

Gradient-Guided Density Peak Clustering

Yikun Zhang^{1,2}

Joint work with *Professor Yen-Chi Chen*³

¹*Department of Statistics, University of Chicago*

²*NSF-Simons AI Institute for the Sky (SkAI Institute)*

³*Department of Statistics, University of Washington*

Computational and Applied Mathematics Colloquium
October 29, 2026

Background and Introduction



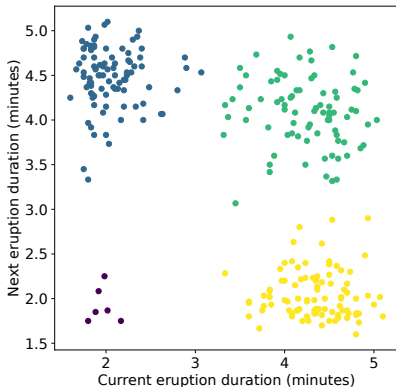
Clustering: a data analysis technique that partitions a set of objects into groups (called **clusters**) based on their similarity.

$$\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\} \subset \mathbb{R}^d \quad \implies \quad \hat{\mathcal{C}}_1, \dots, \hat{\mathcal{C}}_M \quad \text{with} \quad \bigcup_{j=1}^M \hat{\mathcal{C}}_j = \mathbb{X}_n.$$

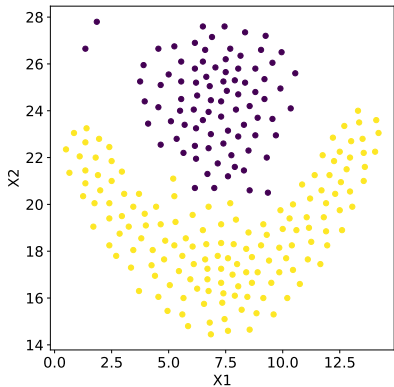
Clustering: Unsupervised Machine Learning

Clustering: a data analysis technique that partitions a set of objects into groups (called **clusters**) based on their similarity.

$$\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\} \subset \mathbb{R}^d \quad \Rightarrow \quad \hat{\mathcal{C}}_1, \dots, \hat{\mathcal{C}}_M \quad \text{with} \quad \bigcup_{i=1}^M \hat{\mathcal{C}}_i = \mathbb{X}_n.$$



Old Faithful dataset ([Azzalini and Bowman, 1990](#)).



Flame dataset ([Fu and Medico, 2007](#)).

What is Density Peak Clustering (DPC)?

- ① DPC (Rodriguez and Laio, 2014) is a type of density-based clustering:

Partition $\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ $\xRightarrow{\text{based on}}$ density function $p : \mathbb{R}^d \rightarrow \mathbb{R}$ of $\mathbb{X}_n$.

- Construct $\hat{p}$ via k -nearest neighbors (k NN) or a kernel density estimator (KDE).

What is Density Peak Clustering (DPC)?

- ① DPC (Rodriguez and Laio, 2014) is a type of density-based clustering:

Partition $\mathbb{X}_n = \{X_1, \dots, X_n\}$ $\xrightarrow{\text{based on}}$ density function $p : \mathbb{R}^d \rightarrow \mathbb{R}$ of $\mathbb{X}_n$.

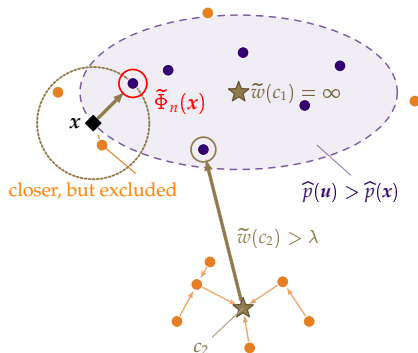
- Construct $\hat{p}$ via k -nearest neighbors (k NN) or a kernel density estimator (KDE).

- ② Given $\mathbb{X}_n = \{X_1, \dots, X_n\}$ and any $x \in \mathbb{R}^d$, define

- 1NN with higher density:

$$\tilde{\Phi}_n(x) = \arg \min_{X_i \in \mathbb{X}_n} \{ \|X_i - x\| : \hat{p}(X_i) > \hat{p}(x) \}.$$

- 1NN uphill distance: $\tilde{w}(x) = \left\| \tilde{\Phi}_n(x) - x \right\|$ or ∞ .



What is Density Peak Clustering (DPC)?

- ① DPC (Rodriguez and Laio, 2014) is a type of density-based clustering:

Partition $\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ $\xrightarrow{\text{based on}}$ density function $p : \mathbb{R}^d \rightarrow \mathbb{R}$ of $\mathbb{X}_n$.

- Construct $\hat{p}$ via k -nearest neighbors (k NN) or a kernel density estimator (KDE).

- ② Given $\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ and any $x \in \mathbb{R}^d$, define

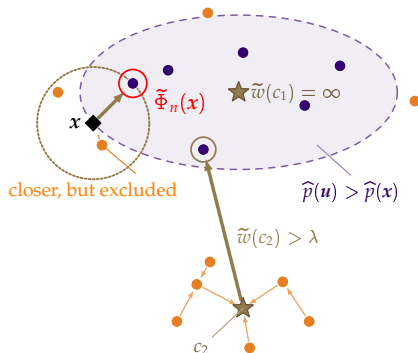
- 1NN with higher density:

$$\tilde{\Phi}_n(x) = \arg \min_{\mathbf{X}_i \in \mathbb{X}_n} \{ \|\mathbf{X}_i - x\| : \hat{p}(\mathbf{X}_i) > \hat{p}(x) \}.$$

- 1NN uphill distance: $\tilde{w}(x) = \|\tilde{\Phi}_n(x) - x\|$ or ∞ .

