a) Simulated data points. (b) Euclidean SCMS. (c) Directional-linear SCMS.

► Our directional-linear SCMS algorithm is stabler than its Euclidean prototype.

Convergence Results of the Proposed Methods

We prove the (local/global) convergence of our directional mean shift, DirSCMS, and DirLinSCMS algorithms under some mild regularity conditions (Zhang and Chen, 2021c,b, 2022, 2021a).

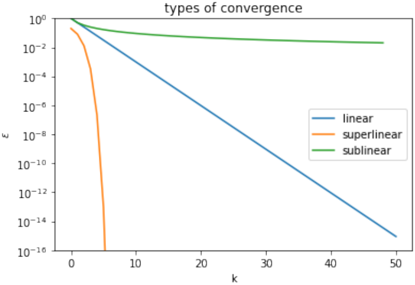
- **Question**: how fast will our proposed algorithms converge?

We prove the (local/global) convergence of our directional mean shift, DirSCMS, and DirLinSCMS algorithms under some mild regularity conditions (Zhang and Chen, 2021c,b, 2022, 2021a).

- **Question**: how fast will our proposed algorithms converge?

Definition (Linear Convergence)

A sequence $\{\boldsymbol{y}_k\}_{k=0,1,..}$ is said to converge *linearly* to $\boldsymbol{y}^*$ if there exists a positive constant $\Upsilon < 1$ (rate of convergence) such that $||\boldsymbol{y}_{k+1} - \boldsymbol{y}^*|| \leq \Upsilon ||\boldsymbol{y}_k - \boldsymbol{y}^*||$ when k is sufficiently large (Boyd et al., 2004).



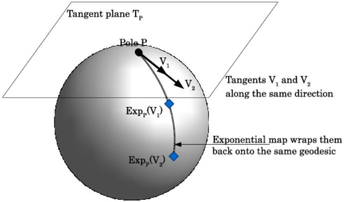
We prove the linear convergence of our proposed algorithms by viewing them as the first-order method and its subspace constrained variant with a (smooth) function f on Ω_q.

We prove the linear convergence of our proposed algorithms by viewing them as the first-order method and its subspace constrained variant with a (smooth) function f on Ω_q.

- Gradient Ascent Algorithm on Ω_q:

$$\boldsymbol{y}_{k+1} = \mathtt{Exp}_{\boldsymbol{y}_k}\left(\eta \cdot \mathtt{grad} f(\boldsymbol{y}_k)\right),$$

where $\eta > 0$ is the step size and $\mathtt{Exp}_{\boldsymbol{x}} : T_{\boldsymbol{x}} \to \Omega_q$ is the *exponential map* at $\boldsymbol{x}$ of a (Riemannian) manifold Ω_q.

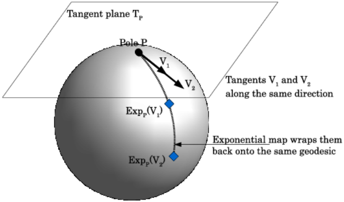


We prove the linear convergence of our proposed algorithms by viewing them as the first-order method and its subspace constrained variant with a (smooth) function f on Ω_q.

- Gradient Ascent Algorithm on Ω_q:

$$\boldsymbol{y}_{k+1} = \mathtt{Exp}_{\boldsymbol{y}_k}\left(\eta \cdot \mathtt{grad} f(\boldsymbol{y}_k)\right),$$

  where $\eta > 0$ is the step size and $\mathtt{Exp}_{\boldsymbol{x}} : T_{\boldsymbol{x}} \to \Omega_q$ is the *exponential map* at $\boldsymbol{x}$ of a (Riemannian) manifold Ω_q.



- Subspace Constrained Gradient Ascent Algorithm on Ω_q: $\boldsymbol{y}_{k+1} = \mathtt{Exp}_{\boldsymbol{y}_k}\left[\eta \cdot V_d(\boldsymbol{y}_k) V_d(\boldsymbol{y}_k)^T \mathtt{grad} f(\boldsymbol{y}_k)\right]$.

Linear Convergence of Gradient Ascent on Ω_q

Under some regularity conditions, we prove the followings (Theorem 12 in Zhang and Chen 2021c):

1. **Linear convergence of gradient ascent with f:** There exists a small radius $r_0 > 0$ such that when the step size $\eta > 0$ is sufficiently small and the initial point $\boldsymbol{y}_0 \in \{\boldsymbol{z} \in \mathbb{M} : d(\boldsymbol{z}, \boldsymbol{m}) < r_0\}$ for some $\boldsymbol{m} \in \Omega_q$,

$$d(\boldsymbol{y}_k, \boldsymbol{m}) \leq \Upsilon^k \cdot d(\boldsymbol{y}_0, \boldsymbol{m}) \quad \text{with} \quad \Upsilon = \sqrt{1 - \frac{\eta \lambda_*}{2}},$$

where $d(\boldsymbol{p}, \boldsymbol{q}) = \left|\left|\mathtt{Exp}_{\boldsymbol{p}}^{-1}(\boldsymbol{q})\right|\right|_2$ and $\lambda_* > 0$ is the eigenvalue bound from 0.

