Geometric Data Analysis Reading Group

Optimization on Manifolds With Applications in Cosmic Web Detection

Yikun Zhang
Joint work with Prof. *Yen-Chi Chen*

Department of Statistics,
University of Washington

October 30, 2023

Outline

Introduction

Motivations: 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 ≈ 3.26 light-years.
- **Cause**: the anisotropic collapse of matter in gravitational instability scenarios at the early stage of the Universe (Zel'Dovich, 1970).

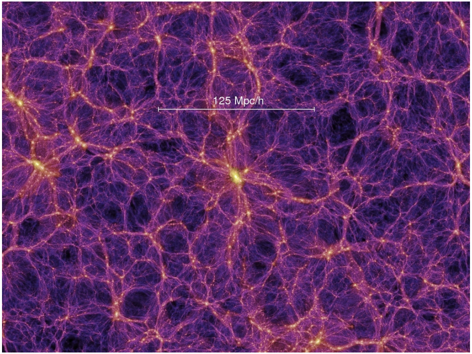


Figure: Visualization of *Cosmic Web* (credited to the Millennium simulation project (Springel et al., 2005)).

Motivations: Characteristics of Cosmic Web

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

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

Motivations: Characteristics of Cosmic Web

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

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

Motivations: 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*.

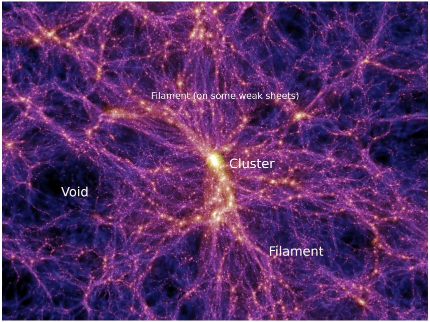


Figure: Characteristics of *Cosmic Web* (credited to the Millennium simulation).

Candidate Data For Cosmic Web Detection

We can reconstruct the underlying cosmic web structures based on

- the observed galaxies in some astronomical surveys (such as the Sloan Digital Sky Survey; SDSS);
- (galactic) halos in some cosmological simulations (such as the Millennium and Illustris simulations).

Candidate Data For Cosmic Web Detection

We can reconstruct the underlying cosmic web structures based on

- the observed galaxies in some astronomical surveys (such as the Sloan Digital Sky Survey; SDSS);
- (galactic) halos in some cosmological simulations (such as the Millennium and Illustris simulations).

Specifically, the positions of observed galaxies (or simulated halos) are recorded as

$$\mathcal{D} = \{(z_1, \alpha_1, \delta_1), ..., (z_n, \alpha_n, \delta_n)\},$$

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

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

Candidate Data For Cosmic Web Detection

The positions of observed galaxies (or simulated halos) are recorded as $\mathcal{D} = \{(z_1, \alpha_1, \delta_1), ..., (z_n, \alpha_n, \delta_n)\}$, where, for $i = 1, ..., n$,

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

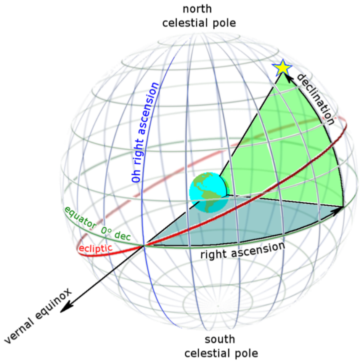


Figure: Illustration of RA and DEC (Image Courtesy of Wikipedia).

General Cosmic Web Detection

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

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

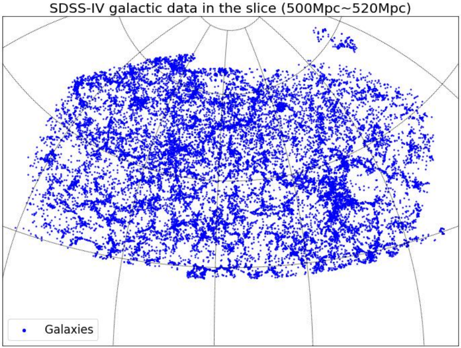


Figure: Distribution of galaxies within a thin redshift slice.

General Cosmic Web Detection

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

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

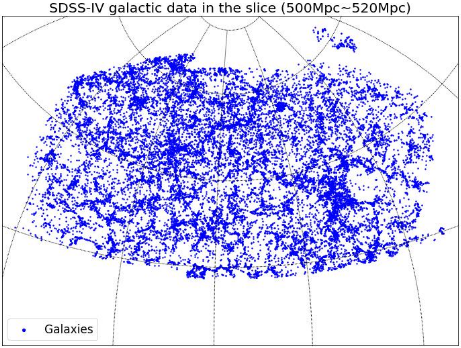


Figure: Distribution of galaxies within a thin redshift slice.

- 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.

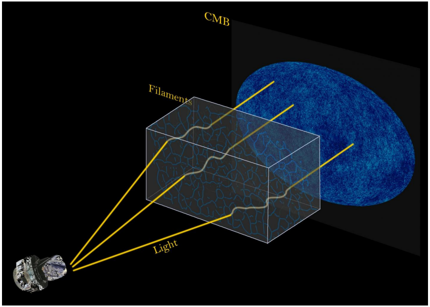


Figure: Illustration of the bending trajectory of CMB lights (credit to Siyu He, Shadab Alam, Wei Chen, and Planck/ESA; see He et al. (2018) for details).

Highlights of the Today's Talk

1. Brief introduction to our cosmic filament model and its statistical estimation theory.
    - Introduce the *directional density ridge* theory.
    - Prove the statistical consistency in estimating the true density ridges with directional kernel density estimator (KDE).

Highlights of the Today's Talk

1. Brief introduction to our cosmic filament model and its statistical estimation theory.
   - Introduce the *directional density ridge* theory.
   - Prove the statistical consistency in estimating the true density ridges with directional kernel density estimator (KDE).
2. Presentation on computational perspectives of identifying cosmic filaments via *Directional (Subspace Constrained) Mean Shift* algorithm.
   - Formulate our DirSCMS algorithm as an example of the general subspace constrained gradient ascent (SCGA) methods on (nonlinear) manifolds.
   - Establish the linear convergence properties of our SCGA algorithm.

Highlights of the Today's Talk

1. Brief introduction to our cosmic filament model and its statistical estimation theory.
   - Introduce the *directional density ridge* theory.
   - Prove the statistical consistency in estimating the true density ridges with directional kernel density estimator (KDE).
2. Presentation on computational perspectives of identifying cosmic filaments via *Directional (Subspace Constrained) Mean Shift* algorithm.
   - Formulate our DirSCMS algorithm as an example of the general subspace constrained gradient ascent (SCGA) methods on (nonlinear) manifolds.
   - Establish the linear convergence properties of our SCGA algorithm.
3. Discussion on some future directions for our SCGA methods.

Cosmic Filament Model: Directional Density Ridges

An Existing Statistical Model for Cosmic Filaments

Given a (smooth) density function p on the Euclidean space $\mathbb{R}^D$ (where $D = 2$ for the 2D cosmic web detection problem), we consider its gradient $\nabla p(\boldsymbol{x}) \in \mathbb{R}^D$ and Hessian matrix $\nabla\nabla p(\boldsymbol{x}) \in \mathbb{R}^{D\times D}$ with spectral decomposition

$$\nabla\nabla p(\boldsymbol{x}) = V(\boldsymbol{x})\Lambda(\boldsymbol{x})V(\boldsymbol{x})^T,$$

where

- $V(\boldsymbol{x}) = [\boldsymbol{v}_1(\boldsymbol{x}), ..., \boldsymbol{v}_D(\boldsymbol{x})] \in \mathbb{R}^{D\times D}$ is a real orthogonal matrix.
- $\Lambda(\boldsymbol{x}) = \mathtt{Diag}\left[\lambda_1(\boldsymbol{x}), ..., \lambda_D(\boldsymbol{x})\right] \in \mathbb{R}^{D\times D}$ is a diagonal matrix with $\lambda_1(\boldsymbol{x}) \geq \cdots \geq \lambda_D(\boldsymbol{x})$.

An Existing Statistical Model for Cosmic Filaments

Given a (smooth) density function p on the Euclidean space $\mathbb{R}^D$ (where $D = 2$ for the 2D cosmic web detection problem), we consider its gradient $\nabla p(\boldsymbol{x}) \in \mathbb{R}^D$ and Hessian matrix $\nabla\nabla p(\boldsymbol{x}) \in \mathbb{R}^{D\times D}$ with spectral decomposition

$$\nabla\nabla p(\boldsymbol{x}) = V(\boldsymbol{x})\Lambda(\boldsymbol{x})V(\boldsymbol{x})^T,$$

where

- $V(\boldsymbol{x}) = [\boldsymbol{v}_1(\boldsymbol{x}), ..., \boldsymbol{v}_D(\boldsymbol{x})] \in \mathbb{R}^{D\times D}$ is a real orthogonal matrix.
- $\Lambda(\boldsymbol{x}) = \mathtt{Diag}\left[\lambda_1(\boldsymbol{x}), ..., \lambda_D(\boldsymbol{x})\right] \in \mathbb{R}^{D\times D}$ is a diagonal matrix with $\lambda_1(\boldsymbol{x}) \geq \cdots \geq \lambda_D(\boldsymbol{x})$.