- ③ Identify the cluster centers

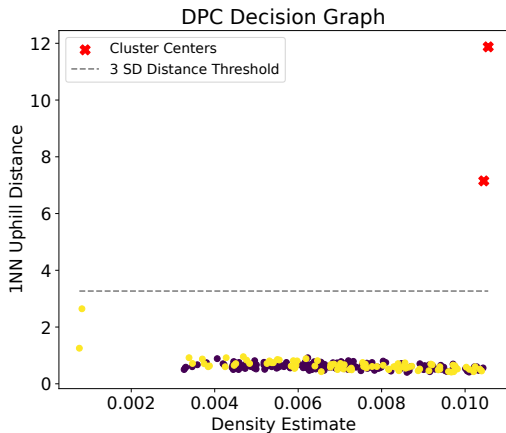
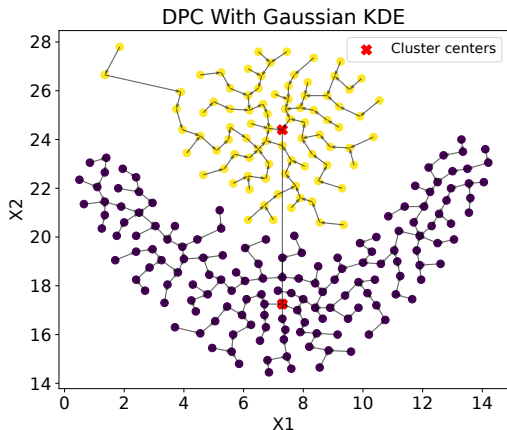
$$\tilde{\mathcal{M}}_\lambda = \{\mathbf{X}_i \in \mathbb{X}_n : \tilde{w}(\mathbf{X}_i) > \lambda\}, \quad \lambda > 0.$$



Cluster Center Identification in DPC

Identify the cluster centers

$$\widetilde{\mathcal{M}}_{\lambda} = \{X_i \in \mathbb{X}_n : \widetilde{w}(X_i) > \lambda\}, \quad \lambda > 0.$$



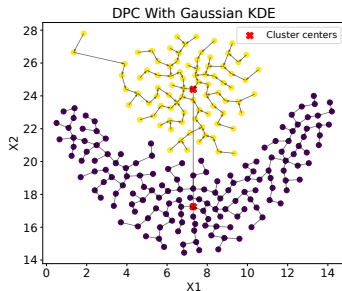
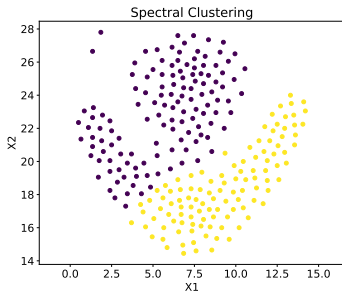
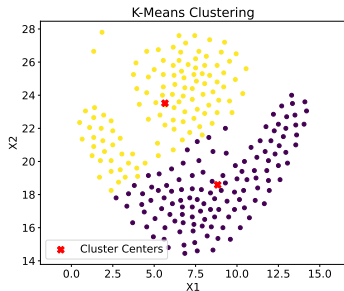
Question: Why do we study or even use DPC?

Question: Why do we study or even use DPC?

- 1 Unlike k -means clustering ([MacQueen, 1967](#); [Lloyd, 1982](#)) and spectral clustering ([von Luxburg, 2007](#)), DPC does not require a pre-specified number of clusters.

Question: Why do we study or even use DPC?

- 1 Unlike k -means clustering (MacQueen, 1967; Lloyd, 1982) and spectral clustering (von Luxburg, 2007), DPC does not require a pre-specified number of clusters.

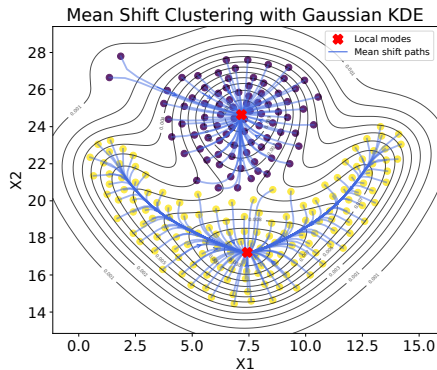


- 2 More adaptive to irregular cluster shapes.

The gradient ascent flow of p starting at $\mathbf{x}$ is defined as

$$\boldsymbol{\gamma}'_x(t) = \nabla p(\boldsymbol{\gamma}_x(t)), \quad \boldsymbol{\gamma}_x(0) = \mathbf{x}.$$

- $\mathcal{M} = \{m_1, \dots, m_M\}$ is the set of local modes of p .

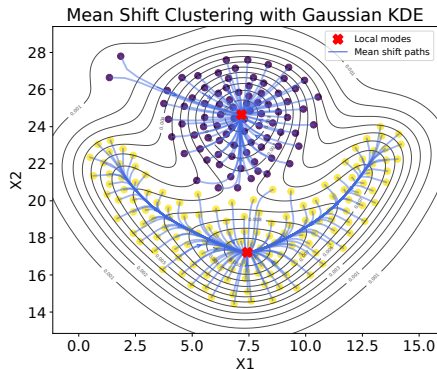


The gradient ascent flow of p starting at $\mathbf{x}$ is defined as

$$\boldsymbol{\gamma}'_x(t) = \nabla p(\boldsymbol{\gamma}_x(t)), \quad \boldsymbol{\gamma}_x(0) = \mathbf{x}.$$

- $\mathcal{M} = \{\mathbf{m}_1, \dots, \mathbf{m}_M\}$ is the set of local modes of p .
- Population clusters (or basins of attraction):

$$\mathcal{C}_j = \left\{ \mathbf{x} \in \mathbb{R}^d : \lim_{t \rightarrow \infty} \boldsymbol{\gamma}_x(t) = \mathbf{m}_j \right\}, j = 1, \dots, M.$$



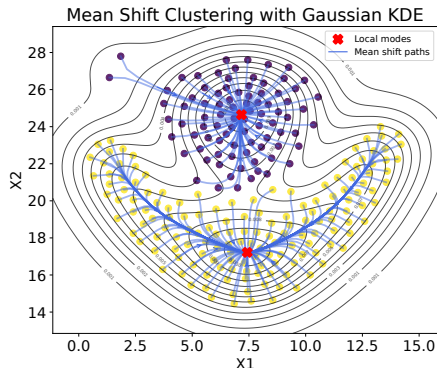
The gradient ascent flow of p starting at $\mathbf{x}$ is defined as

$$\boldsymbol{\gamma}'_x(t) = \nabla p(\boldsymbol{\gamma}_x(t)), \quad \boldsymbol{\gamma}_x(0) = \mathbf{x}.$$

- $\mathcal{M} = \{\mathbf{m}_1, \dots, \mathbf{m}_M\}$ is the set of local modes of p .
- Population clusters (or basins of attraction):

$$\mathcal{C}_j = \left\{ \mathbf{x} \in \mathbb{R}^d : \lim_{t \rightarrow \infty} \boldsymbol{\gamma}_x(t) = \mathbf{m}_j \right\}, j = 1, \dots, M.$$

- Estimate p by KDE and obtain $\widehat{\mathcal{M}}$ by the mean shift algorithm (Fukunaga and Hostetler, 1975; Comaniciu and Meer, 2002).