Linear Convergence of Gradient Ascent on Ω_q

Under some regularity conditions, we prove the followings (Theorem 12 in Zhang and Chen 2021c):

1. **Linear convergence of gradient ascent with f:** There exists a small radius $r_0 > 0$ such that when the step size $\eta > 0$ is sufficiently small and the initial point $\boldsymbol{y}_0 \in \{\boldsymbol{z} \in \mathbb{M} : d(\boldsymbol{z}, \boldsymbol{m}) < r_0\}$ for some $\boldsymbol{m} \in \Omega_q$,

$$d(\boldsymbol{y}_k, \boldsymbol{m}) \leq \Upsilon^k \cdot d(\boldsymbol{y}_0, \boldsymbol{m}) \quad \text{with} \quad \Upsilon = \sqrt{1 - \frac{\eta\lambda_*}{2}},$$

   where $d(\boldsymbol{p}, \boldsymbol{q}) = \left|\left|\mathtt{Exp}_{\boldsymbol{p}}^{-1}(\boldsymbol{q})\right|\right|_2$ and $\lambda_* > 0$ is the eigenvalue bound from 0.

2. **Linear convergence of gradient ascent with $\widehat{f}_h$:** let the sample-based gradient ascent update on Ω_q be

$$\widehat{\boldsymbol{y}}_{k+1} = \mathtt{Exp}_{\boldsymbol{y}_k}\left(\eta \cdot \mathtt{grad}\,\widehat{f}_h(\widehat{\boldsymbol{y}}_k)\right).$$

   When the step size $\eta > 0$ is sufficiently small and the initial point $\widehat{\boldsymbol{y}}_0 \in \{\boldsymbol{z} \in \Omega_q : d(\boldsymbol{z}, \boldsymbol{m}) < r_0\}$ for some $\boldsymbol{m} \in \mathcal{M}$,

$$d\left(\widehat{\boldsymbol{y}}_k, \boldsymbol{m}\right) \leq \Upsilon^k \cdot d\left(\widehat{\boldsymbol{y}}_0, \boldsymbol{m}\right) + O(h^2) + O_P\left(\sqrt{\frac{|\log h|}{nh^{q+2}}}\right)$$

   with probability tending to 1, as $h \to 0$ and $\frac{nh^{q+2}}{|\log h|} \to \infty$.

Python Implementation: SCONCE-SCMS

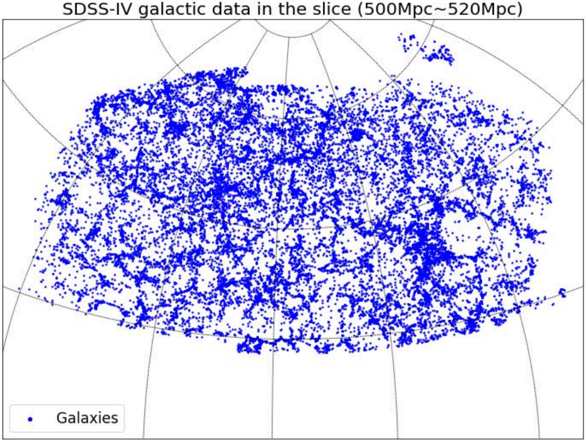
All of our proposed methods are encapsulated in a Python package called **SCONCE-SCMS** (**S**pherical and **CON**ic **C**osmic w**E**b finder with the extended **SCMS** algorithms; Zhang et al. 2022).

- **Python Package Index**: https://pypi.org/project/sconce-scms/.
- **Documentation**: https://sconce-scms.readthedocs.io/en/latest/.

SDSS-IV Cosmic Web Catalog

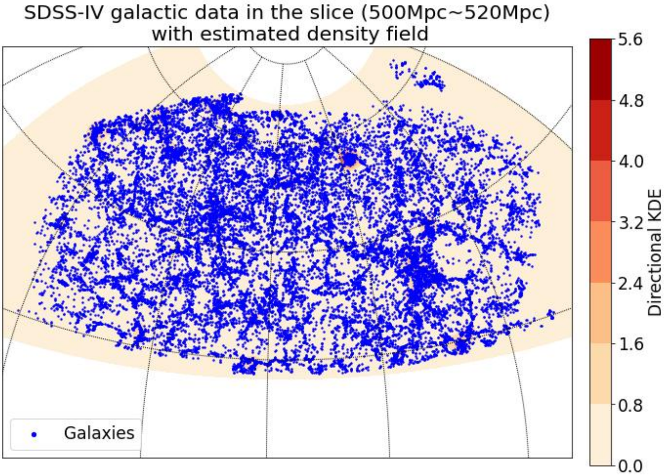
Step 1 (Slicing the Universe): Partition the redshift range into 325 spherical slices based on the comoving distance $\Delta L = 20\,\text{Mpc}$.

- Within each slice, we consider the redshifts of galaxies to be the same so that the galaxies are located on Ω_2.