The cosmic web can be characterized as the order-d density ridge (Genovese et al., 2014; Chen et al., 2015):

$$R_d(p) = \left\{\boldsymbol{x} \in \mathbb{R}^D : V_d(\boldsymbol{x})^T \nabla p(\boldsymbol{x}) = \mathbf{0}, \lambda_{d+1}(\boldsymbol{x}) < 0\right\},$$

where $V_d(\boldsymbol{x}) = [\boldsymbol{v}_{d+1}(\boldsymbol{x}), ..., \boldsymbol{v}_D(\boldsymbol{x})] \in \mathbb{R}^{D\times(D-d)}$ has its columns as the unit eigenvectors of $\nabla\nabla p(\boldsymbol{x})$ associated with the last $D-d$ eigenvalues.

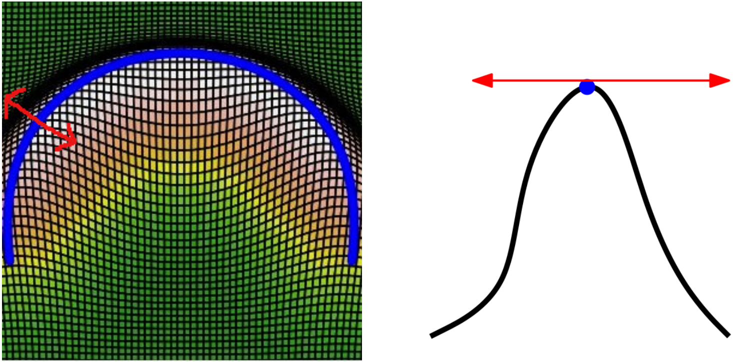


Figure: Density ridge (lifted onto the density function p; Chen et al. 2015)

The order-d density ridge of p in $\mathbb{R}^D$:

$$R_d(p) = \left\{ \boldsymbol{x} \in \mathbb{R}^D : V_d(\boldsymbol{x})^T \nabla p(\boldsymbol{x}) = \mathbf{0}, \lambda_{d+1}(\boldsymbol{x}) < 0 \right\}.$$

An Existing Statistical Model for Cosmic Filaments

- When $d = 0$, $R_d(p)$ reduces to the set of local modes of p as:

$$R_0(p) = \left\{\boldsymbol{x} \in \mathbb{R}^D : \nabla p(\boldsymbol{x}) = \mathbf{0}, \lambda_1(\boldsymbol{x}) < 0\right\},$$

which marks the set of candidate *galaxy clusters*.

- When $d = 1$, $R_d(p)$ characterizes the collection of one-dimensional *cosmic filaments*.

- When $d = 2$, $R_d(p)$ specifies the two-dimensional *cosmic sheets/walls*.

Drawback of the Standard Density Ridge Model

By controlling the redshift value, the locations of galaxies are recorded as $\{(\alpha_i, \delta_i)\}_{i=1}^n$ on a (unit) sphere Ω_2, where $\Omega_q = \left\{\boldsymbol{x} \in \mathbb{R}^{q+1} : ||x||_2 = 1\right\}$.

- The observed galaxies do not lie on a flat 2D plane but some **spherical shell**, which have a *nonlinear* curvature!

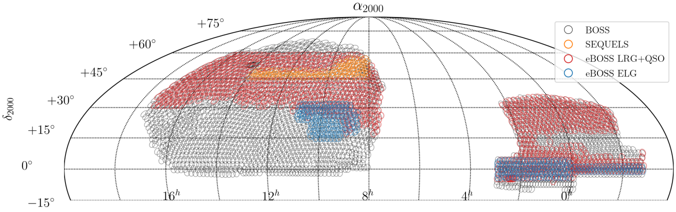


Figure: BOSS/eBOSS Spectroscopic Footprint as of DR16 (credit to SDSS)

► **Question**: How do we generalize the density ridge model in order to handle the nonlinear geometry?

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.

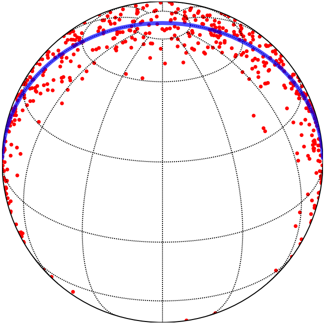


Figure: Noisy observations (red points) and the underlying true ring/filament structure (blue line)

Methods:

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

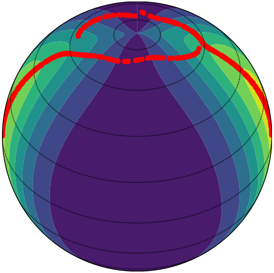
Or

- We consider those data points on the unit sphere $\Omega_2 = \left\{\boldsymbol{x} \in \mathbb{R}^3 : ||\boldsymbol{x}||_2 = 1\right\}$, and recover the ring/filament structure using DirSCMS algorithm (Zhang and Chen, 2022).

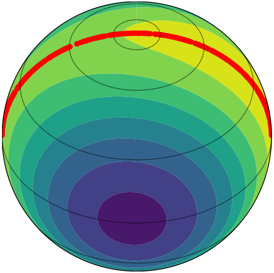
We will discuss more on the directional subspace constrained mean shift (DirSCMS) algorithm soon.

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

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



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



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

To define our density ridge model on Ω_q or general manifolds, we need some technical concepts from differential geometry.

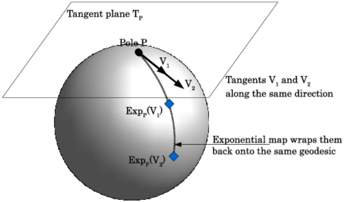
To define our density ridge model on Ω_q or general manifolds, we need some technical concepts from differential geometry.

Definition (Tangent space of Ω_q)

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

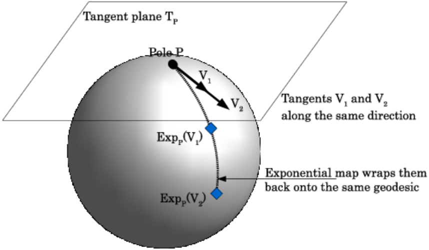
$$T_{\boldsymbol{x}} \equiv T_{\boldsymbol{x}}(\Omega_q) \simeq \left\{ \boldsymbol{v} \in \mathbb{R}^{q+1} : \boldsymbol{x}^T \boldsymbol{v} = 0 \right\},$$

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 $\boldsymbol{x}$.



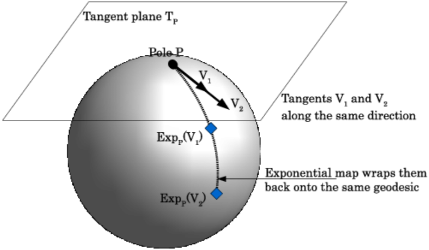
Definition (Exponential Map)

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



Definition (Exponential Map)

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



The inverse of an exponential map is known as the *logarithmic map* $\texttt{Exp}_{\boldsymbol{x}}^{-1} : U \to T_{\boldsymbol{x}}$ for some neighborhood $U \subset \Omega_q$ so that $\texttt{Exp}_{\boldsymbol{x}}^{-1}(\boldsymbol{y})$ is a tangent vector in $T_{\boldsymbol{x}}$ pointing to $\boldsymbol{y} \in \Omega_q$ with $\left|\left|Exp_{\boldsymbol{x}}^{-1}(\boldsymbol{y})\right|\right|_2$ being the geodesic distance between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Riemannian Gradient on Ω_q

Given a smooth function $f : \Omega_q \to \mathbb{R}$ and 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 $f : \Omega_q \to \mathbb{R}$ and 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

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

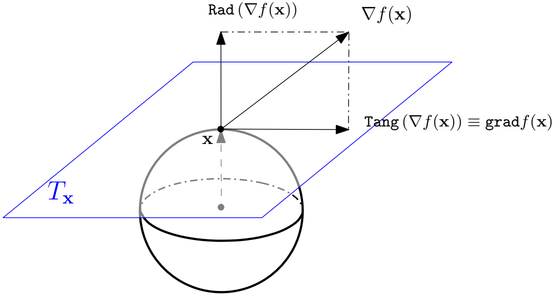
for any $\boldsymbol{v} \in T_{\boldsymbol{x}}$ and the predefined Riemannian metric $\langle \cdot, \cdot, \rangle_{\boldsymbol{x}}$.

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:

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

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, $\boldsymbol{I}_{q+1}$ 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

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

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, 2021b):

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

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

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

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, 2021b):
$$\mathcal{H}f(\boldsymbol{x}) = \left(\boldsymbol{I}_{q+1} - \boldsymbol{x}\boldsymbol{x}^T\right)\left[\nabla\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{x})^T \boldsymbol{x} \cdot \boldsymbol{I}_{q+1}\right]\left(\boldsymbol{I}_{q+1} - \boldsymbol{x}\boldsymbol{x}^T\right).$$
3. Taylor's expansion (Pennec, 2006):
$(f \circ \mathtt{Exp}_{\boldsymbol{x}})(\boldsymbol{v}) = f(\boldsymbol{x}) + \langle \mathtt{grad} f(\boldsymbol{x}), \boldsymbol{v} \rangle_{\boldsymbol{x}} + \frac{1}{2}\langle \mathcal{H}f(\boldsymbol{x})[\boldsymbol{v}], \boldsymbol{v} \rangle_{\boldsymbol{x}} + O\left(||\boldsymbol{v}||^3\right).$

Directional Density Ridges on Ω_q

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 $\mathcal{H}f(\boldsymbol{x})$.