DPC vs. Mean Shift Clustering

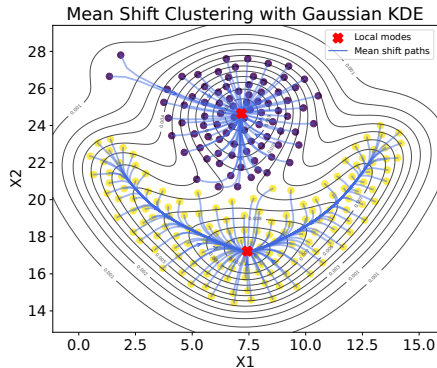
The gradient ascent flow of p starting at $\mathbf{x}$ is defined as

$$\boldsymbol{\gamma}'_x(t) = \nabla p(\boldsymbol{\gamma}_x(t)), \quad \boldsymbol{\gamma}_x(0) = \mathbf{x}.$$

- $\mathcal{M} = \{\mathbf{m}_1, \dots, \mathbf{m}_M\}$ is the set of local modes of p .
- Population clusters (or basins of attraction):

$$\mathcal{C}_j = \left\{ \mathbf{x} \in \mathbb{R}^d : \lim_{t \rightarrow \infty} \boldsymbol{\gamma}_x(t) = \mathbf{m}_j \right\}, j = 1, \dots, M.$$

- Estimate p by KDE and obtain $\widehat{\mathcal{M}}$ by the mean shift algorithm (Fukunaga and Hostetler, 1975; Comaniciu and Meer, 2002).



Drawback of Mean Shift Clustering:

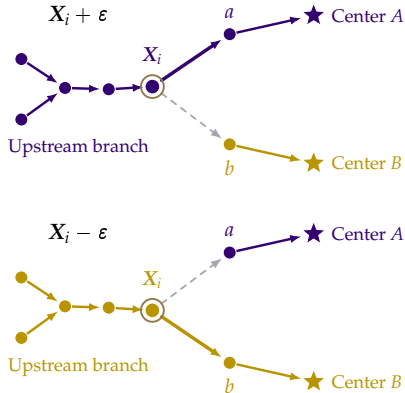
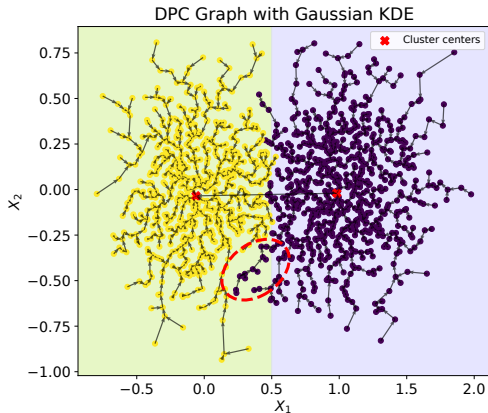
- From *every* X_i , it *repeatedly* estimates/evaluates the gradient until convergence.

However, 1NN shift can make DPC **unstable** in low-density regions.

Domino Effect in DPC

However, 1NN shift can make DPC **unstable** in low-density regions.

- Gaussian mixture: $0.5 \cdot \mathcal{N}(\boldsymbol{\mu}_1, 0.09\mathbf{I}_2) + 0.5 \cdot \mathcal{N}(\boldsymbol{\mu}_2, 0.09\mathbf{I}_2)$, $\boldsymbol{\mu}_1 = (0, 0)^T$, $\boldsymbol{\mu}_2 = (1, 0)^T$.



A single perturbation redirects the whole upstream branch (Xie et al., 2016; Seyedi et al., 2019).

Open Problem: Pathwise Consistency of DPC

Domino effect obscures a *population counterpart* of the DPC path from $\mathbf{X}_i \in \mathbb{X}_n$ to a cluster center in $\widetilde{\mathcal{M}}_\lambda$.

Domino effect obscures a *population counterpart* of the DPC path from $\mathbf{X}_i \in \mathbb{X}_n$ to a cluster center in $\widetilde{\mathcal{M}}_\lambda$.

December 2025

Clustering by hill-climbing: Consistency results

[Ery Arias-Castro](#), [Wanli Qiao](#)

[Author Affiliations -](#)

Ery Arias-Castro,¹ Wanli Qiao²

¹Department of Mathematics and Halicioglu Data Science Institute, University of California

²Department of Statistics, George Mason University

Ann. Statist. 53(6): 2536-2562 (December 2025). DOI: 10.1214/25-AOS2554

ABOUT	FIRST PAGE	CITED BY	REFERENCES	SUPPLEMENTAL CONTENT
-------	------------	----------	------------	----------------------

Abstract

We consider several hill-climbing approaches to clustering as formulated by Fukunaga and Hostetler (*IEEE Trans. Inf. Theory* **IT-21** (1975) 32–40) in the 1970s. We study both continuous-space and discrete-space (i.e., medoid) variants and establish their consistency.

Section 5.4.2: “A population analog of quick shift[†] is not *straightforward to define* ... As far as we know, proving (or disproving) the consistency of quick shift over the entire basins of attraction *remains an open problem.*”

Domino effect obscures a *population counterpart* of the DPC path from $\mathbf{X}_i \in \mathbb{X}_n$ to a cluster center in $\widetilde{\mathcal{M}}_\lambda$.

