Geometric Data Analysis Reading Group

SCONCE: **S**pherical and **CON**ic **C**osmic W**E**b Detection Through Directional Density Ridges

Yikun Zhang
Joint work with *Yen-Chi Chen*

Department of Statistics,
University of Washington

April 4, 2022

Background: What Is Cosmic Web?

Background: What Is Cosmic Web?

Cosmic Web is a large-scale network structure revealing that the matter in our Universe is not uniformly distributed (Zel'Dovich, 1970; Shandarin and Zeldovich, 1989; Bond et al., 1996).

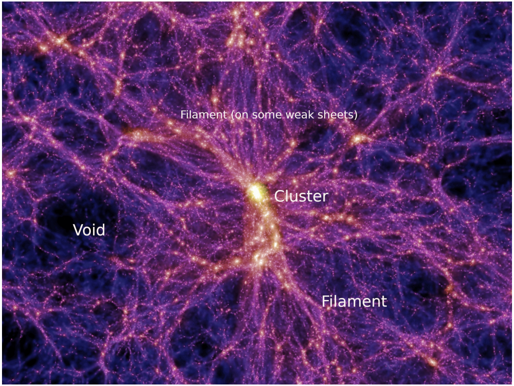


Figure 1: Characteristics of *Cosmic Web* (credited to the millennium simulation project (Springel et al., 2005)).

The Objective of Our Research

Our studies focus on detecting the (one-dimensional) cosmic filaments and (zero-dimensional) cosmic nodes on the filaments from some astronomical survey data.

The Objective of Our Research

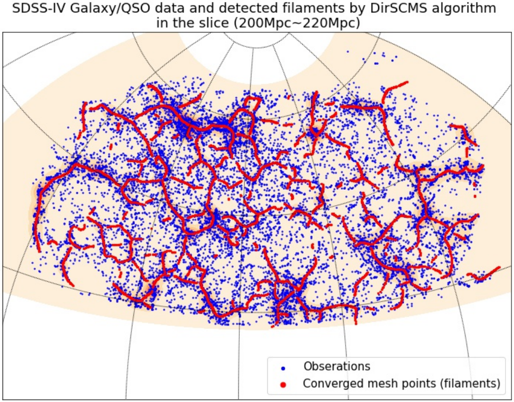
Our studies focus on detecting the (one-dimensional) cosmic filaments and (zero-dimensional) cosmic nodes on the filaments from some astronomical survey data. In particular,

- we propose a statistical model to characterize the cosmic filaments.

The Objective of Our Research

Our studies focus on detecting the (one-dimensional) cosmic filaments and (zero-dimensional) cosmic nodes on the filaments from some astronomical survey data. In particular,

- we propose a statistical model to characterize the cosmic filaments.
- we also develop a fast algorithm to estimate the filamentary structures from a set of discrete observations.



The Significance of Cosmic Filaments

- The filaments has impacts on the stellar properties of their nearby galaxies, such as stellar mass, spinning orientation, and star forming rate (Chen et al., 2017; Malavasi et al., 2022).

The Significance of Cosmic Filaments

- The filaments has impacts on the stellar properties of their nearby galaxies, such as stellar mass, spinning orientation, and star forming rate (Chen et al., 2017; Malavasi et al., 2022).

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

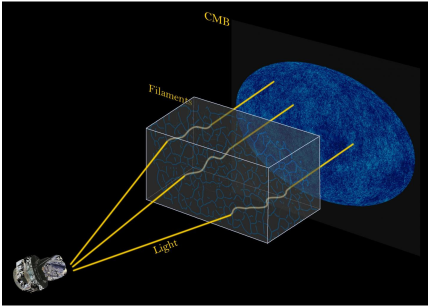


Figure 2: Illustration of the bending trajectory of CMB lights (credit to Siyu He, Shadab Alam, Wei Chen, and Planck/ESA; see He et al. (2018) for details).

Astronomical Survey Data

In astronomical survey data, the positions of observed objects are recorded as

$$\{(\alpha_1, \delta_1, Z_1), ..., (\alpha_n, \delta_n, Z_n)\},$$

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

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

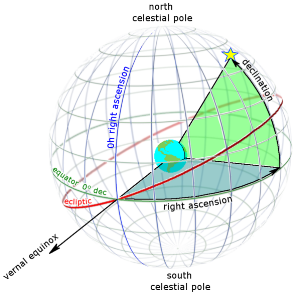


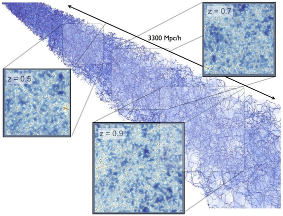
Figure 3: Illustration of RA and DEC (Image Courtesy of Wikipedia).