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

Directional Density Ridges on Ω_q

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 $\mathcal{H}f(\boldsymbol{x})$.

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

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

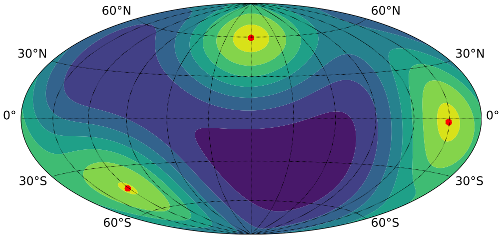
$$\mathcal{R}_d \equiv \texttt{Ridge}(f) = \left\{\boldsymbol{x} \in \Omega_q : V_d(\boldsymbol{x})V_d(\boldsymbol{x})^T \texttt{grad} f(\boldsymbol{x}) = \mathbf{0}, \lambda_{d+1}(\boldsymbol{x}) < 0\right\},$$

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

Local Modes of f on Ω_q

When $d = 1$, the directional density ridge $\mathcal{R}_1$ becomes the set of local modes $\mathcal{M}$ of f on Ω_q as:

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



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

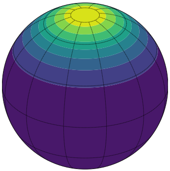
Directional Density Estimation

How do we estimate the directional density ridge $\mathcal{R}_d$ on Ω_q from directional data $\{(\alpha_1, \delta_1), ..., (\alpha_n, \delta_n)\} \to \{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\} \subset \Omega_q$?

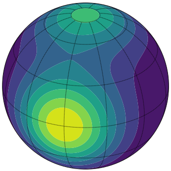
We first estimate the density function f on Ω_q via the directional KDE (Hall et al., 1987; Bai et al., 1988; García-Portugués, 2013) as:

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

- $L : [0, \infty) \to [0, \infty)$ is a directional kernel, *i.e.*, a rapidly decaying nonnegative function (e.g., 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

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

and

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

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

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

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

- **Question:** How do we identify the sets of directional local modes $\widehat{\mathcal{M}}$ and ridge $\widehat{\mathcal{R}}_d$ in practice?

▶ **Question:** How do we identify the sets of directional local modes $\widehat{\mathcal{M}}$ and ridge $\widehat{\mathcal{R}}_d$ in practice?

1. (**Mode-Seeking**) We develop the directional mean shift procedure to estimate $\widehat{\mathcal{M}}$ as (Section 3 in Zhang and Chen 2021b):

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

2. (**Ridge-Finding**) We also generalize it to the directional subspace constrained mean shift (SCMS) algorithm as (Section 4.2 in Zhang and Chen 2022):

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

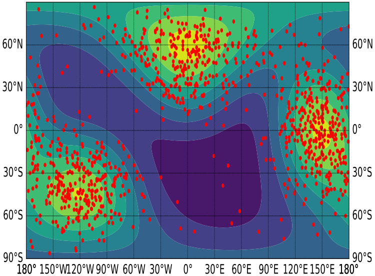
for $t = 0, 1, ...$

Directional Mean Shift: Simulation Study

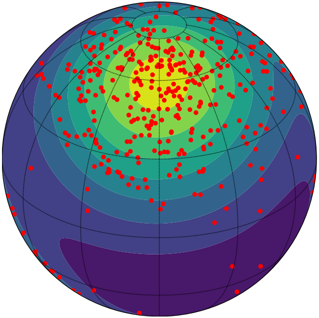
We simulate 1000 data points from the following density

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

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

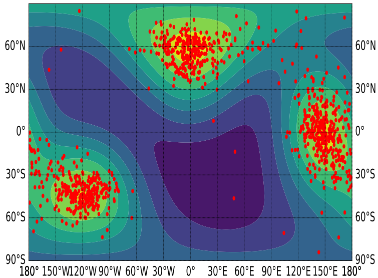


(b) Step 0

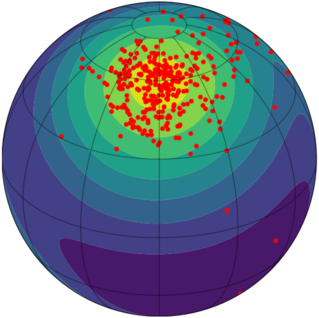
We simulate 1000 data points from the following density

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

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



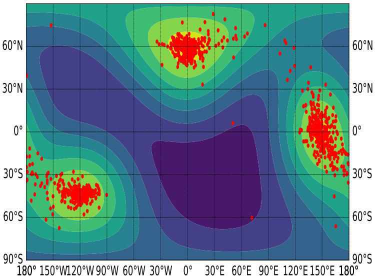
(b) Step 1

Directional Mean Shift: Simulation Study

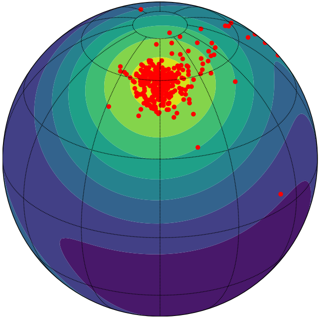
We simulate 1000 data points from the following density

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

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



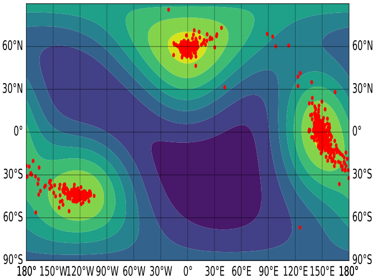
(b) Step 2

Directional Mean Shift: Simulation Study

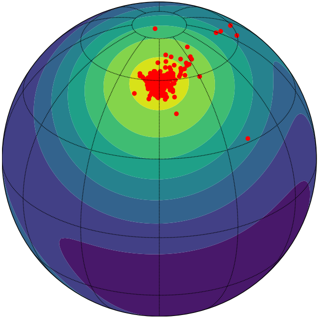
We simulate 1000 data points from the following density

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

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 3

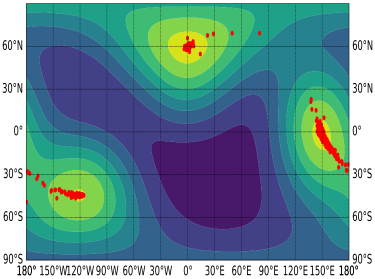


(b) Step 3

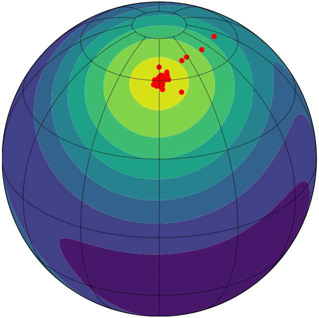
We simulate 1000 data points from the following density

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

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 5



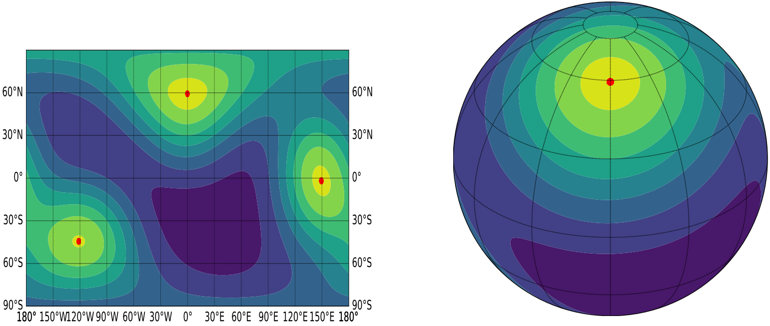
(b) Step 5

Directional Mean Shift: Simulation Study

We simulate 1000 data points from the following density

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

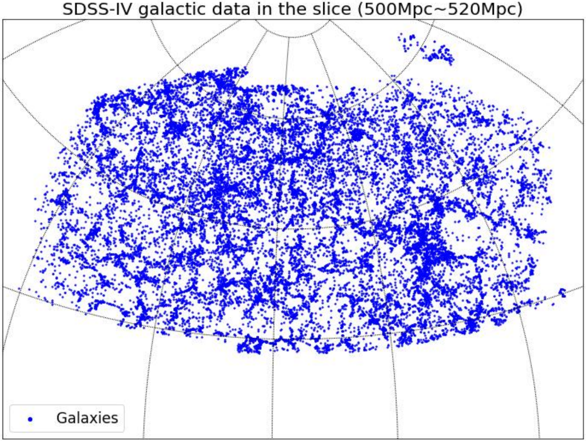
with $\boldsymbol{\mu}_1 = [-120°, -45°]$, $\boldsymbol{\mu}_2 = [0°, 60°]$, $\boldsymbol{\mu}_3 = [150°, 0°]$, and $\nu_1 = \nu_2 = 8$, $\nu_3 = 5$.