December 2025

Clustering by hill-climbing: Consistency results

Ery Arias-Castro, Wanli Qiao

Author Affiliations -

Ery Arias-Castro,¹ Wanli Qiao²

¹Department of Mathematics and Halicioglu Data Science Institute, University of California

²Department of Statistics, George Mason University

Ann. Statist. 53(6): 2536-2562 (December 2025). DOI: 10.1214/25-AOS2554

ABOUT	FIRST PAGE	CITED BY	REFERENCES	SUPPLEMENTAL CONTENT
-------	------------	----------	------------	----------------------

Abstract

We consider several hill-climbing approaches to clustering as formulated by Fukunaga and Hostetler (*IEEE Trans. Inf. Theory* **IT-21** (1975) 32–40) in the 1970s. We study both continuous-space and discrete-space (i.e., medoid) variants and establish their consistency.

Section 5.4.2: “A population analog of quick shift[†] is not *straightforward to define* ... As far as we know, proving (or disproving) the consistency of quick shift over the entire basins of attraction *remains an open problem.*”

[†]Quick shift (Vedaldi and Soatto, 2008) is essentially the same algorithm as DPC.

Goal: Resolve the instability of DPC with a minimal fix and study its (pathwise) stability theory.

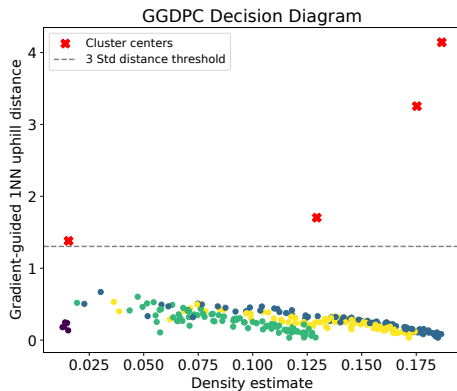
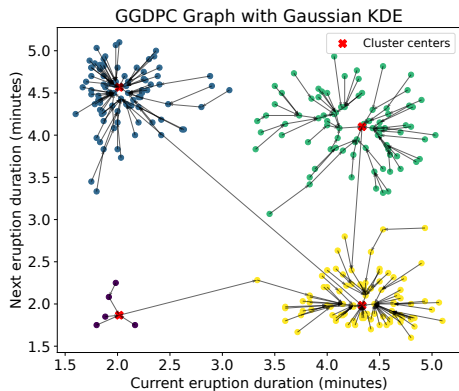
Overview of Today's Talk

High-Level Idea: Use a **one-step gradient ascent** from each observation X_i to **guide** its 1NN uphill search.

Overview of Today's Talk

High-Level Idea: Use a **one-step gradient ascent** from each observation X_i to **guide** its 1NN uphill search.

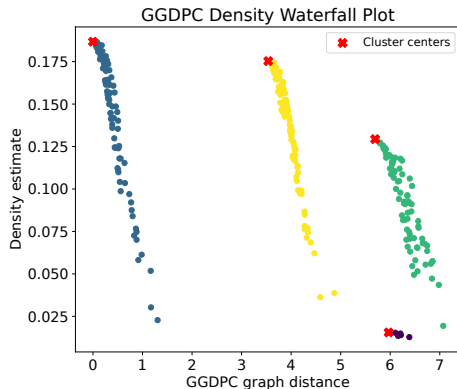
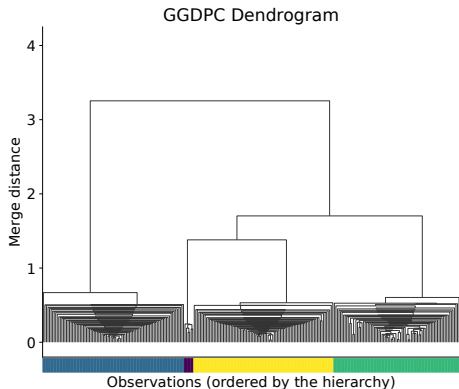
① **“Good Clustering”:** convergence of density modes and adjusted Rand index (ARI).



Overview of Today's Talk

High-Level Idea: Use a **one-step gradient ascent** from each observation X_i to **guide** its 1NN uphill search.

- 1 “*Good Clustering*”: convergence of density modes and adjusted Rand index (ARI).
- 2 “*Informative Clustering*”:



High-Level Idea: Use a **one-step gradient ascent** from each observation X_i to **guide** its 1NN uphill search.

① *“Good Clustering”*: convergence of density modes and adjusted Rand index (ARI).

② *“Informative Clustering”*:

- GGDPC dendrogram $\xRightarrow{\text{Gromov-Hausdorff}}$ Modal distance dendrogram.
- Empirical density waterfall measure $\xRightarrow{\text{Wasserstein-1}}$ Population waterfall measure.

High-Level Idea: Use a **one-step gradient ascent** from each observation X_i to **guide** its 1NN uphill search.

① *“Good Clustering”*: convergence of density modes and adjusted Rand index (ARI).

② *“Informative Clustering”*:

- GGDPC dendrogram $\xRightarrow{\text{Gromov-Hausdorff}}$ Modal distance dendrogram.
- Empirical density waterfall measure $\xRightarrow{\text{Wasserstein-1}}$ Population waterfall measure.

③ *“Stable Clustering”*:

GGDPC ascending paths $\approx$ Gradient ascent flow of p as $n \rightarrow \infty$.