Challenge: The filamentary structures are overwhelmingly complex (Cautun et al., 2013). The existing methods come from two categories:

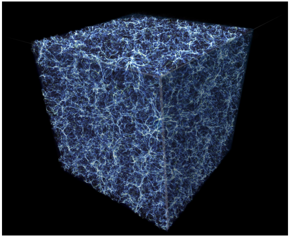
Some Previous Works of Filament Detection

Challenge: The filamentary structures are overwhelmingly complex (Cautun et al., 2013). The existing methods come from two categories:

- **2D method**: Partition the Universe into thin redshift slices (Chen et al., 2015b; Duque et al., 2021).
- **3D method**: Convert redshifts into (comoving) distances (Tempel et al., 2014; Sousbie et al., 2011).



(a) 2D method by slicing the Universe (credit to Laigle et al. 2018).



(b) 3D method in a cubic region.

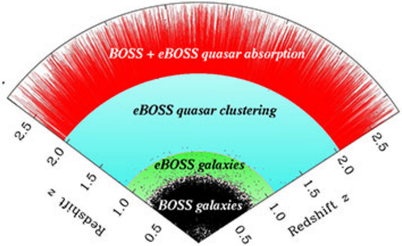
Highlight: Our method can easily switch between these two categories!

Drawback of Existing Methods on 2D Slices

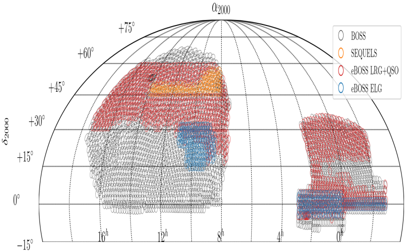
Drawback of Existing Methods on 2D Slices

The slices ($\Delta z = 0.005$) in the survey data are not some flat 2D planes, but some **spherical shells**, which have a *nonlinear* curvature!

- Recall that the locations of astronomical objects in a slice are recorded by $\{(\alpha_i, \delta_i)\}_{i=1}^{n}$ 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)

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.

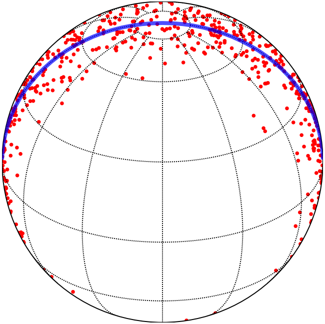
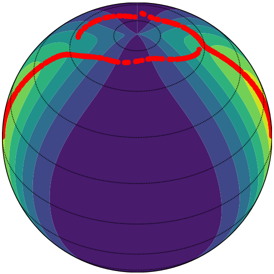


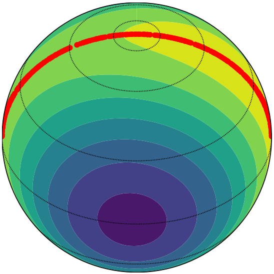
Figure 6: Noisy observations (red points) and the underlying true ring/filament structure (blue line)

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.

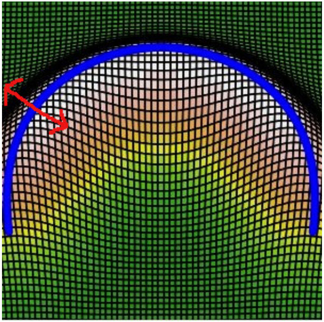


(b) Directional SCMS Method.

$*$ SCMS: subspace constrained mean shift (Ozertem and Erdogmus, 2011).

Our Filament Model: Directional Density Ridges

Directional density ridges are generalized local maxima (within some subspaces) of the underlying density function on the unit hypersphere $\Omega_q = \left\{ \boldsymbol{x} \in \mathbb{R}^{q+1} : ||\boldsymbol{x}||_2 = 1 \right\}$.