(a) Step 22 (converged) (b) Step 22 (converged)

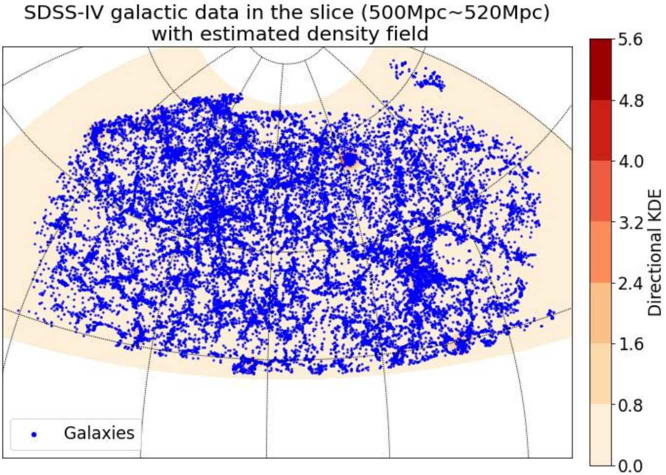
Step 1 (Slicing the Universe): Partition the redshift range into 325 spherical slices based on the comoving distance $\Delta L = 20\,\text{Mpc}$.

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



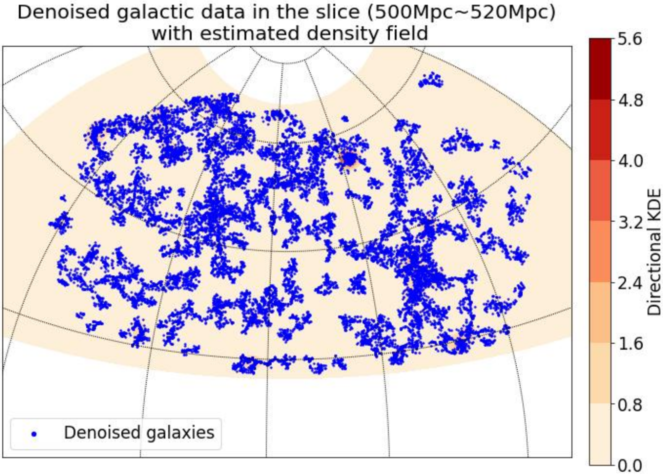
Step 2 (Density Estimation): Estimate the galaxy density field within each spherical slice by directional KDE.

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

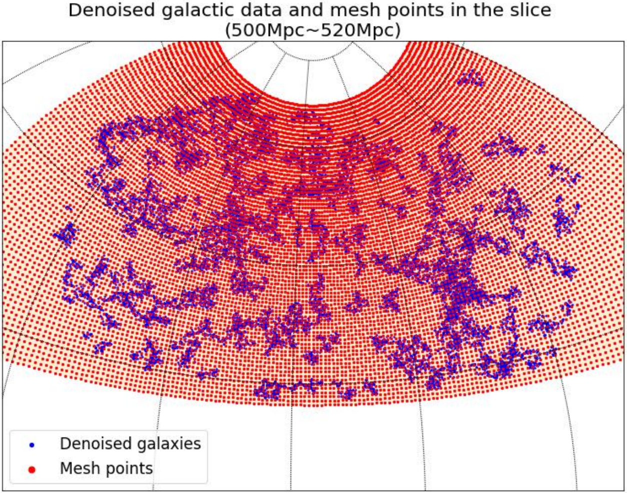


Step 3 (Denoising): Remove the observations with low-density values.

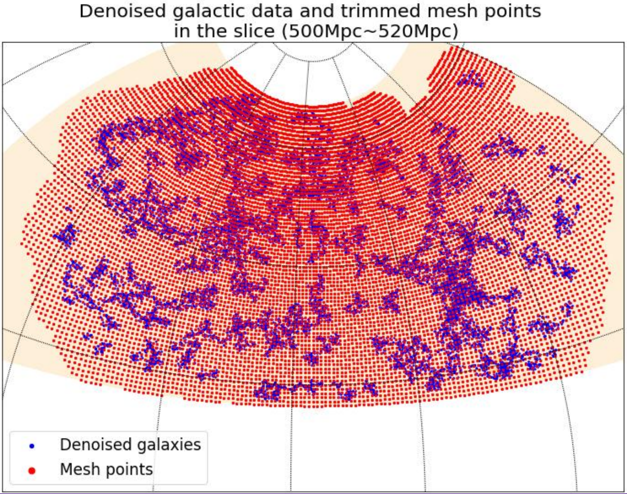
- We keep at least 80% of the original galaxy data in the slice.



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.

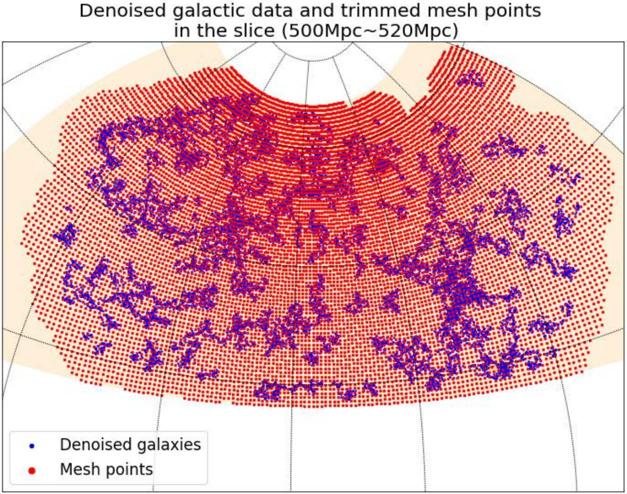


Figure: DirSCMS Iterations (Step 0).

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

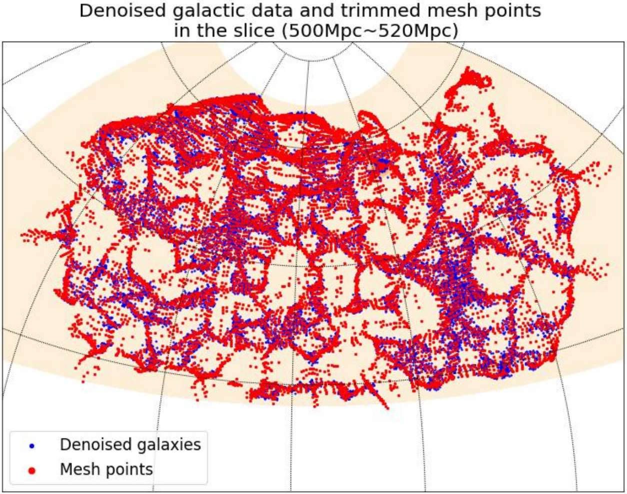


Figure: DirSCMS Iterations (Step 1).

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

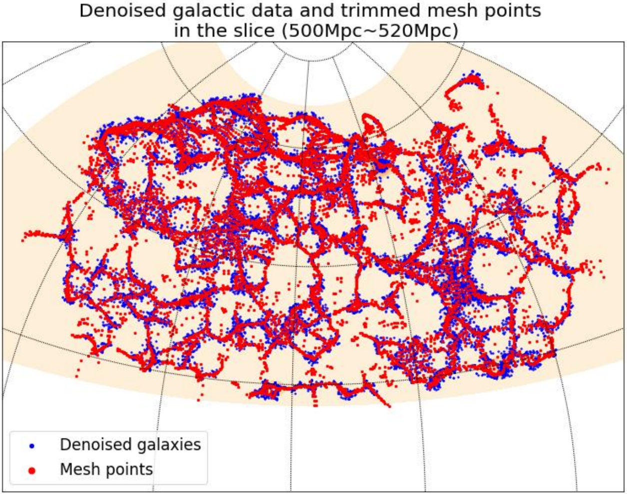


Figure: DirSCMS Iterations (Step 2).

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

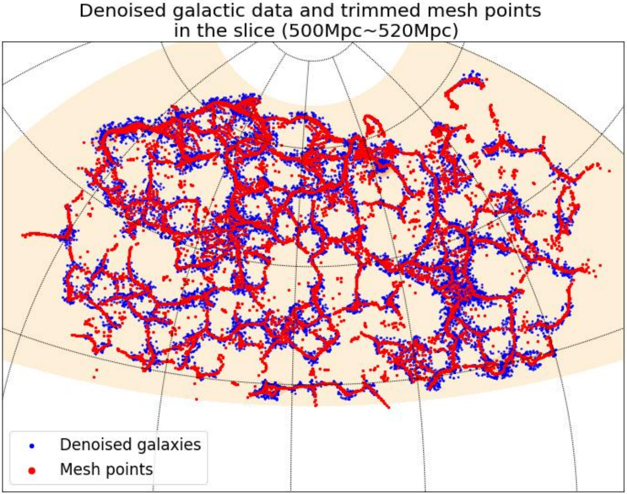


Figure: DirSCMS Iterations (Step 3).

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

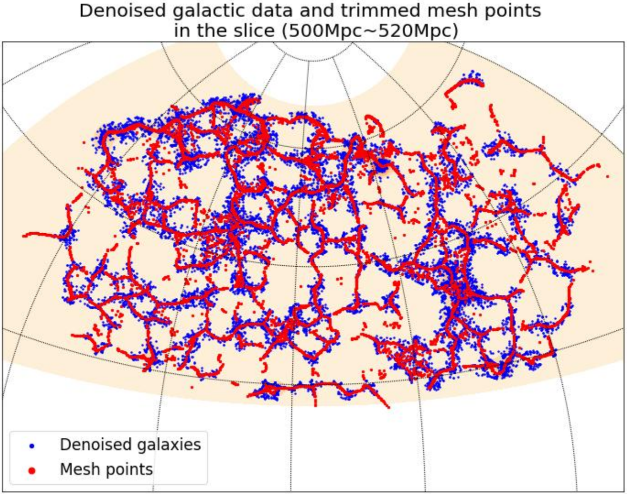


Figure: DirSCMS Iterations (Step 5).