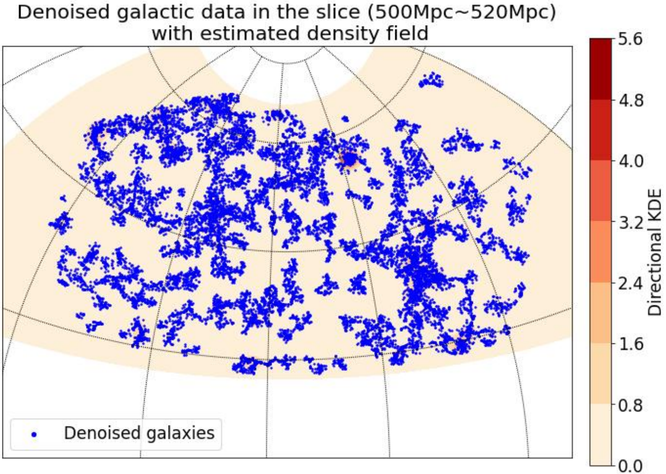
Step 2 (Density Estimation): Estimate the galaxy density field within each spherical slice by directional KDE.

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

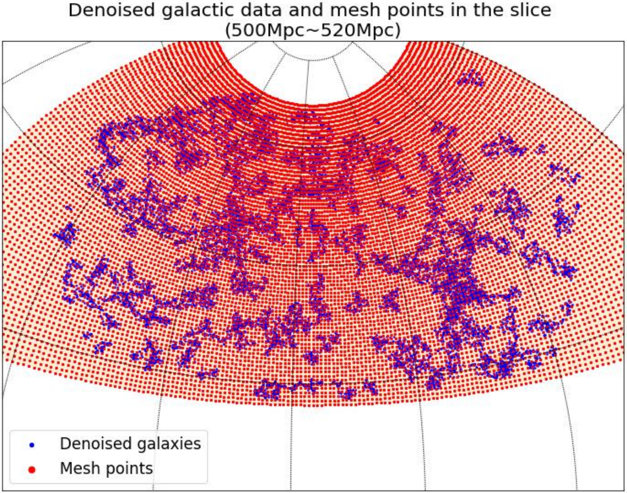


Step 3 (Denoising): Remove the observations with low-density values.

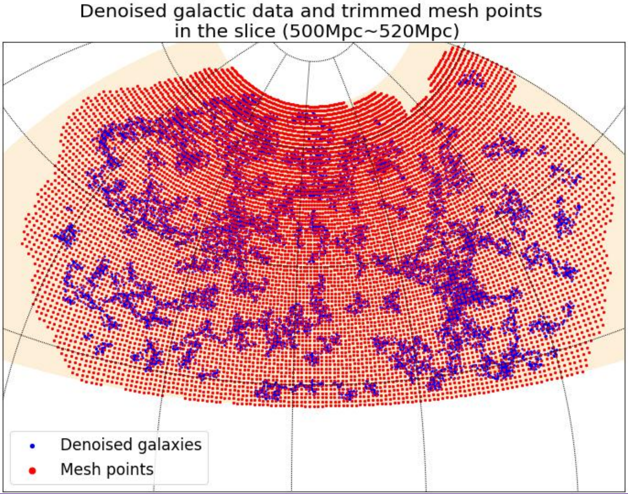
- We keep at least 80% of the original galaxy data in the slice.



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.

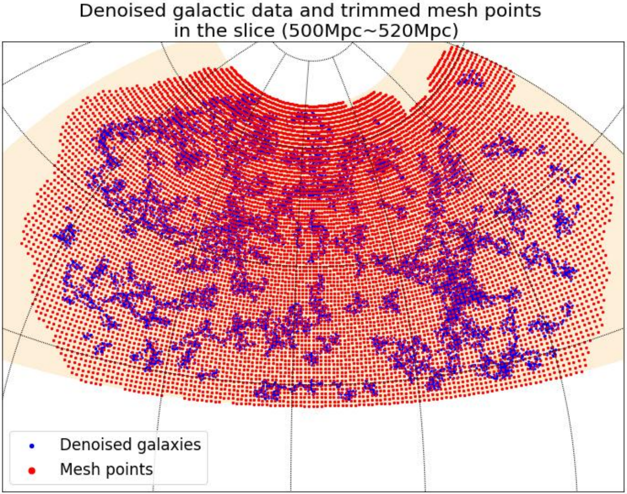


Figure 19: DirSCMS Iterations (Step 0).

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

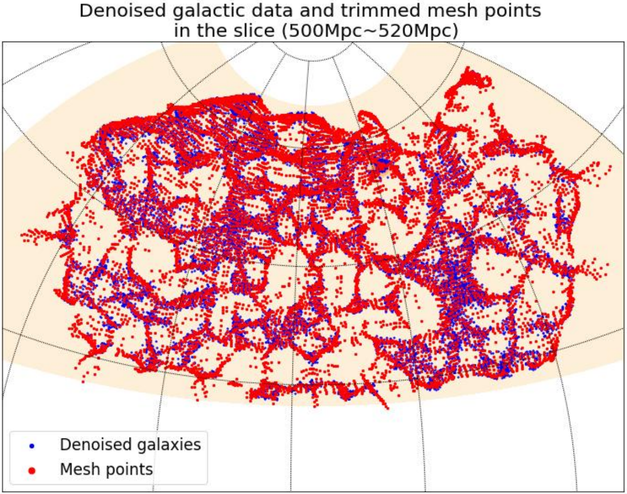


Figure 19: DirSCMS Iterations (Step 1).