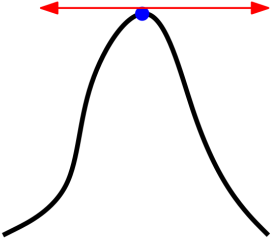


Figure 8: Density ridge (lifted onto the underlying density function; Chen et al. 2015a)

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

- A smooth density function $f : \Omega_q \to \mathbb{R}$. ($q = 2$ in a spherical slice.)
- Riemannian gradient $\texttt{grad} f(\boldsymbol{x})$ and Riemannian Hessian $\mathcal{H}f(\boldsymbol{x})$.
- Denote $V_d(\boldsymbol{x}) = \left[\boldsymbol{v}_{d+1}(\boldsymbol{x}), ..., \boldsymbol{v}_q(\boldsymbol{x})\right] \in \mathbb{R}^{(q+1)\times(q-d)}$ with columns as the last $(q-d)$ eigenvectors of $\mathcal{H}f(\boldsymbol{x})$ lying within the tangent space $T_{\boldsymbol{x}}$ at $\boldsymbol{x} \in \Omega_q$.

$\Longrightarrow$

Formal Definitions of Directional Density Ridges

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

- A smooth density function $f : \Omega_q \to \mathbb{R}$. ($q = 2$ in a spherical slice.)
- Riemannian gradient $\texttt{grad} f(\boldsymbol{x})$ and Riemannian Hessian $\mathcal{H}f(\boldsymbol{x})$.
- Denote $V_d(\boldsymbol{x}) = \left[\boldsymbol{v}_{d+1}(\boldsymbol{x}), ..., \boldsymbol{v}_q(\boldsymbol{x})\right] \in \mathbb{R}^{(q+1)\times(q-d)}$ with columns as the last $(q-d)$ eigenvectors of $\mathcal{H}f(\boldsymbol{x})$ lying within the tangent space $T_{\boldsymbol{x}}$ at $\boldsymbol{x} \in \Omega_q$.

$\Longrightarrow$

Local modes of f on Ω_q:

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

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

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

$*$ 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.

Filament Estimation in Practice

Given some discrete observations $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\} \subset \Omega_q$,

1. **Density Estimation**: We estimate the underlying density function via the *directional* kernel density estimator (KDE; Hall et al. 1987; Bai et al. 1988):

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

   where
   - L is a directional kernel, *e.g.*, the von Mises kernel $L(r) = e^{-r}$,
   - $h > 0$ is the bandwidth, and $c_{L,q}(h)$ is a normalizing constant.

Filament Estimation in Practice

Given some discrete observations $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\} \subset \Omega_q$,

1. **Density Estimation**: We estimate the underlying density function via the *directional* kernel density estimator (KDE; Hall et al. 1987; Bai et al. 1988):

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

   where

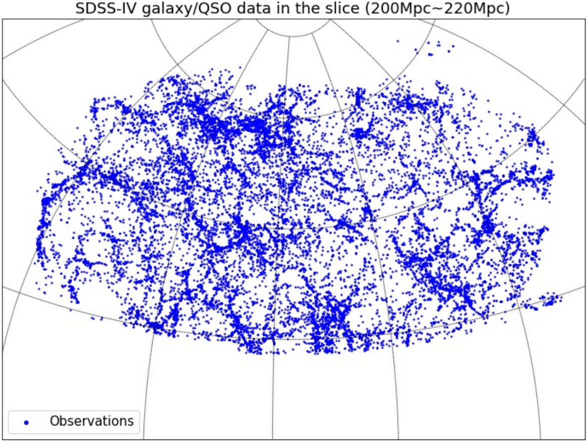
   - L is a directional kernel, *e.g.*, the von Mises kernel $L(r) = e^{-r}$,
   - $h > 0$ is the bandwidth, and $c_{L,q}(h)$ is a normalizing constant.