Gradient-Guided Density Peak Clustering (GGDPC)



Let $\hat{p}$ and $\hat{g}$ be estimators of p and its gradient ∇p . Given $\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$,

$$\text{DPC: } \tilde{\Phi}_n(\mathbf{x}) = \arg \min_{\mathbf{X}_i \in \mathbb{X}_n} \{ \|\mathbf{X}_i - \mathbf{x}\| : \hat{p}(\mathbf{X}_i) > \hat{p}(\mathbf{x}) \}, \quad \tilde{w}(\mathbf{x}) = \left\| \tilde{\Phi}_n(\mathbf{x}) - \mathbf{x} \right\|.$$

Gradient-Guided 1NN Uphill Search

Let $\hat{p}$ and $\hat{g}$ be estimators of p and its gradient ∇p . Given $\mathbb{X}_n = \{X_1, \dots, X_n\}$,

DPC: $\tilde{\Phi}_n(x) = \arg \min_{X_i \in \mathbb{X}_n} \{ \|X_i - x\| : \hat{p}(X_i) > \hat{p}(x) \}, \quad \tilde{w}(x) = \left\| \tilde{\Phi}_n(x) - x \right\|.$

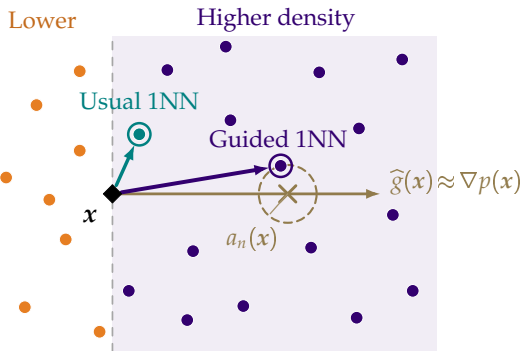
GGDPC:

- Guided 1NN with higher density:

$$a_n(x) = x + \eta_n \hat{g}(x), \quad \eta_n > 0,$$

$$\hat{\Phi}_n(x) = \arg \min_{X_i \in \mathbb{X}_n} \{ \|X_i - a_n(x)\| : \hat{p}(X_i) > \hat{p}(x) \},$$

$$\hat{w}_n(x) = \left\| \hat{\Phi}_n(x) - x \right\| \text{ or } \infty.$$



Gradient-Guided 1NN Uphill Search

Let $\hat{p}$ and $\hat{g}$ be estimators of p and its gradient ∇p . Given $\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$,

DPC: $\tilde{\Phi}_n(\mathbf{x}) = \arg \min_{\mathbf{X}_i \in \mathbb{X}_n} \{ \|\mathbf{X}_i - \mathbf{x}\| : \hat{p}(\mathbf{X}_i) > \hat{p}(\mathbf{x}) \}, \quad \tilde{w}(\mathbf{x}) = \|\tilde{\Phi}_n(\mathbf{x}) - \mathbf{x}\|.$

GGDPC:

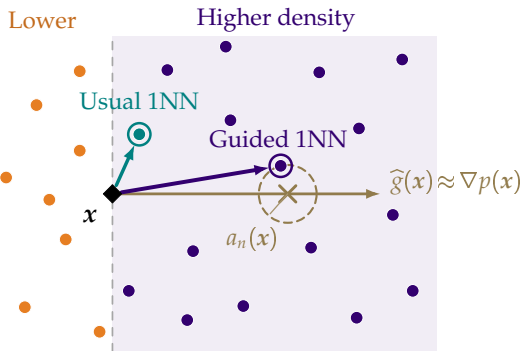
- Guided 1NN with higher density:

$$\mathbf{a}_n(\mathbf{x}) = \mathbf{x} + \eta_n \hat{g}(\mathbf{x}), \quad \eta_n > 0,$$

$$\hat{\Phi}_n(\mathbf{x}) = \arg \min_{\mathbf{X}_i \in \mathbb{X}_n} \{ \|\mathbf{X}_i - \mathbf{a}_n(\mathbf{x})\| : \hat{p}(\mathbf{X}_i) > \hat{p}(\mathbf{x}) \},$$

$$\hat{w}_n(\mathbf{x}) = \|\hat{\Phi}_n(\mathbf{x}) - \mathbf{x}\| \text{ or } \infty.$$

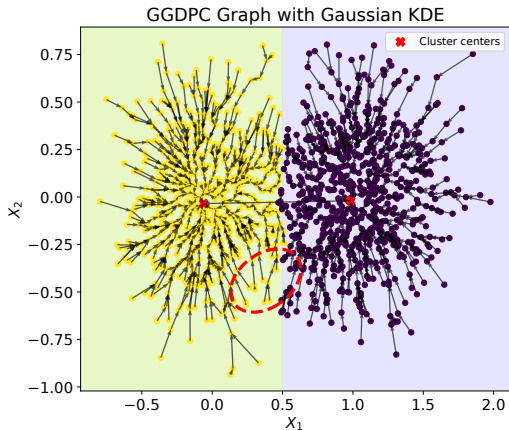
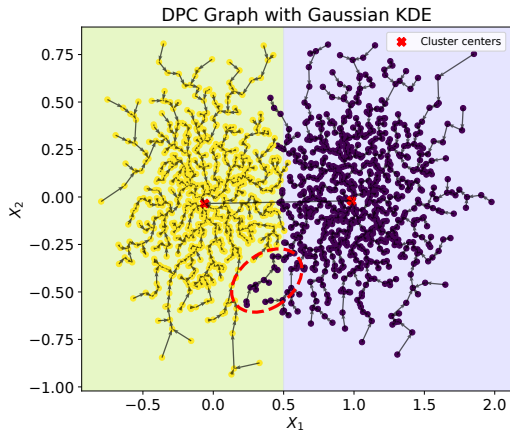
- Density is compared with the *original* point.
- $\hat{\mathcal{M}}_\lambda = \{\mathbf{X}_i \in \mathbb{X}_n : \hat{w}_n(\mathbf{X}_i) > \lambda\}$ for $\lambda > 0$.



Comparisons Between DPC and GGDPC

Gaussian mixture: $0.5 \cdot \mathcal{N}(\boldsymbol{\mu}_1, 0.09\mathbf{I}_2) + 0.5 \cdot \mathcal{N}(\boldsymbol{\mu}_2, 0.09\mathbf{I}_2)$, $\boldsymbol{\mu}_1 = (0, 0)^T$, $\boldsymbol{\mu}_2 = (1, 0)^T$.

- $n = 1500$ and population separatrix is $x_1 = 0.5$.



On the same sample, the GGDPC graph follows the modal basins better.