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

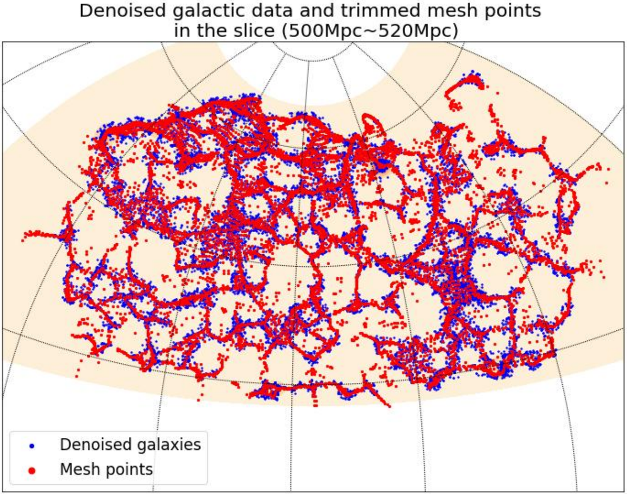


Figure 19: DirSCMS Iterations (Step 2).

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

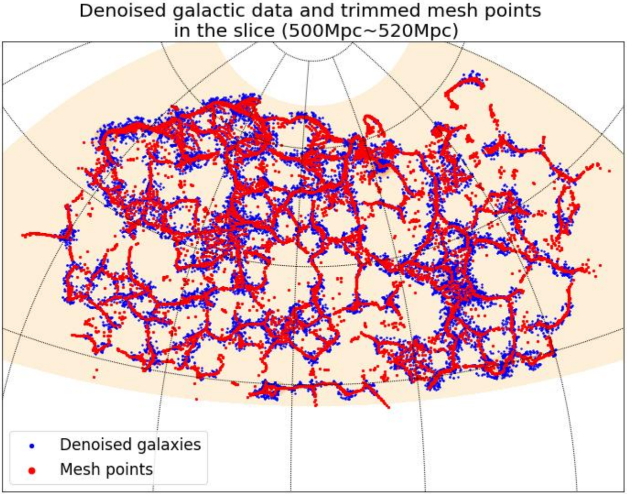


Figure 19: DirSCMS Iterations (Step 3).

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

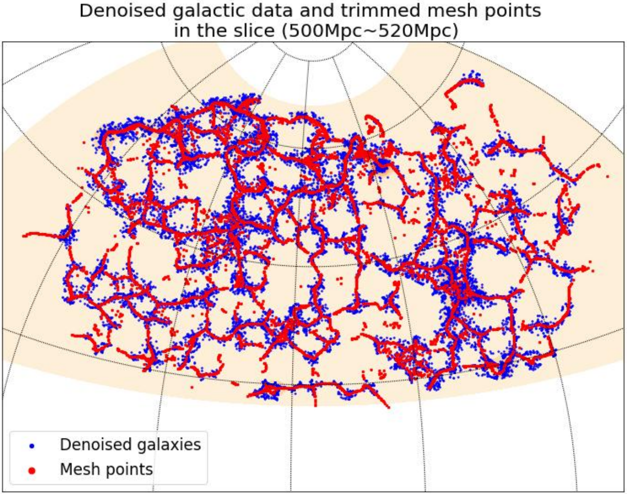


Figure 19: DirSCMS Iterations (Step 5).

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

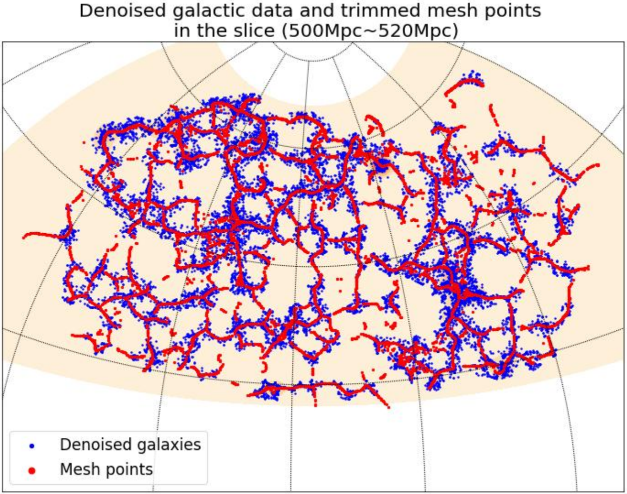


Figure 19: DirSCMS Iterations (Step 8).

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

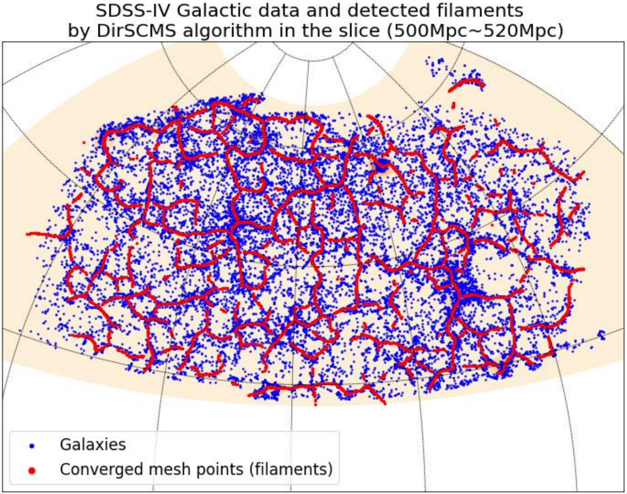


Figure 19: DirSCMS Iterations (Final).