2. **Filament Estimation**: We propose the directional subspace constrained mean shift (DirSCMS) algorithm (Zhang and Chen, 2021c), which iterates a sequence $\left\{\boldsymbol{x}^{(t)}\right\}_{t=0}^{\infty} \subset \Omega_q$ that converges *linearly* to the density ridges of directional KDE:

$$\widehat{\boldsymbol{x}}^{(t+1)} \leftarrow \widehat{\boldsymbol{x}}^{(t)} - \widehat{V}_d(\widehat{\boldsymbol{x}}^{(t)}) \widehat{V}_d(\widehat{\boldsymbol{x}}^{(t)})^T \left[\frac{\sum_{i=1}^n \boldsymbol{X}_i L'\left(\frac{1-\boldsymbol{X}_i^T \widehat{\boldsymbol{x}}^{(t)}}{h^2}\right)}{\sum_{i=1}^n \boldsymbol{X}_i L'\left(\frac{1-\boldsymbol{X}_i^T \widehat{\boldsymbol{x}}^{(t)}}{h^2}\right)}\right] \text{ with } \boldsymbol{x}^{(t+1)} \leftarrow \frac{\boldsymbol{x}^{(t+1)}}{\left|\left|\boldsymbol{x}^{(t+1)}\right|\right|_2}.$$

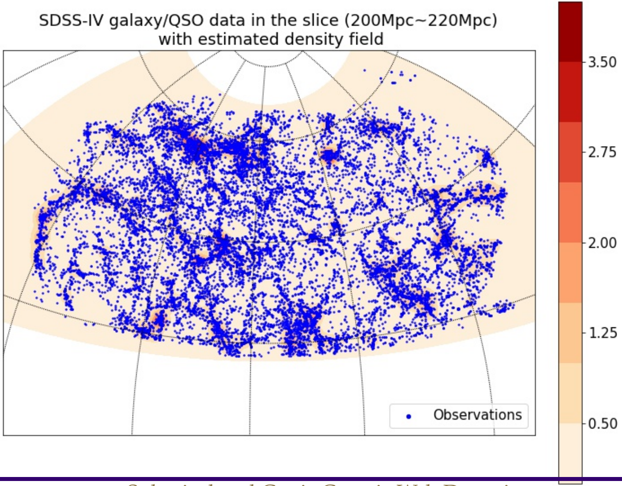
Step 1 (Slicing the Universe): Partition the redshift range into spherical slices based on the comoving distance $\Delta L = 20\,\text{Mpc}$.

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



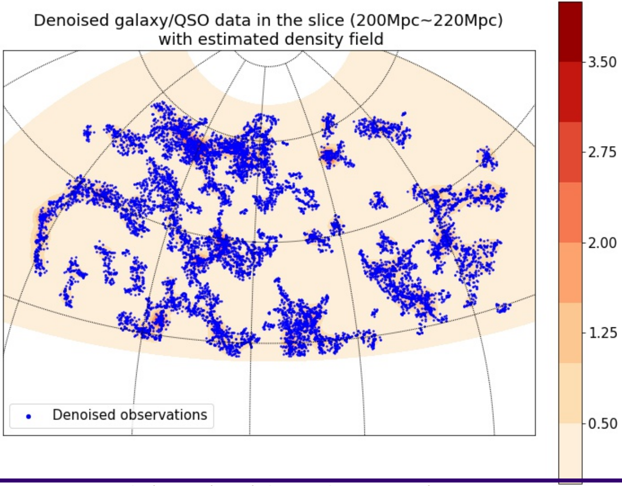
Step 2 (Density Estimation): Estimate the galaxy density field via directional KDE.

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

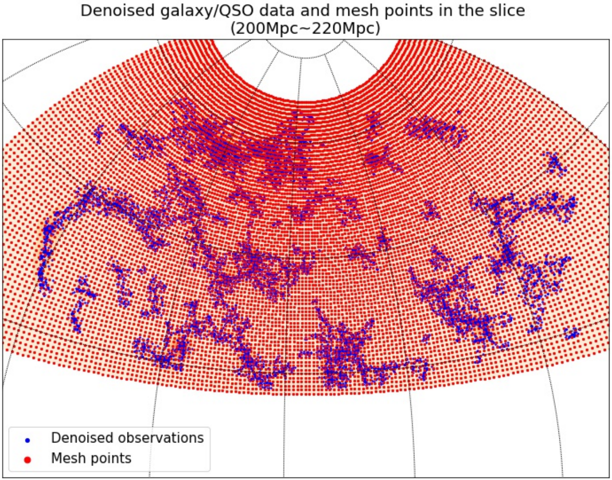


Step 3 (Denoising): Remove the observations with low-density values.