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

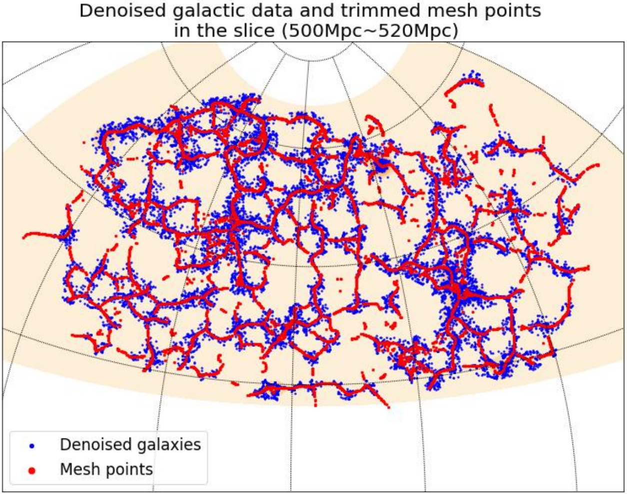


Figure: DirSCMS Iterations (Step 8).

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

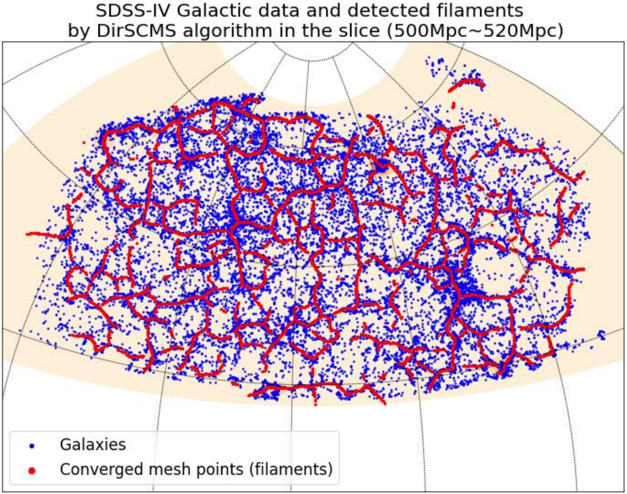


Figure: DirSCMS Iterations (Final).

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

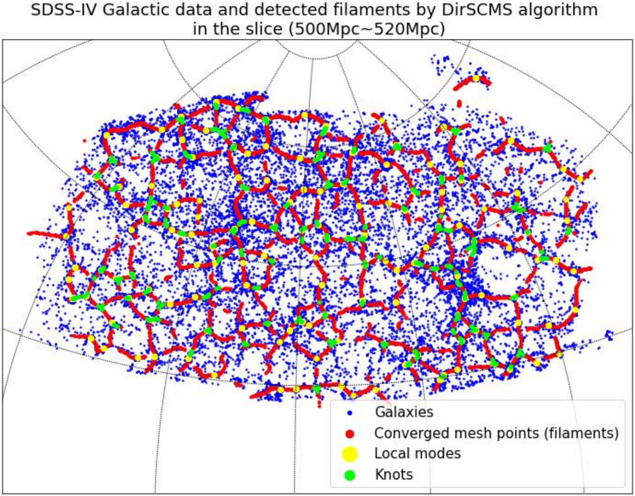


Figure: Nodes on the detected filaments.

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).

SCONCE
SCMS

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

Final Cosmic Web Catalog on SDSS-IV Data

- The final catalog is available at https://doi.org/10.5281/zenodo.6244866.

Published June 10, 2022 | Version 0.0.1

Dataset | Open

SDSS-IV Cosmic Web Catalog

Zhang, Yikun[1]

Show affiliations

Supervisors: Chen, Yen-Chi[1]; de Souza, Rafael S.[2]

Show affiliations

This repository contains the cosmic web catalog data released in the paper "Cosmic Web Catalog on SDSS-IV Data with SCONCE" (preparing).

The catalog is constructed on the SDSS-IV galaxies and quasars (QSO) using our proposed Directional Subspace Constrained Mean Shift (DirSCMS) algorithm. We release both the cosmic filaments and local modes (i.e., local maxima of the estimated galaxy/QSO density field, which serves as candidates of galaxy clusters) within 325 thin redshift slices, each of which spans 20Mpc under the Planck15 cosmology. The entire catalog covers the redshift range from \(z=0\) to \(z=3\).

1. "**Cosmic_filaments_2D_DirSCMS_new1**": The file contains some discrete realizations of the estimated cosmic filaments in some particular redshift slices. The meaning of each column in the file is described as follows:

- **RA** -- right ascension.
- **DEC** -- declination.
- **z_low** -- lower limit of the redshift slice.
- **z_high** -- upper limit of the redshift slice.
- **comov_dist_low** -- lower limit of the comoving distance in the redshift slice under the Planck15 cosmology.
- **comov_dist_high** -- upper limit of the comoving distance in the redshift slice under the Planck15 cosmology.
- **bw** -- smoothing bandwidth parameter for the DirSCMS algorithm in the redshift slice.
- **unc_meas** -- uncertainty measure of the filamentary point by the nonparametric bootstrap techinque.
- **density** -- (proportional) estimated galaxy/QSO density value at the filamentary point.
- **grad_Dir1** -- (Riemannian) gradient of the estimated density field (first direction).
- **grad_Dir2** -- (Riemannian) gradient of the estimated density field (second direction).
- **grad_Dir3** -- (Riemannian) gradient of the estimated density field (third direction).
- **knot_label** -- indicator of whether the filamentary point is a knot (i.e., the intersection of several filaments) or not.

Versions

Version 0.0.1 — Jun 10, 2022

10.5281/zenodo.6244866

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.6244865. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

Indexed in

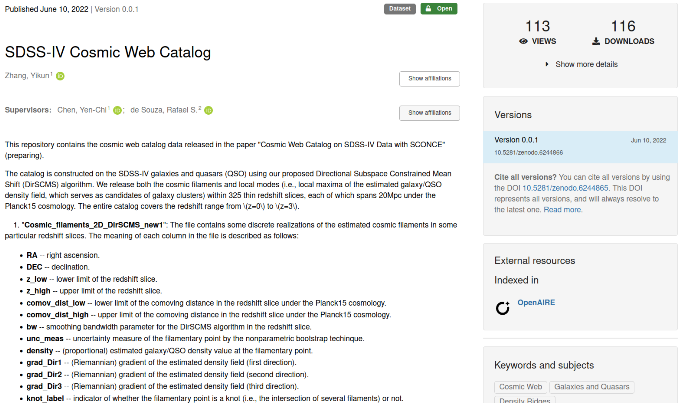
OpenAIRE

Keywords and subjects

Cosmic Web | Galaxies and Quasars | Density Ridges

Final Cosmic Web Catalog on SDSS-IV Data

- The final catalog is available at https://doi.org/10.5281/zenodo.6244866.



Published June 10, 2022 | Version 0.0.1

Dataset Open

SDSS-IV Cosmic Web Catalog

Zhang, Yikun[1]

Show affiliations

Supervisors: Chen, Yen-Chi[1]; de Souza, Rafael S.[2]

Show affiliations

This repository contains the cosmic web catalog data released in the paper "Cosmic Web Catalog on SDSS-IV Data with SCONCE" (preparing).

The catalog is constructed on the SDSS-IV galaxies and quasars (QSO) using our proposed Directional Subspace Constrained Mean Shift (DirSCMS) algorithm. We release both the cosmic filaments and local modes (i.e., local maxima of the estimated galaxy/QSO density field, which serves as candidates of galaxy clusters) within 325 thin redshift slices, each of which spans 20Mpc under the Planck15 cosmology. The entire catalog covers the redshift range from \(z=0\) to \(z=3\).

1. "**Cosmic_filaments_2D_DirSCMS_new1**": The file contains some discrete realizations of the estimated cosmic filaments in some particular redshift slices. The meaning of each column in the file is described as follows:

- **RA** -- right ascension.
- **DEC** -- declination.
- **z_low** -- lower limit of the redshift slice.
- **z_high** -- upper limit of the redshift slice.
- **comov_dist_low** -- lower limit of the comoving distance in the redshift slice under the Planck15 cosmology.
- **comov_dist_high** -- upper limit of the comoving distance in the redshift slice under the Planck15 cosmology.
- **bw** -- smoothing bandwidth parameter for the DirSCMS algorithm in the redshift slice.
- **unc_meas** -- uncertainty measure of the filamentary point by the nonparametric bootstrap techinque.
- **density** -- (proportional) estimated galaxy/QSO density value at the filamentary point.
- **grad_Dir1** -- (Riemannian) gradient of the estimated density field (first direction).
- **grad_Dir2** -- (Riemannian) gradient of the estimated density field (second direction).
- **grad_Dir3** -- (Riemannian) gradient of the estimated density field (third direction).
- **knot_label** -- indicator of whether the filamentary point is a knot (i.e., the intersection of several filaments) or not.

113 VIEWS | 116 DOWNLOADS

Show more details

Versions

Version 0.0.1 Jun 10, 2022
10.5281/zenodo.6244866

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.6244865. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

Indexed in

OpenAIRE

Keywords and subjects

Cosmic Web | Galaxies and Quasars | Density Ridges

Advertisement: We use this cosmic web catalog to study the relation between stellar masses of galaxies and their distances to filaments despite some missingness in stellar masses (next Wednesday 5-6pm).