Step 7 (Mode and Knot Estimation): We seek out the local modes and knots on the filaments as cosmic nodes.

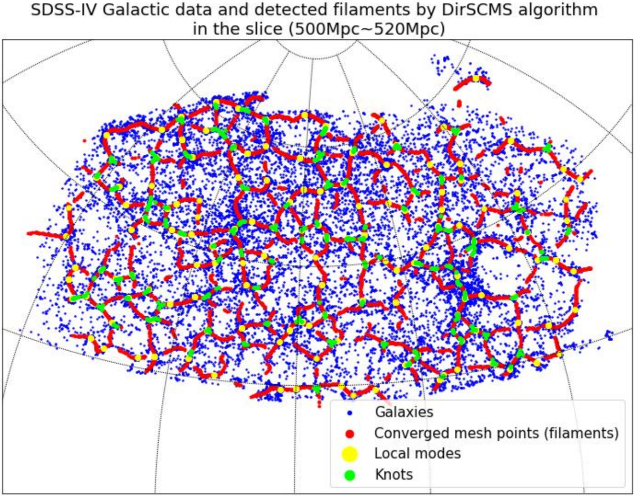
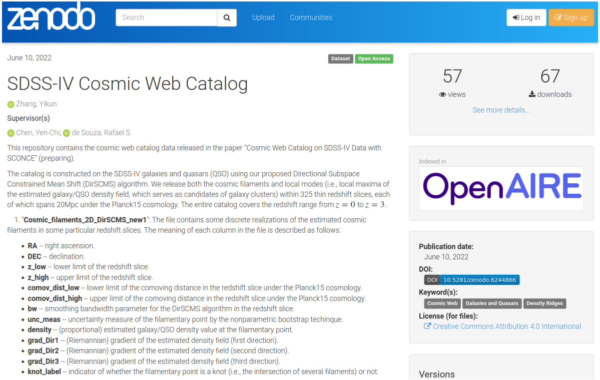


Figure 20: Nodes on the detected filaments.

Final Cosmic Web Catalog on SDSS-IV Data

- The input data incorporate not only galaxy but also quasar (QSO) observations so as to dive deeper into the Universe.
- We compute the uncertainty measure and other features for each detected filamentary point.
- The final catalog is available at https://doi.org/10.5281/zenodo.6244866.



Conclusion and Future Works

Conclusions

In this talk, we discussed our method for estimating cosmic filament structures from observed galactic data and its statistical theory.

Conclusions

In this talk, we discussed our method for estimating cosmic filament structures from observed galactic data and its statistical theory.

1. The cosmic filaments is modeled by directional density ridges, which can be consistently estimated by directional KDE.

Conclusions

In this talk, we discussed our method for estimating cosmic filament structures from observed galactic data and its statistical theory.

1. The cosmic filaments is modeled by directional density ridges, which can be consistently estimated by directional KDE.

2. We designed an efficient algorithm (DirSCMS) to find the directional density ridges in practical applications.

Conclusions

In this talk, we discussed our method for estimating cosmic filament structures from observed galactic data and its statistical theory.

1. The cosmic filaments is modeled by directional density ridges, which can be consistently estimated by directional KDE.
2. We designed an efficient algorithm (DirSCMS) to find the directional density ridges in practical applications.
3. The cosmic web catalog based on our proposed method is publicly available.

Conclusions

In this talk, we discussed our method for estimating cosmic filament structures from observed galactic data and its statistical theory.

1. The cosmic filaments is modeled by directional density ridges, which can be consistently estimated by directional KDE.
2. We designed an efficient algorithm (DirSCMS) to find the directional density ridges in practical applications.
3. The cosmic web catalog based on our proposed method is publicly available.

Future Work: Cosmic Void Detection

Along this line of research, we are planning to

- Leverage our cosmic filament catalog to identify cosmic voids and infer the precise cosmology (Sánchez et al., 2016).

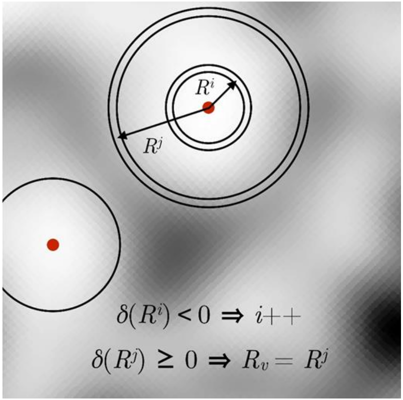
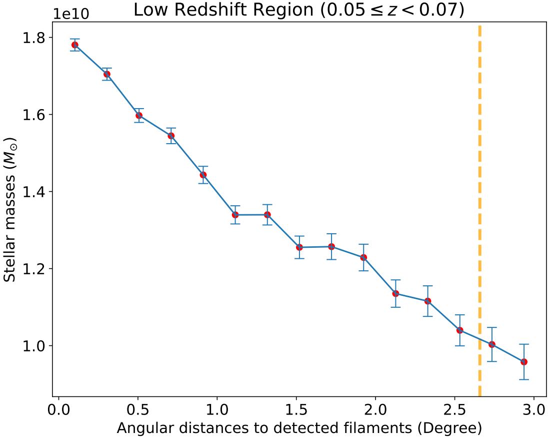


Figure 21: Simple void-finding algorithm (Sánchez et al., 2016).