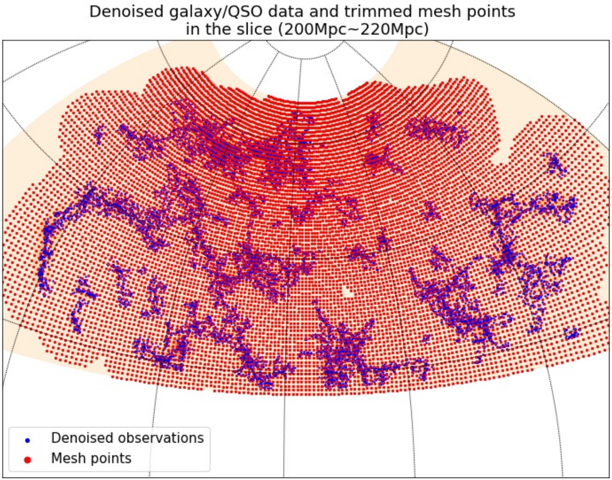
- We keep at least 80% of the original galaxy data in the slice.



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.

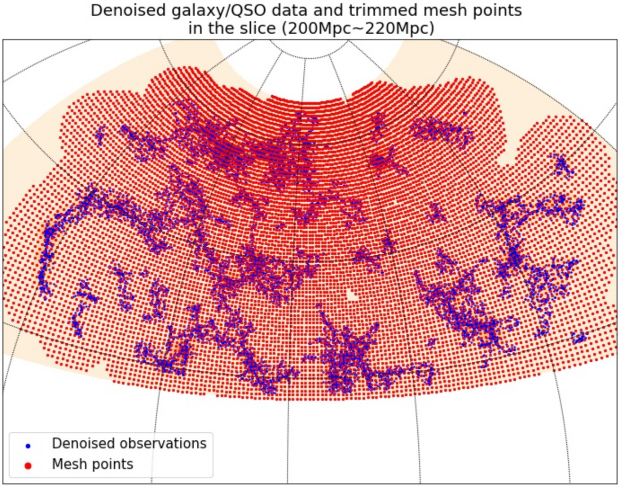


Figure 9: DirSCMS Iterations (Step 0).

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

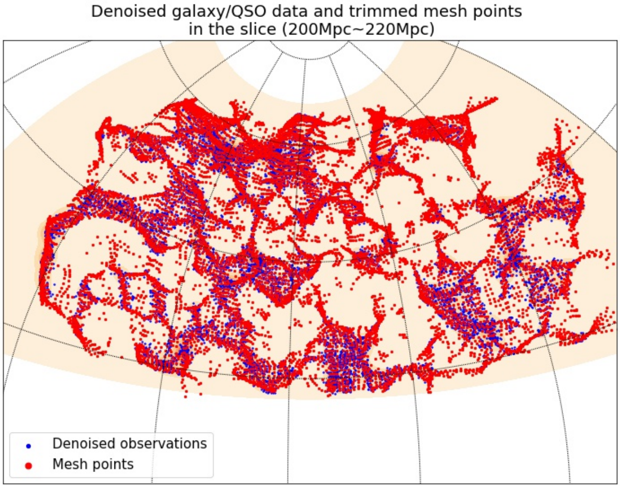


Figure 9: DirSCMS Iterations (Step 1).

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

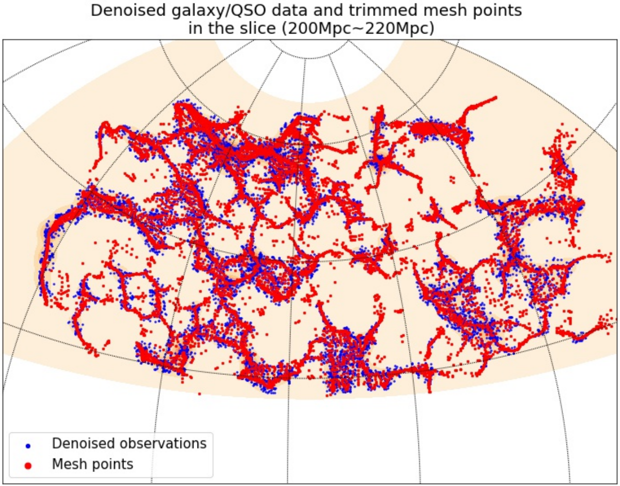


Figure 9: DirSCMS Iterations (Step 2).