Convergence of GGDPC Clustering

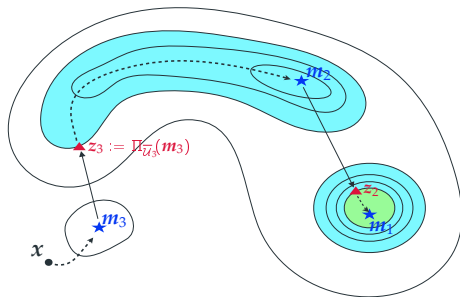


Question: Is $\widehat{\mathcal{M}}_\lambda = \{\mathbf{X}_i \in \mathbb{X}_n : \widehat{w}_n(\mathbf{X}_i) > \lambda\}$ statistically consistent?

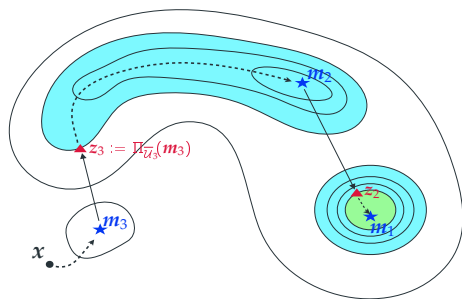
Question: Is $\widehat{\mathcal{M}}_\lambda = \{\mathbf{X}_i \in \mathbb{X}_n : \widehat{w}_n(\mathbf{X}_i) > \lambda\}$ statistically consistent?

Assume p is Morse with a compact support.

- $\mathcal{M} = \{\mathbf{m}_1, \dots, \mathbf{m}_M\} : p(\mathbf{m}_1) > \dots > p(\mathbf{m}_M)$.



Question: Is $\widehat{\mathcal{M}}_\lambda = \{\mathbf{X}_i \in \mathbb{X}_n : \widehat{w}_n(\mathbf{X}_i) > \lambda\}$ statistically consistent?



Assume p is Morse with a compact support.

- $\mathcal{M} = \{\mathbf{m}_1, \dots, \mathbf{m}_M\} : p(\mathbf{m}_1) > \dots > p(\mathbf{m}_M)$.

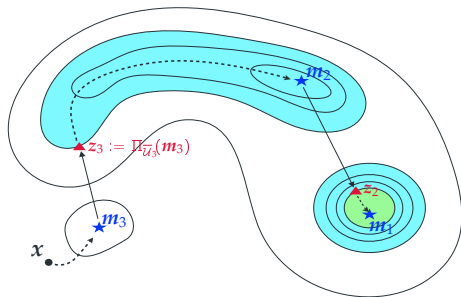
- Define $\overline{\mathcal{U}}_j := \overline{\{x \in \mathbb{R}^d : p(x) > p(\mathbf{m}_j)\}}$ and

$$w(\mathbf{m}_j) := d(\mathbf{m}_j, \overline{\mathcal{U}}_j).$$

- Population λ -separated upper mode set:

$$\mathcal{M}_\lambda = \{\mathbf{m}_j \in \mathcal{M} : w(\mathbf{m}_j) > \lambda\}.$$

Question: Is $\widehat{\mathcal{M}}_\lambda = \{\mathbf{X}_i \in \mathbb{X}_n : \widehat{w}_n(\mathbf{X}_i) > \lambda\}$ statistically consistent?



Assume p is Morse with a compact support.

- $\mathcal{M} = \{\mathbf{m}_1, \dots, \mathbf{m}_M\}$: $p(\mathbf{m}_1) > \dots > p(\mathbf{m}_M)$.

- Define $\overline{\mathcal{U}}_j := \overline{\{x \in \mathbb{R}^d : p(x) > p(\mathbf{m}_j)\}}$ and

$$w(\mathbf{m}_j) := d(\mathbf{m}_j, \overline{\mathcal{U}}_j).$$

- Population λ -separated upper mode set:

$$\mathcal{M}_\lambda = \{\mathbf{m}_j \in \mathcal{M} : w(\mathbf{m}_j) > \lambda\}.$$

Hausdorff Distance:

$$\text{Haus}(\widehat{\mathcal{M}}_\lambda, \mathcal{M}_\lambda) = \max \left\{ \sup_{x \in \widehat{\mathcal{M}}_\lambda} d(x, \mathcal{M}_\lambda), \sup_{y \in \mathcal{M}_\lambda} d(y, \widehat{\mathcal{M}}_\lambda) \right\}.$$

Theorem (Zhang and Chen 2026; see also Jiang 2017; Verdinelli and Wasserman 2018)

- ① p is a Morse function supported on a compact regular set $\mathcal{C}$, and $\inf_{x \in \mathcal{C}} p(x) \geq p_{\min} > 0$.
- ② $\eta_n = o(1)$, $\|\hat{p} - p\|_{\infty} = o_P(1)$, $\|\hat{g} - \nabla p\|_{\infty} = o_P(1)$, and $q_n := \left(\frac{\log n}{n}\right)^{1/d}$.

Theorem (Zhang and Chen 2026; see also Jiang 2017; Verdinelli and Wasserman 2018)

- ① p is a Morse function supported on a compact regular set $\mathcal{C}$, and $\inf_{x \in \mathcal{C}} p(x) \geq p_{\min} > 0$.
- ② $\eta_n = o(1)$, $\|\hat{p} - p\|_{\infty} = o_P(1)$, $\|\hat{g} - \nabla p\|_{\infty} = o_P(1)$, and $q_n := \left(\frac{\log n}{n}\right)^{1/d}$.

Then, for any $\lambda \notin \{w(\mathbf{m}_j) : 1 < j \leq M\}$,

$$|\widehat{\mathcal{M}}_{\lambda}| = |\mathcal{M}_{\lambda}| \quad w.p. \rightarrow 1, \quad \text{Haus}(\mathcal{M}_{\lambda}, \widehat{\mathcal{M}}_{\lambda}) = O_P\left(q_n + \|\hat{p} - p\|_{\infty}^{1/2}\right).$$

Theorem (Zhang and Chen 2026; see also Jiang 2017; Verdinelli and Wasserman 2018)

- ① p is a Morse function supported on a compact regular set $\mathcal{C}$, and $\inf_{x \in \mathcal{C}} p(x) \geq p_{\min} > 0$.
- ② $\eta_n = o(1)$, $\|\hat{p} - p\|_{\infty} = o_P(1)$, $\|\hat{g} - \nabla p\|_{\infty} = o_P(1)$, and $q_n := \left(\frac{\log n}{n}\right)^{1/d}$.