Optimization Theory on Manifolds

Convergence Results of the Proposed Methods

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

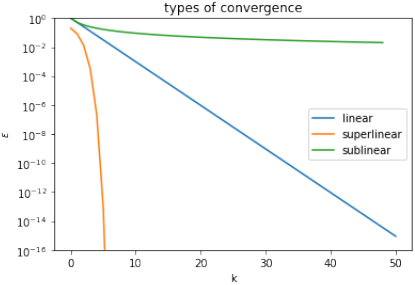
- **Question**: how fast will our proposed algorithms converge?

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

- **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).



Optimization Problems on Manifolds

Given a function $f : \mathbb{M} \to \mathbb{R}$ defined on a (smooth) manifold $\mathbb{M}$, we consider an unconstrained optimization problem on $\mathbb{M}$ as:

$$\max_{x \in \mathbb{M}} f(x).$$

Optimization Problems on Manifolds

Given a function $f : \mathbb{M} \to \mathbb{R}$ defined on a (smooth) manifold $\mathbb{M}$, we consider an unconstrained optimization problem on $\mathbb{M}$ as:

$$\max_{x \in \mathbb{M}} f(x).$$

1. Stiefel manifold: $St(D_1, D_2) = \left\{ \boldsymbol{X} \in \mathbb{R}^{D_1 \times D_2} : \boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{I}_{D_2} \right\}$ (e.g., Ω_q).

Optimization Problems on Manifolds

Given a function $f : \mathbb{M} \to \mathbb{R}$ defined on a (smooth) manifold $\mathbb{M}$, we consider an unconstrained optimization problem on $\mathbb{M}$ as:

$$\max_{x \in \mathbb{M}} f(x).$$

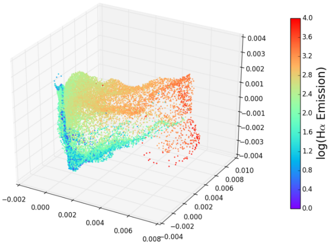
1. Stiefel manifold: $St(D_1, D_2) = \left\{ \boldsymbol{X} \in \mathbb{R}^{D_1 \times D_2} : \boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{I}_{D_2} \right\}$ (e.g., Ω_q).
2. Gasserman manifold: $\{$linear subspaces of dimension d in $\mathbb{R}^D\}$.

Optimization Problems on Manifolds

Given a function $f : \mathbb{M} \to \mathbb{R}$ defined on a (smooth) manifold $\mathbb{M}$, we consider an unconstrained optimization problem on $\mathbb{M}$ as:

$$\max_{x \in \mathbb{M}} f(x).$$

1. Stiefel manifold: $St(D_1, D_2) = \left\{ \boldsymbol{X} \in \mathbb{R}^{D_1 \times D_2} : \boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{I}_{D_2} \right\}$ (e.g., Ω_q).
2. Gasserman manifold: $\{$linear subspaces of dimension d in $\mathbb{R}^D\}$.
3. The fitted manifold from the data (Izenman, 2012).



A three-dimensional embedding of the main sample of galaxy spectra from SDSS (approximately 675,000 spectra observed in 3750 dimensions; McQueen et al. 2016).

Optimization Problems on Manifolds

Given a function $f : \mathbb{M} \to \mathbb{R}$ defined on a (smooth) manifold $\mathbb{M}$, we consider an unconstrained optimization problem on $\mathbb{M}$ as:

$$\max_{x \in \mathbb{M}} f(x).$$

Applications of optimization on manifolds include

- low-rank matrix completion and tensor factorization (Vandereycken, 2013; Mishra et al., 2013).

Optimization Problems on Manifolds

Given a function $f : \mathbb{M} \to \mathbb{R}$ defined on a (smooth) manifold $\mathbb{M}$, we consider an unconstrained optimization problem on $\mathbb{M}$ as:

$$\max_{x \in \mathbb{M}} f(x).$$

Applications of optimization on manifolds include

- low-rank matrix completion and tensor factorization (Vandereycken, 2013; Mishra et al., 2013).
- dictionary learning (Harandi et al., 2012).

Optimization Problems on Manifolds

Given a function $f : \mathbb{M} \to \mathbb{R}$ defined on a (smooth) manifold $\mathbb{M}$, we consider an unconstrained optimization problem on $\mathbb{M}$ as:

$$\max_{x \in \mathbb{M}} f(x).$$

Applications of optimization on manifolds include

- low-rank matrix completion and tensor factorization (Vandereycken, 2013; Mishra et al., 2013).
- dictionary learning (Harandi et al., 2012).
- Fréchet mean or geometric median on manifolds (Lin et al., 2020).

Optimization Problems on Manifolds

Given a function $f : \mathbb{M} \to \mathbb{R}$ defined on a (smooth) manifold $\mathbb{M}$, we consider an unconstrained optimization problem on $\mathbb{M}$ as:

$$\max_{x \in \mathbb{M}} f(x).$$

Applications of optimization on manifolds include

- low-rank matrix completion and tensor factorization (Vandereycken, 2013; Mishra et al., 2013).
- dictionary learning (Harandi et al., 2012).
- Fréchet mean or geometric median on manifolds (Lin et al., 2020).
- accelerating the parameter estimation for Gaussian mixture models (Hosseini and Sra, 2015).

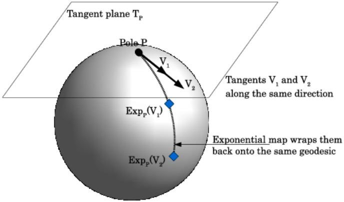
First-Order Methods on Manifolds

To solve the optimization problem on a manifold, $\max_{x \in \mathbb{M}} f(x)$, under a smooth function $f : \mathbb{M} \to \mathbb{R}$, we consider generalizing the gradient ascent method from $\mathbb{R}^D$ to $\mathbb{M}$.

- Gradient Ascent Algorithm on $\mathbb{M}$:

$$\boldsymbol{x}^{(t+1)} = \mathtt{Exp}_{\boldsymbol{x}^{(t)}} \left(\eta \cdot \mathtt{grad} f(\boldsymbol{x}^{(t)}) \right),$$

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



Establishing the linear convergence of gradient ascent methods on manifolds is challenging!

- The law of cosines in Euclidean space

$$a^2 = b^2 + c^2 - 2bc\cos(A)$$

  is no longer valid on nonlinear manifolds.

Establishing the linear convergence of gradient ascent methods on manifolds is challenging!

- The law of cosines in Euclidean space

$$a^2 = b^2 + c^2 - 2bc\cos(A)$$

is no longer valid on nonlinear manifolds.

Lemma (Lemma 5 in Zhang and Sra 2016)

If a, b, c are the sides (i.e., side lengths) of a geodesic triangle in an Alexandrov space with sectional curvature lower bounded by κ, and A is the angle between sides b and c, then

$$a^2 \leq \frac{\sqrt{|\kappa|}c}{\tanh(\sqrt{|\kappa|}c)} b^2 + c^2 - 2bc\cos(A).$$

Linear Convergence of Gradient Ascent Method on $\mathbb{M}$

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

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 $\boldsymbol{x}^{(0)} \in \left\{\boldsymbol{z} \in \mathbb{M} : ||\boldsymbol{z} - \boldsymbol{m}_k||_2 < r_0\right\}$ for some $\boldsymbol{m}_k \in \mathcal{M}$,

$$d(\boldsymbol{x}^{(t)}, \boldsymbol{m}_k) \leq \Upsilon^t \cdot d(\boldsymbol{x}^{(0)}, \boldsymbol{m}_k) \quad \text{with} \quad \Upsilon = \sqrt{1 - \frac{\eta\lambda_*}{2}},$$

where $d(\boldsymbol{p}, \boldsymbol{q}) = \left|\left|\mathtt{Exp}_{\boldsymbol{p}}^{-1}(\boldsymbol{q})\right|\right|_2$.

Linear Convergence of Gradient Ascent Method on $\mathbb{M}$

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

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 $\boldsymbol{x}^{(0)} \in \left\{\boldsymbol{z} \in \mathbb{M} : ||\boldsymbol{z} - \boldsymbol{m}_k||_2 < r_0\right\}$ for some $\boldsymbol{m}_k \in \mathcal{M}$,
$$d(\boldsymbol{x}^{(t)}, \boldsymbol{m}_k) \leq \Upsilon^t \cdot d(\boldsymbol{x}^{(0)}, \boldsymbol{m}_k) \quad \text{with} \quad \Upsilon = \sqrt{1 - \frac{\eta\lambda_*}{2}},$$
where $d(\boldsymbol{p}, \boldsymbol{q}) = \left|\left|\mathtt{Exp}_{\boldsymbol{p}}^{-1}(\boldsymbol{q})\right|\right|_2$.
2. **Linear convergence of gradient ascent with $\widehat{f}_h$:** let the gradient ascent update on $\mathbb{M}$ be
$$\widehat{\boldsymbol{x}}^{(t+1)} = \mathtt{Exp}_{\widehat{\boldsymbol{x}}^{(t)}}\left(\eta \cdot \mathtt{grad}\,\widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})\right).$$
When the step size $\eta > 0$ is sufficiently small and the initial point $\widehat{\boldsymbol{x}}^{(0)} \in \left\{\boldsymbol{z} \in \mathbb{M} : ||\boldsymbol{z} - \boldsymbol{m}_k||_2 < r_0\right\}$ for some $\boldsymbol{m}_k \in \mathcal{M}$,
$$d\left(\widehat{\boldsymbol{x}}^{(t)}, \boldsymbol{m}_k\right) \leq \Upsilon^t \cdot d\left(\widehat{\boldsymbol{x}}^{(0)}, \boldsymbol{m}_k\right) + O(h^2) + O_P\left(\sqrt{\frac{|\log h|}{nh^{q+2}}}\right)$$
with probability tending to 1, as $h \to 0$ and $\frac{nh^{q+2}}{|\log h|} \to \infty$.