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

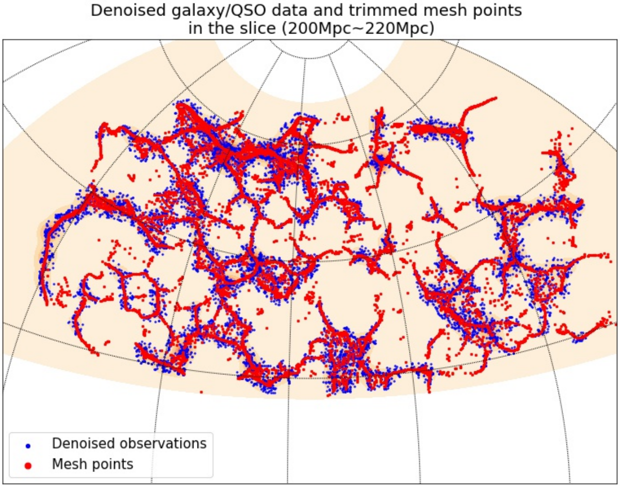


Figure 9: DirSCMS Iterations (Step 3).

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

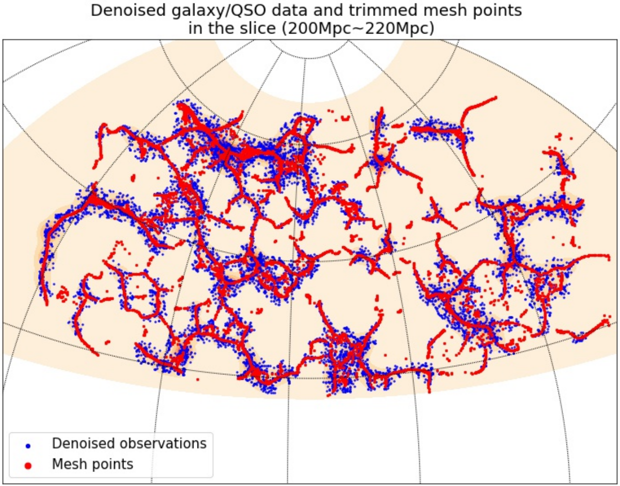


Figure 9: DirSCMS Iterations (Step 5).

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

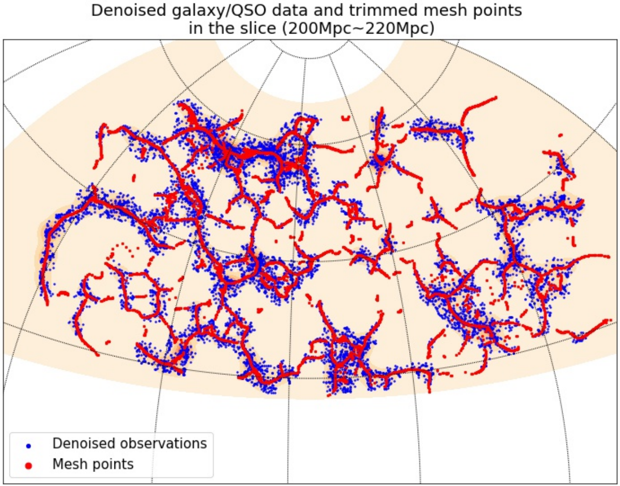


Figure 9: DirSCMS Iterations (Step 8).

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

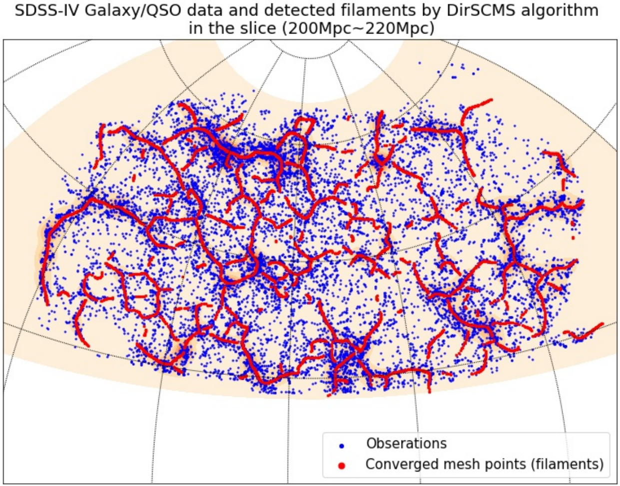


Figure 9: DirSCMS Iterations (Final).