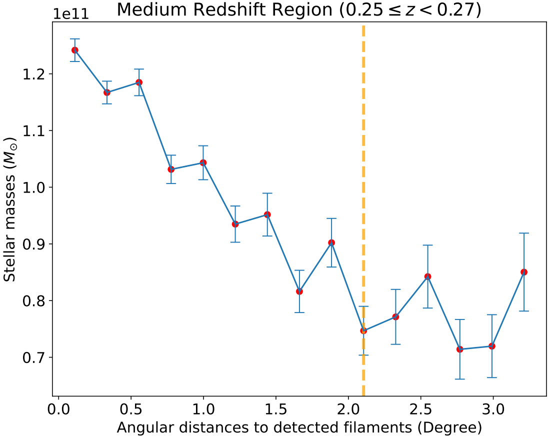
Future Work: Filament Effects of Galaxy Properties

- Analyze if galaxy properties, such as stellar mass, color, and star formation rate, are correlated with our detected cosmic web structures (Chen et al., 2017; Kotecha, 2020)...



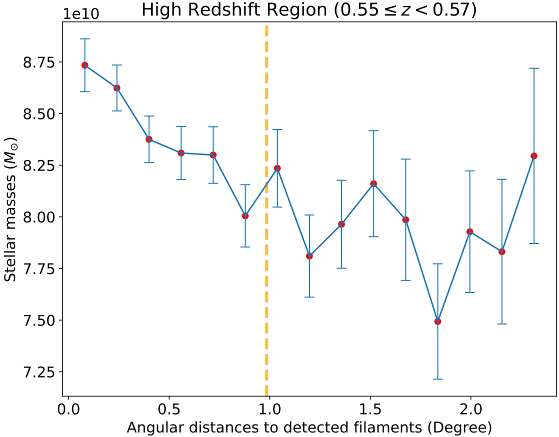
Future Work: Filament Effects of Galaxy Properties

- Analyze if galaxy properties, such as stellar mass, color, and star formation rate, are correlated with our detected cosmic web structures (Chen et al., 2017; Kotecha, 2020)...



Future Work: Filament Effects of Galaxy Properties

- Analyze if galaxy properties, such as stellar mass, color, and star formation rate, are correlated with our detected cosmic web structures (Chen et al., 2017; Kotecha, 2020)...



Thank you!

More details can be found in

[1] Y. Zhang and Y.-C. Chen. Kernel Smoothing, Mean Shift, and Their Learning Theory with Directional Data. *Journal of Machine Learning Research*, 22(154):1–92, 2021. https://arxiv.org/abs/2010.13523

[2] Y. Zhang and Y.-C. Chen. The EM Perspective of Directional Mean Shift Algorithm. arXiv preprint, 2021. https://arxiv.org/abs/2101.10058

[3] Y. Zhang and Y.-C. Chen. Linear Convergence of the Subspace Constrained Mean Shift Algorithm: From Euclidean to Directional Data. *Information and Inference: A Journal of the IMA*, 2022. https://arxiv.org/abs/2104.14977

[4] Y. Zhang and Y.-C. Chen. Mode and Ridge Estimation in Euclidean and Directional Product Spaces: A Mean Shift Approach. arXiv preprint, 2021. https://arxiv.org/abs/2110.08505

[5] Y. Zhang, R. S. de Souza, and Y.-C. Chen. SCONCE: A Cosmic Web Finder for Spherical and Conic Geometries. arXiv preprint, 2022. (To appear in the *Monthly Notices of the Royal Astronomical Society*) https://arxiv.org/abs/2207.07001

Reference I

P.-A. Absil, R. Mahony, and R. Sepulchre. Optimization algorithms on matrix manifolds. In *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.

P. A. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Banday, R. Barreiro, J. Bartlett, N. Bartolo, et al. Planck 2015 results-xiii. cosmological parameters. *Astronomy & Astrophysics*, 594:A13, 2016.

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.

S. Boyd, S. P. Boyd, and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

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, 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.

Reference II

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. C. Duque, M. Migliaccio, D. Marinucci, and N. Vittorio. A novel cosmic filament catalogue from sdss data. *Astronomy & Astrophysics*, 659:A166, 2022.

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

E. García-Portugués, R. M. Crujeiras, and W. González-Manteiga. Central limit theorems for directional and linear random variables with applications. *Statistica Sinica*, pages 1207–1229, 2015.

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.

G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. Bennett, J. Dunkley, M. Nolta, M. Halpern, R. Hill, N. Odegard, et al. Nine-year wilkinson microwave anisotropy probe (wmap) observations: cosmological parameter results. *The Astrophysical Journal Supplement Series*, 208(2):19, 2013.

R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

N. Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227(1):1–21, 1987.