- **Question:** Why is the linear convergence of gradient ascent methods on $\mathbb{M}$ only valid around a small neighborhood of $\boldsymbol{m}_k \in \mathcal{M}$?

▶ **Question:** Why is the linear convergence of gradient ascent methods on $\mathbb{M}$ only valid around a small neighborhood of $\boldsymbol{m}_k \in \mathcal{M}$?

• **Geodesically Strong Concavity**. A differentiable function $f : \mathbb{M} \to \mathbb{R}$ is said to be *geodesically concave* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{M}$, it holds that

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \leq \left\langle \mathtt{grad} f(\boldsymbol{x}), \mathtt{Exp}_{\boldsymbol{x}}^{-1}(\boldsymbol{y}) \right\rangle .$$

A function $f : \mathbb{M} \to \mathbb{R}$ is said to be *geodesically μ_g-strongly concave* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{M}$, it holds that

$$f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \left\langle \mathtt{grad} f(\boldsymbol{x}), \mathtt{Exp}_{\boldsymbol{x}}^{-1}(\boldsymbol{y}) \right\rangle - \frac{\mu_g}{2} \cdot d_g(\boldsymbol{x}, \boldsymbol{y})^2.$$

Key Results for the Linear Convergence on $\mathbb{M}$

▶ **Question:** Why is the linear convergence of gradient ascent methods on $\mathbb{M}$ only valid around a small neighborhood of $\boldsymbol{m}_k \in \mathcal{M}$?

• **Geodesically Strong Concavity**. A differentiable function $f : \mathbb{M} \to \mathbb{R}$ is said to be *geodesically concave* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{M}$, it holds that

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \leq \left\langle \texttt{grad} f(\boldsymbol{x}), \texttt{Exp}_{\boldsymbol{x}}^{-1}(\boldsymbol{y}) \right\rangle .$$

A function $f : \mathbb{M} \to \mathbb{R}$ is said to be *geodesically μ_g-strongly concave* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{M}$, it holds that

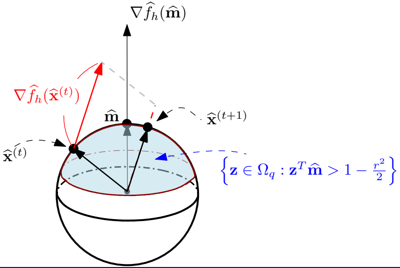
$$f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \left\langle \texttt{grad} f(\boldsymbol{x}), \texttt{Exp}_{\boldsymbol{x}}^{-1}(\boldsymbol{y}) \right\rangle - \frac{\mu_g}{2} \cdot d_g(\boldsymbol{x}, \boldsymbol{y})^2.$$

- Compact manifolds, such as spheres, do not admit globally geodesically convex functions other than the constant function (Yau, 1974).

1. The directional mean shift algorithms can be viewed as a variant of the gradient ascent methods on Ω_q but with adaptive step sizes.

$$\widehat{\boldsymbol{x}}^{(t+1)} = \frac{\nabla \widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})}{\left|\left|\nabla \widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})\right|\right|_2} = \mathtt{Exp}_{\widehat{\boldsymbol{x}}^{(t)}}\left[\eta_t \cdot \mathtt{grad}\widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})\right] \quad \text{for } t = 0, 1, ...,$$

where $\eta_t = \frac{\widehat{\theta}_t}{\left|\left|\nabla \widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})\right|\right|_2 \sin\left(\widehat{\theta}_t\right)}$ with $\widehat{\theta}_t$ being the angle between $\widehat{\boldsymbol{x}}^{(t)}$ and $\nabla \widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})$.



2. The step size

$$\eta_t = \frac{\widehat{\theta}_t}{\left|\left|\nabla \widehat{f}_h(\widehat{\boldsymbol{x}}^{(t)})\right|\right|_2 \sin\left(\widehat{\theta}_t\right)}$$

can be made sufficiently small when the bandwidth h is small and the sample size n is large, but is also universally bounded away from 0 with respect to the iteration number t.

3. Therefore, the linear convergence of directional mean shift algorithm follow from the previous results for gradient ascent methods on $\mathbb{M}$.

Generalization to SCGA Algorithms (on $\mathbb{M}$)

To establish the linear convergence of subspace constrained gradient ascent methods (on $\mathbb{M}$), we intend to generalize the (geodesically) strong concavity to its subspace constrained version.

- Subspace constrained gradient ascent (SCGA) algorithm in $\mathbb{R}^D$:

$$\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} + \eta \cdot V_d(\boldsymbol{x}^{(t)}) V_d(\boldsymbol{x}^{(t)})^T \nabla f(\boldsymbol{x}^{(t)}).$$

- Subspace constrained gradient ascent (SCGA) algorithm on $\mathbb{M}$:

$$\boldsymbol{x}^{(t+1)} = \mathtt{Exp}_{\boldsymbol{x}^{(t)}} \left[\eta \cdot V_d(\boldsymbol{x}^{(t)}) V_d(\boldsymbol{x}^{(t)})^T \mathtt{grad} f(\boldsymbol{x}^{(t)}) \right].$$

Generalization to SCGA Algorithms (on $\mathbb{M}$)

To establish the linear convergence of subspace constrained gradient ascent methods (on $\mathbb{M}$), we intend to generalize the (geodesically) strong concavity to its subspace constrained version.

- Subspace constrained gradient ascent (SCGA) algorithm in $\mathbb{R}^D$:

$$\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} + \eta \cdot V_d(\boldsymbol{x}^{(t)}) V_d(\boldsymbol{x}^{(t)})^T \nabla f(\boldsymbol{x}^{(t)}).$$

- Subspace constrained gradient ascent (SCGA) algorithm on $\mathbb{M}$:

$$\boldsymbol{x}^{(t+1)} = \mathtt{Exp}_{\boldsymbol{x}^{(t)}} \left[\eta \cdot V_d(\boldsymbol{x}^{(t)}) V_d(\boldsymbol{x}^{(t)})^T \mathtt{grad} f(\boldsymbol{x}^{(t)}) \right].$$

Subspace constrained strong concavity (SCSC):

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) \leq \nabla f(\boldsymbol{y})^T V_d(\boldsymbol{y}) V_d(\boldsymbol{y})^T (\boldsymbol{x} - \boldsymbol{y}) - \frac{\mu}{2} ||\boldsymbol{x} - \boldsymbol{y}||_2^2$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \text{domain}(f)$ and some constant $\mu > 0$.

Regularity Conditions for SCSC

- **(A1)** (*Eigengap*) There exist constants $\rho > 0$ and $\beta_0 > 0$ such that $\lambda_{d+1}(\boldsymbol{y}) \leq -\beta_0$ and $\lambda_d(\boldsymbol{y}) - \lambda_{d+1}(\boldsymbol{y}) \geq \beta_0$ for any $\boldsymbol{y} \in R_d \oplus \rho$.

- **(A2)** (*Path Smoothness*) There exists a constant $\beta_1 \in (0, \beta_0)$ such that

$$D^{\frac{3}{2}} \left|\left| U_d^{\perp}(\boldsymbol{y}) \nabla f(\boldsymbol{y}) \right|\right|_2 \left|\left| \nabla^3 f(\boldsymbol{y}) \right|\right|_{\max} \leq \frac{\beta_0^2}{2},$$
$$d \cdot D^{\frac{3}{2}} \left|\left| \nabla f(\boldsymbol{x}) \right|\right|_2 \left|\left| \nabla^3 f(\boldsymbol{x}) \right|\right|_{\max} \leq \beta_0(\beta_0 - \beta_1)$$

  for all $\boldsymbol{y} \in R_d \oplus \rho$ and $\boldsymbol{x} \in R_d$, where $U_d^{\perp}(\boldsymbol{y}) = \boldsymbol{I}_D - V_d(\boldsymbol{y}) V_d(\boldsymbol{y})^T$.

- **(A3)** (*Quadratic Behaviors of Residual Vectors*) There exists a constant $\beta_2 > 0$ such that the SCGA sequence $\left\{\boldsymbol{x}^{(t)}\right\}_{t=0}^{\infty}$ with step size $0 < \eta \leq \min\left\{\frac{4}{\beta_0}, \frac{1}{D||f||_{\infty}^{(2)}}\right\}$ and $\boldsymbol{x}^* \in R_d$ as its limiting point satisfies

$$\nabla f(\boldsymbol{x}^{(t)})^T U_d^{\perp}(\boldsymbol{x}^{(t)})(\boldsymbol{x}^* - \boldsymbol{x}^{(t)}) \leq \frac{\beta_0}{4} \left|\left| \boldsymbol{x}^* - \boldsymbol{x}^{(t)} \right|\right|_2^2,$$
$$\left|\left| U_d^{\perp}(\boldsymbol{x}^{(t)})(\boldsymbol{x}^* - \boldsymbol{x}^{(t)}) \right|\right|_2 \leq \beta_2 \left|\left| \boldsymbol{x}^* - \boldsymbol{x}^{(t)} \right|\right|_2^2.$$

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 $\frac{nh^{q+4}}{|\log h|} \to 0$.