Then, for any $\lambda \notin \{w(\mathbf{m}_j) : 1 < j \leq M\}$,

$$|\widehat{\mathcal{M}}_{\lambda}| = |\mathcal{M}_{\lambda}| \quad w.p. \rightarrow 1, \quad \text{Haus}(\mathcal{M}_{\lambda}, \widehat{\mathcal{M}}_{\lambda}) = O_P(q_n + \|\hat{p} - p\|_{\infty}^{1/2}).$$

For a smooth estimator $\hat{p}$ with $\hat{g} = \nabla \hat{p}$,

$$\text{Haus}(\mathcal{M}_{\lambda}, \widehat{\mathcal{M}}_{\lambda}) = O_P(q_n + \min\{\|\hat{p} - p\|_{\infty}^{1/2}, \|\hat{g} - \nabla p\|_{\infty}\}).$$

- Minimavity when $d \leq 6$ (Romano, 1988; Arias-Castro et al., 2022):

$$\|\hat{g} - \nabla p\|_{\infty} = O_P\left(\left(\frac{\log n}{n}\right)^{2/(d+6)}\right) \implies \text{Haus}(\mathcal{M}_{\lambda}, \widehat{\mathcal{M}}_{\lambda}) = O_P\left(\left(\frac{\log n}{n}\right)^{\min\{2/(d+6), 1/d\}}\right).$$

Question: Does GGDPC agree with the population (oracle) clustering?

Question: Does GGDPC agree with the population (oracle) clustering?

- Population basins of attraction $\mathcal{C}_1, \dots, \mathcal{C}_M$ of p :

$$\mathcal{C}_j = \left\{ x \in \mathcal{C} : \lim_{t \rightarrow \infty} \boldsymbol{\gamma}_x(t) = \boldsymbol{m}_j \right\}, \quad \boldsymbol{\gamma}'_x(t) = \nabla p(\boldsymbol{\gamma}_x(t)) \quad \text{and} \quad \boldsymbol{\gamma}_x(0) = x.$$

- Define the population modal partition of $\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ as

$$\mathcal{P}_n^* := \{\mathbb{X}_n \cap \mathcal{C}_1, \dots, \mathbb{X}_n \cap \mathcal{C}_M\}.$$

Question: Does GGDPC agree with the population (oracle) clustering?

- Population basins of attraction $\mathcal{C}_1, \dots, \mathcal{C}_M$ of p :

$$\mathcal{C}_j = \left\{ x \in \mathcal{C} : \lim_{t \rightarrow \infty} \boldsymbol{\gamma}_x(t) = \mathbf{m}_j \right\}, \quad \boldsymbol{\gamma}'_x(t) = \nabla p(\boldsymbol{\gamma}_x(t)) \quad \text{and} \quad \boldsymbol{\gamma}_x(0) = x.$$

- Define the population modal partition of $\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ as

$$\mathcal{P}_n^* := \{\mathbb{X}_n \cap \mathcal{C}_1, \dots, \mathbb{X}_n \cap \mathcal{C}_M\}.$$

- GGDPC has its own partition $\hat{\mathcal{P}}_{n,\lambda}$ of $\mathbb{X}_n$.

Question: Does GGDPC agree with the population (oracle) clustering?

- Population basins of attraction $\mathcal{C}_1, \dots, \mathcal{C}_M$ of p :

$$\mathcal{C}_j = \left\{ x \in \mathcal{C} : \lim_{t \rightarrow \infty} \boldsymbol{\gamma}_x(t) = \mathbf{m}_j \right\}, \quad \boldsymbol{\gamma}'_x(t) = \nabla p(\boldsymbol{\gamma}_x(t)) \quad \text{and} \quad \boldsymbol{\gamma}_x(0) = x.$$

- Define the population modal partition of $\mathbb{X}_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ as

$$\mathcal{P}_n^* := \{\mathbb{X}_n \cap \mathcal{C}_1, \dots, \mathbb{X}_n \cap \mathcal{C}_M\}.$$

- GGDPC has its own partition $\hat{\mathcal{P}}_{n,\lambda}$ of $\mathbb{X}_n$.

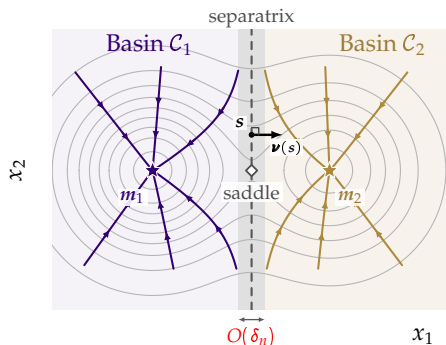
ARI (Hubert and Arabie, 1985): $\text{ARI}(\hat{\mathcal{P}}_{n,\lambda}, \mathcal{P}_n^*) = \frac{\text{RI} - \mathbb{E}_0(\text{RI})}{1 - \mathbb{E}_0(\text{RI})}$ with $\text{RI} = \frac{TP + TN}{\binom{n}{2}}$.

- ARI = 1: identical clusterings; $\mathbb{E}_0[\text{ARI}] = 0$: random labeling.

Normal Repulsion Condition: For $s \in \text{Reg}(\partial\mathcal{C}_j)$,

$$\exists \rho_S > 0 \quad \text{s.t.} \quad \nu(s)^T \nabla^2 p(s) \nu(s) \geq \rho_S,$$

$\nu(s)$: a unit normal vector to $\partial\mathcal{C}_j$ pointing into $\mathcal{C}_j$.