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

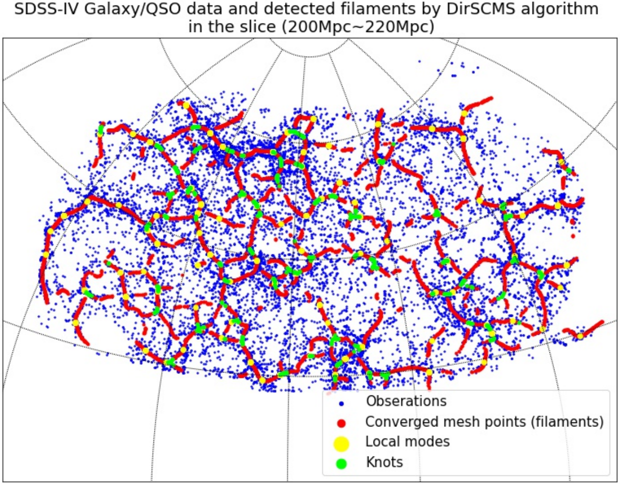


Figure 10: Nodes on the detected filaments.

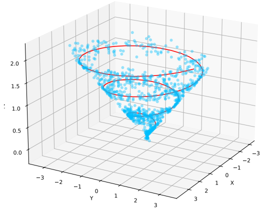
Recall that the survey data $\{(\alpha_i, \delta_i, Z_i)\}_{i=1}^n \in \Omega_2 \times \mathbb{R}^+$ is directional-linear.

- We consider extending our DirSCMS algorithm to estimate the cosmic filaments (*i.e.*, density ridges) in a directional-linear product space (Zhang and Chen, 2021a).
- We adopt the directional-linear KDE (García-Portugués et al., 2015) with $\boldsymbol{X}_i \in \Omega_2$ being the Cartesian coordinate of (ϕ_i, η_i) for $i = 1, ..., n$:

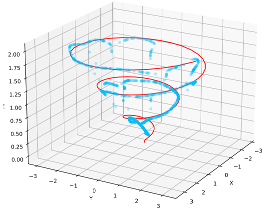
$$\widehat{f}_{\boldsymbol{h}}(\boldsymbol{x}, z) = \frac{C_{L,2}(h_1)}{nh_2} \sum_{i=1}^{n} L\left(\frac{1 - \boldsymbol{x}^T \boldsymbol{X}_i}{h_1^2}\right) K\left(\frac{z - Z_i}{h_2}\right)$$

where $L(r) = e^{-r}$ and $K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ are the kernel functions.

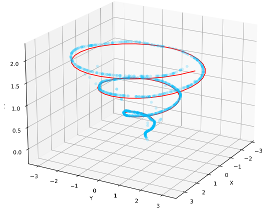
Our directional-linear SCMS algorithm is stabler than its Euclidean prototype.



(a) Simulated data points.



(b) Euclidean SCMS.



(c) Directional-linear SCMS.

Application to SDSS-IV Galaxy Data

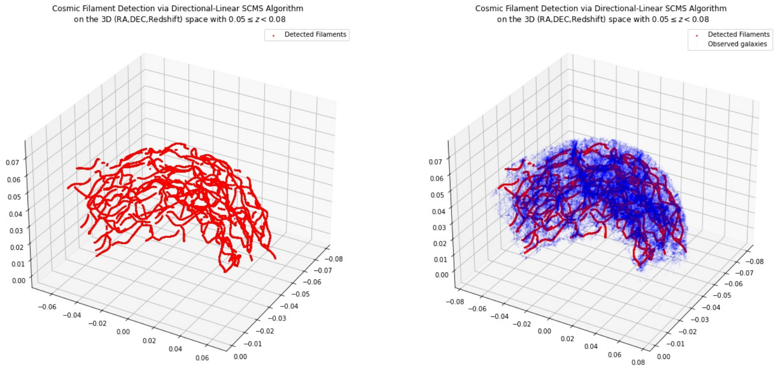


Figure 12: Cosmic filament detection in the 3D (RA,DEC,Redshift) space with our directional-linear SCMS algorithm.

Thank you!

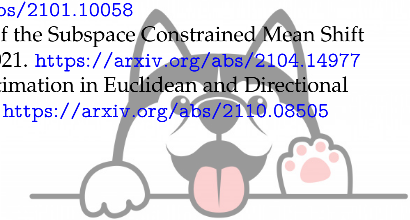
More details can be found in

[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. 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. 2021. 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. 2021. https://arxiv.org/abs/2110.08505



Reference I

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.