S. Kotecha. *The Effect of Cosmic Web Filaments on Quenching in Galaxy Clusters*. PhD thesis, 2020.

U. Kuchner, A. Aragón-Salamanca, A. Rost, F. R. Pearce, M. E. Gray, W. Cui, A. Knebe, E. Rasia, and G. Yepes. Cosmic filaments in galaxy cluster outskirts: quantifying finding filaments in redshift space. *Monthly Notices of the Royal Astronomical Society*, 503(2):2065–2076, 2021.

Reference III

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.

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.

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

X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

S. Pfeifer, N. I. Libeskind, Y. Hoffman, W. A. Hellwing, M. Bilicki, and K. Naidoo. Cows: A filament finder for hessian cosmic web identifiers. *Monthly Notices of the Royal Astronomical Society*, 514(1): 470–479, 2022.

C. Sánchez, J. Clampitt, A. Kovacs, B. Jain, J. García-Bellido, S. Nadathur, D. Gruen, N. Hamaus, D. Huterer, P. Vielzeuf, et al. Cosmic voids and void lensing in the dark energy survey science verification data. *Monthly Notices of the Royal Astronomical Society*, 465(1):746–759, 10 2016.

W. Sargent and E. Turner. A statistical method for determining the cosmological density parameter from the redshifts of a complete sample of galaxies. *The Astrophysical Journal*, 212:L3–L7, 1977.

T. Sousbie, C. Pichon, and H. Kawahara. The persistent cosmic web and its filamentary structure–ii. illustrations. *Monthly Notices of the Royal Astronomical Society*, 414(1):384–403, 2011.

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.

M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, D. Nelson, and L. Hernquist. Introducing the illustris project: simulating the coevolution of dark and visible matter in the universe. *Monthly Notices of the Royal Astronomical Society*, 444(2):1518–1547, 2014.

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. Mode and ridge estimation in euclidean and directional product spaces: A mean shift approach. *arXiv preprint arXiv:2110.08505*, 2021a. URL https://arxiv.org/abs/2110.08505.

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

Y. Zhang and Y.-C. Chen. Kernel smoothing, mean shift, and their learning theory with directional data. *Journal of Machine Learning Research*, 22(154):1–92, 2021c.

Y. Zhang and Y.-C. Chen. Linear convergence of the subspace constrained mean shift algorithm: From euclidean to directional data. *Information and Inference: A Journal of the IMA*, 2022. URL https://doi.org/10.1093/imaiai/iaac005.

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, R. S. de Souza, and Y.-C. Chen. Sconce: A cosmic web finder for spherical and conic geometries. *arXiv preprint arXiv:2207.07001*, 2022. URL https://arxiv.org/abs/2207.07001.

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.

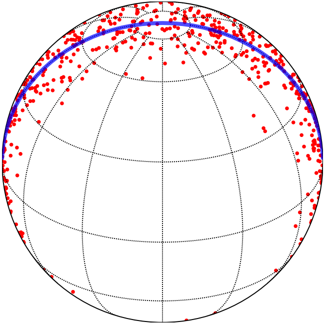
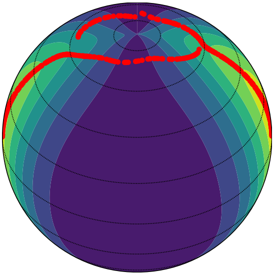


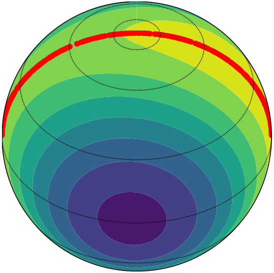
Figure 22: Noisy observations (red points) and the underlying true ring/filament structure (blue line)

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

The background contour plots are kernel density estimators on the flat plane $[-90°, 90°] \times [0°, 360°)$ and unit sphere $\Omega_2 = \left\{ \boldsymbol{x} \in \mathbb{R}^3 : ||\boldsymbol{x}||_2 = 1 \right\}$, respectively.



(a) Euclidean SCMS Method.



(b) Directional SCMS Method.

$*$ Euclidean SCMS method is the original subspace constrained mean shift algorithm proposed by Ozertem and Erdogmus (2011).

Under some regularity conditions (Hall et al., 1987; Bai et al., 1988; García-Portugués, 2013; Zhang and Chen, 2021c), we have

- **Pointwise Consistency**: for any fixed $\boldsymbol{x} \in \Omega_q$,

$$\widehat{f}_h(\boldsymbol{x}) - f(\boldsymbol{x}) = O(h^2) + O_P\left(\sqrt{\frac{1}{nh^q}}\right)$$

as $h \to 0$ and $nh^q \to \infty$;

$$\mathtt{grad}\,\widehat{f}_h(\boldsymbol{x}) - \mathtt{grad}\,f(\boldsymbol{x}) = O(h^2) + O_P\left(\sqrt{\frac{1}{nh^{q+2}}}\right)$$

as $h \to 0$ and $nh^{q+2} \to \infty$;

$$\mathcal{H}\widehat{f}_h(\boldsymbol{x}) - \mathcal{H}f(\boldsymbol{x}) = O(h^2) + O_P\left(\sqrt{\frac{1}{nh^{q+4}}}\right)$$

as $h \to 0$ and $nh^{q+4} \to \infty$.