Experiments of the Euclidean SCMS Algorithm

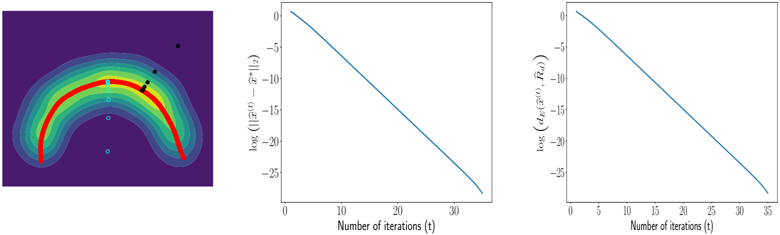


Figure: Density ridges estimated by the Euclidean SCMS algorithm on a simulated dataset and its (linear) convergence plot.

Experiments of the Directional SCMS Algorithm

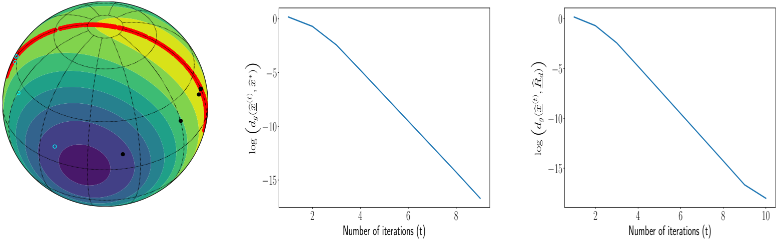


Figure: Density ridges estimated by the directional SCMS algorithm performed on a simulated dataset and its (linear) convergence plots.

Future Research Directions

Future Works on SCGA Algorithms on Manifolds

1. The current linear convergence theory of the SCGA method on $\mathbb{R}^D$ or general manifolds requires some regularity condition on the iterative sequence $\{\boldsymbol{x}^{(t)}\}_{t=0}^{\infty}$; recall Assumption (A3).

   - It is worth thinking about the conditions on f under which the SCGA algorithm (on manifolds) can converge linearly.

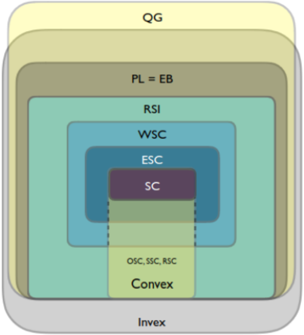
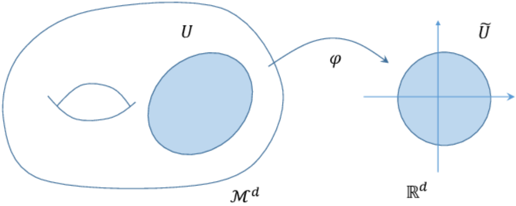


Figure: Different conditions in optimization theory (Karimi et al., 2016).

2. Can the stochastic gradient ascent methods bring some benefits towards the above iteration?

3. More broadly, the subspace projection matrix $V_d(\boldsymbol{x})V_d(\boldsymbol{x})^T$ does not need to come from the Hessian of f.

   - An interesting application of the above problem is the coordinate descent algorithm on manifolds (Gutman and Ho-Nguyen, 2023; Peng and Vidal, 2023).



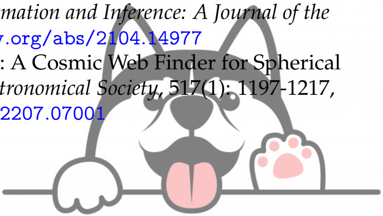
Thank you!

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. Linear Convergence of the Subspace Constrained Mean Shift Algorithm: From Euclidean to Directional Data. *Information and Inference: A Journal of the IMA*, 12(1): 210-311, 2022. https://arxiv.org/abs/2104.14977

[3] Y. Zhang, R. S. de Souza, and Y.-C. Chen. SCONCE: A Cosmic Web Finder for Spherical and Conic Geometries. *Monthly Notices of the Royal Astronomical Society*, 517(1): 1197-1217, 2022. https://arxiv.org/abs/2207.07001



Reference

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.

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.

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

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

C. R. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. Nonparametric ridge estimation. *The Annals of Statistics*, 42(4):1511 – 1545, 2014.

D. H. Gutman and N. Ho-Nguyen. Coordinate descent without coordinates: Tangent subspace descent on riemannian manifolds. *Mathematics of Operations Research*, 48(1): 127–159, 2023.

Reference

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.

M. T. Harandi, C. Sanderson, R. Hartley, and B. C. Lovell. Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part II 12*, pages 216–229. Springer, 2012.

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.

R. Hosseini and S. Sra. Matrix manifold optimization for gaussian mixtures. *Advances in neural information processing systems*, 28, 2015.

A. J. Izenman. Introduction to manifold learning. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(5):439–446, 2012.

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

Reference

H. Karimi, J. Nutini, and M. Schmidt. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part I 16*, pages 795–811. Springer, 2016.

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.

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.

L. Lin, D. Lazar, B. Sarpabayeva, and D. B. Dunson. Robust optimization and inference on manifolds. *arXiv preprint arXiv:2006.06843*, 2020.

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.

J. McQueen, M. Meilă, J. VanderPlas, and Z. Zhang. Megaman: Scalable manifold learning in python. *The Journal of Machine Learning Research*, 17(1):5176–5180, 2016.

B. Mishra, G. Meyer, F. Bach, and R. Sepulchre. Low-rank optimization with trace norm penalty. *SIAM Journal on Optimization*, 23(4):2124–2149, 2013.

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

L. Peng and R. Vidal. Block coordinate descent on smooth manifolds. *arXiv preprint arXiv:2305.14744*, 2023.

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

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.

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

B. Vandereycken. Low-rank matrix completion by riemannian optimization. *SIAM Journal on Optimization*, 23(2):1214–1236, 2013.

S.-T. Yau. Non-existence of continuous convex functions on certain riemannian manifolds. *Mathematische Annalen*, 207:269–270, 1974.

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

H. Zhang and S. Sra. First-order methods for geodesically convex optimization. In *Conference on Learning Theory*, pages 1617–1638. PMLR, 2016.

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

Y. Zhang and Y.-C. Chen. 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. *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.

Drawback of 3D Cosmic Web Detection 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).

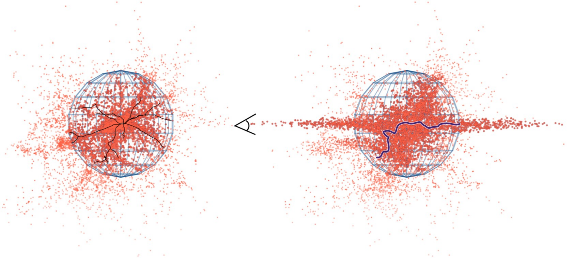


Figure: Redshift distortions along the line of sight (Kuchner et al., 2021).

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

2D Method for Detecting filaments
Slicing the Universe (Tomographic Analysis)

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

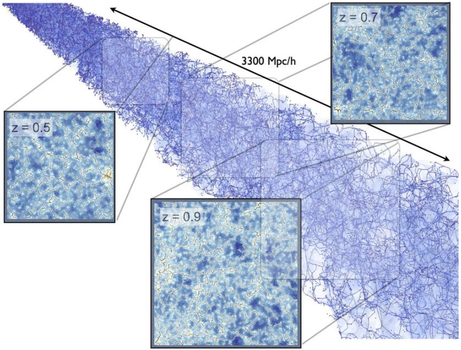


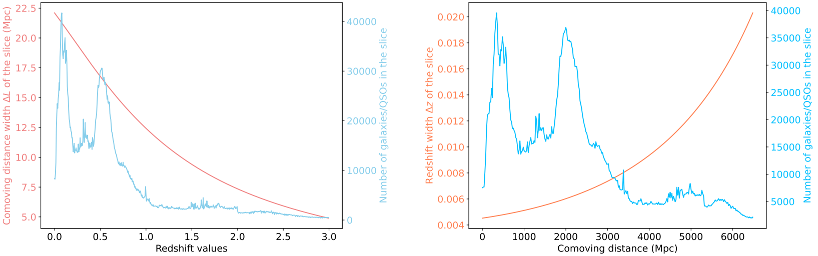
Figure: Illustration of slicing the Universe (credit to Laigle et al. 2018)

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 $\Delta L = 20\,\text{Mpc}$.

Under some regularity conditions (Hall et al., 1987; Bai et al., 1988; García-Portugués, 2013; Zhang and Chen, 2021b), 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 $nh^q \to \infty$;

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

  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 $\frac{nh^q}{\log n} \to \infty$;

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

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

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

Detailed Procedures of DirSCMS Algorithm

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

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

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

Detailed Procedures of DirSCMS Algorithm

- **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.