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.

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, 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. *arXiv preprint arXiv:2106.05253*, 2021.

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.

Reference II

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. Malavasi, M. Langer, N. Aghanim, D. Galárraga-Espinosa, and C. Gouin. Relative effect of nodes and filaments of the cosmic web on the quenching of galaxies and the orientation of their spin. *Astronomy & Astrophysics*, 658:A113, 2022.

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

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.

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.

Reference III

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

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*, 2021c. URL https://arxiv.org/abs/2104.14977.

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* $\texttt{grad} f(\boldsymbol{x})$ on Ω_q:

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

  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 (Zhang and Chen, 2021b):

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

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 $\mathbb{R}^{q+1}$.

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:

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

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**:

Detailed Procedures of DirSCMS Algorithm II

- **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.

- Recall that the directional-linear KDE at $(\boldsymbol{x}, z) \in \Omega_2 \times \mathbb{R}$ is defined as:

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

- Directional-linear mean shift iteration:

$$\begin{aligned}\left(\boldsymbol{x}^{(t+1)}, z^{(t+1)}\right)^T &\leftarrow \Xi(\boldsymbol{x}^{(t)}, z^{(t)}) + \left(\boldsymbol{x}^{(t)}, z^{(t)}\right)^T \\ &= \begin{pmatrix} \dfrac{\sum_{i=1}^{n} \boldsymbol{X}_i \cdot L'\left(\frac{1-\boldsymbol{X}_i^T \boldsymbol{x}^{(t)}}{h_1}\right) K\left(\frac{z^{(t)} - Z_i}{h_2}\right)}{\sum_{i=1}^{n} L'\left(\frac{1-\boldsymbol{X}_i^T \boldsymbol{x}^{(t)}}{h_1}\right) K\left(\frac{z^{(t)} - Z_i}{h_2}\right)} \\ \dfrac{\sum_{i=1}^{n} Z_i \cdot L\left(\frac{1-\boldsymbol{X}_i^T \boldsymbol{x}^{(t)}}{h_1}\right) K\left(\left|\left|\frac{z^{(t)} - Z_i}{h_2}\right|\right|_2^2\right)}{\sum_{i=1}^{n} L\left(\frac{1-\boldsymbol{X}_i^T \boldsymbol{x}^{(t)}}{h_1}\right) K\left(\left|\left|\frac{z^{(t)} - Z_i}{h_2}\right|\right|_2^2\right)} \end{pmatrix}\end{aligned}$$

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

- Directional-linear SCMS algorithm iteration at $\boldsymbol{y}^{(t)} = \left(\boldsymbol{x}^{(t+1)}, z^{(t+1)}\right)^T$:

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

where $\boldsymbol{H} = \mathtt{Diag}(h_1^2, h_1^2, h_2^2) \in \mathbb{R}^{3\times 3}$ is a diagonal matrix and

$$\Xi(\boldsymbol{y}^{(t)}) = \Xi(\boldsymbol{x}^{(t)}, z^{(t)}) = \begin{pmatrix} \frac{\sum_{i=1}^n \boldsymbol{X}_i \cdot L'\left(\frac{1-\boldsymbol{X}_i^T\boldsymbol{x}^{(t)}}{h_1}\right) K\left(\frac{z^{(t)}-Z_i}{h_2}\right)}{\sum_{i=1}^n L'\left(\frac{1-\boldsymbol{X}_i^T\boldsymbol{x}^{(t)}}{h_1}\right) K\left(\frac{z^{(t)}-Z_i}{h_2}\right)} - \boldsymbol{x}^{(t)} \\ \frac{\sum_{i=1}^n Z_i \cdot L\left(\frac{1-\boldsymbol{X}_i^T\boldsymbol{x}^{(t)}}{h_1}\right) K\left(\left|\left|\frac{z^{(t)}-Z_i}{h_2}\right|\right|_2^2\right)}{\sum_{i=1}^n L\left(\frac{1-\boldsymbol{X}_i^T\boldsymbol{x}^{(t)}}{h_1}\right) K\left(\left|\left|\frac{z^{(t)}-Z_i}{h_2}\right|\right|_2^2\right)} - z^{(t)} \end{pmatrix}.$$

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

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