- **Uniform Consistency**:

$$\|\widehat{f}_h - f\|_\infty = O(h^2) + O_P\left(\sqrt{\frac{\log n}{nh^q}}\right)$$

as $h \to 0$ and $\frac{nh^q}{\log n} \to \infty$;

$$\sup_{\boldsymbol{x}\in\Omega_q} \left|\left|\mathtt{grad}\,\widehat{f}_h(\boldsymbol{x}) - \mathtt{grad}\,f(\boldsymbol{x})\right|\right|_{\max} = O(h^2) + O_P\left(\sqrt{\frac{\log n}{nh^{q+2}}}\right)$$

as $h \to 0$ and $\frac{nh^{q+2}}{\log n} \to \infty$;

$$\sup_{\boldsymbol{x}\in\Omega_q} \left|\left|\mathcal{H}\widehat{f}_h(\boldsymbol{x}) - \mathcal{H}f(\boldsymbol{x})\right|\right|_{\max} = O(h^2) + O_P\left(\sqrt{\frac{\log n}{nh^{q+4}}}\right)$$

as $h \to 0$ and $\frac{nh^{q+4}}{\log n} \to \infty$, where $\|g\|_\infty = \sup_{\boldsymbol{x}\in\Omega_q} |g(\boldsymbol{x})|$ and $||A||_{\max}$ is the elementwise maximum norm for a matrix $A \in \mathbb{R}^{(q+1)\times(q+1)}$.

Detailed Procedures of DirSCMS Algorithm I

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 $T_{\widehat{\boldsymbol{x}}^{(t)}}$.

- **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.

Under some regularity conditions, we prove the following (Theorem 4.6 in Zhang and Chen 2022):

1. **R-Linear convergence of $d(\boldsymbol{x}^{(k)}, \mathcal{R}_d)$ with f.** When the step size $\underline{\eta} > 0$ is sufficiently small and the initial point $\boldsymbol{x}^{(0)}$ lies within a small neighborhood of its limiting point $\boldsymbol{x}^*$ in $\mathcal{R}_d$,

$$d\left(\boldsymbol{x}^{(k)}, \mathcal{R}_d\right) \leq \underline{\Upsilon}^k \cdot d\left(\boldsymbol{x}^{(0)}, \boldsymbol{x}^*\right) \quad \text{with} \quad \underline{\Upsilon} = \sqrt{1 - \frac{\underline{\Upsilon}\beta_0}{4}},$$

where $\beta_0 > 0$ is the eigengap between the d-th and $(d+1)$-th eigenvalues of $\mathcal{H}f(\boldsymbol{x})$.

2. **R-Linear convergence of $d(\widehat{\boldsymbol{x}}^{(k)}, \mathcal{R}_d)$ with $\widehat{f}_h$.** When the step size $\underline{\eta} > 0$ is sufficiently small and the initial point $\widehat{\boldsymbol{x}}^{(0)}$ lies within a small neighborhood of $\boldsymbol{x}^*$ in $\mathcal{R}_d$,

$$d\left(\boldsymbol{x}^{(k)}, \mathcal{R}_d\right) \leq \underline{\Upsilon}^k \cdot d\left(\boldsymbol{x}^{(0)}, \boldsymbol{x}^*\right) + O(h^2) + O_P\left(\sqrt{\frac{|\log h|}{nh^{q+4}}}\right)$$

with probability tending to 1, as $h \to 0$ and $\frac{nh^{q+4}}{|\log h|} \to 0$.

- The linear convergence results can also be proved for the subspace constrained gradient ascent method but under some stricter conditions (Zhang and Chen, 2022).
- The (directional) mean shift and SCMS algorithms can be viewed as variants of the (subspace constrained) gradient ascent methods (on Ω_q) but with adaptive step sizes.
- The step sizes can be made sufficiently small as the bandwidth h is small and the sample size n is large, but also universally bounded away from 0 with respect to the iteration number t.

Application of DirLinSCMS to SDSS-IV Galaxy Data

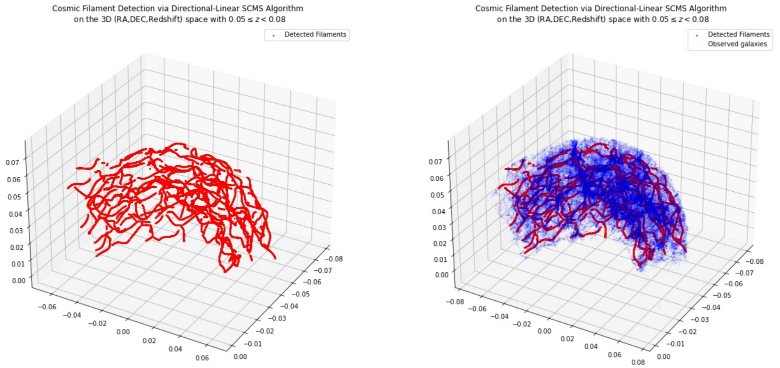


Figure 24: Cosmic filament detection in the 3D (RA,DEC,Redshift) space with our directional-linear SCMS algorithm.