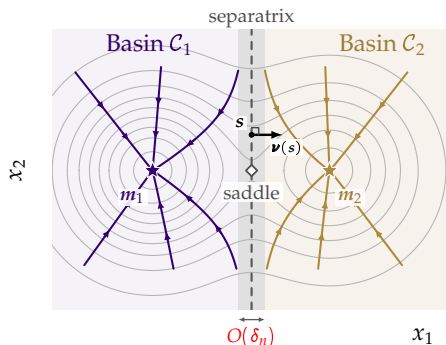


Normal Repulsion Condition: For $s \in \text{Reg}(\partial\mathcal{C}_j)$,

$$\exists \rho_S > 0 \quad \text{s.t.} \quad \nu(s)^T \nabla^2 p(s) \nu(s) \geq \rho_S,$$

$\nu(s)$: a unit normal vector to $\partial\mathcal{C}_j$ pointing into $\mathcal{C}_j$.

- Keep paths in their basin outside an $O(\delta_n)$ tube.
- The sample mass in the tube is $O_P(\delta_n)$, $\delta_n \downarrow 0$.



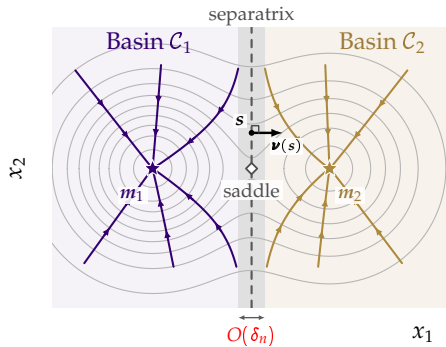
Convergence of Adjusted Rand Index (ARI)

Normal Repulsion Condition: For $s \in \text{Reg}(\partial\mathcal{C}_j)$,

$$\exists \rho_s > 0 \quad \text{s.t.} \quad \nu(s)^T \nabla^2 p(s) \nu(s) \geq \rho_s,$$

$\nu(s)$: a unit normal vector to $\partial\mathcal{C}_j$ pointing into $\mathcal{C}_j$.

- Keep paths in their basin outside an $O(\delta_n)$ tube.
- The sample mass in the tube is $O_P(\delta_n)$, $\delta_n \downarrow 0$.



Theorem (Zhang and Chen 2026)

- 1 Morse function p + normal repulsion + smooth estimator $\hat{p}$ with $\hat{g} = \nabla \hat{p}$.
- 2 $\eta_n + \frac{q_n}{\eta_n} = o(1)$, $\|\hat{p} - p\|_\infty = o_P(1)$, $\|\hat{g} - \nabla p\|_\infty = o_P(1)$, and $q_n := \left(\frac{\log n}{n}\right)^{1/d}$.

$$1 - \text{ARI}(\hat{\mathcal{P}}_{n,\lambda}, \mathcal{P}_n^*) = O_P\left(\eta_n + \frac{q_n}{\eta_n} + \|\hat{g} - \nabla p\|_\infty\right), \quad \forall \lambda \in \left(0, \min_{2 \leq j \leq |\mathcal{M}|} \psi_j\right).$$

Convergence Rates: Modal Consistency vs. ARI

Let $r_n = \frac{\log n}{n}$, $q_n = r_n^{\frac{1}{d}}$, and $e_n = \|\widehat{g} - \nabla p\|_\infty$.

Criterion	Step size	Error bound
Modal consistency (Hausdorff)	$\eta_n = o(1)$	$O_P(q_n + e_n)$
Clustering (1 - ARI)	$\eta_n \asymp \sqrt{q_n}$	$O_P(\sqrt{q_n} + e_n)$

Convergence Rates: Modal Consistency vs. ARI

Let $r_n = \frac{\log n}{n}$, $q_n = r_n^{\frac{1}{d}}$, and $e_n = \|\widehat{g} - \nabla p\|_\infty$.

Criterion	Step size	Error bound
Modal consistency (Hausdorff)	$\eta_n = o(1)$	$O_P(q_n + e_n)$
Clustering (1 - ARI)	$\eta_n \asymp \sqrt{q_n}$	$O_P(\sqrt{q_n} + e_n)$

For (boundary-corrected) KDE $\widehat{p}_n$ with $e_n = O_P\left(r_n^{\frac{2}{d+6}}\right)$:

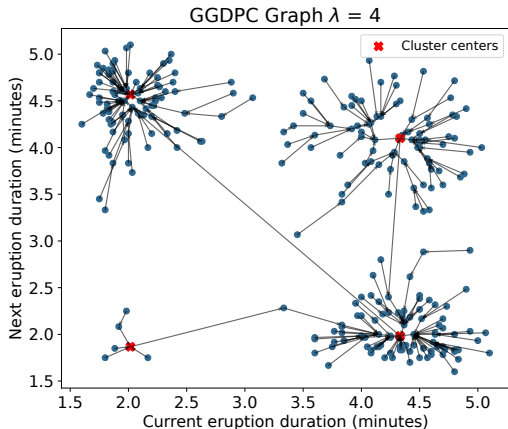
Criterion	Rate	Interpretation
Modal consistency (Hausdorff)	$O_P\left(r_n^{\min\{\frac{2}{d+6}, \frac{1}{d}\}}\right)$	Minimax up to logs for $d \leq 6$
Clustering (1 - ARI)	$O_P\left(r_n^{\min\{\frac{2}{d+6}, \frac{1}{2d}\}}\right)$	1NN error dominates for $d \geq 3$

GGDPC Dendrogram



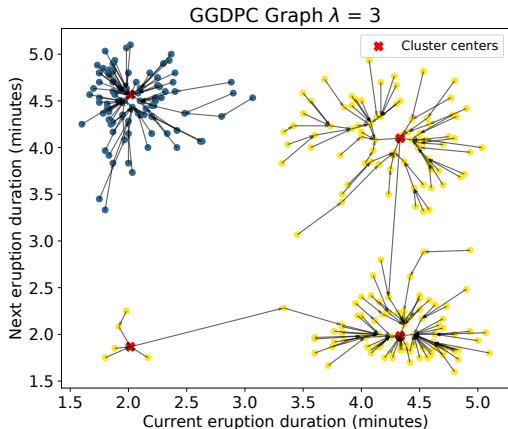
Hierarchy of the GGDPC Graph

G_λ : trim the outgoing edge from *every* X_i with $\hat{w}_n(X_i) > \lambda$ in the GGDPC graph G .



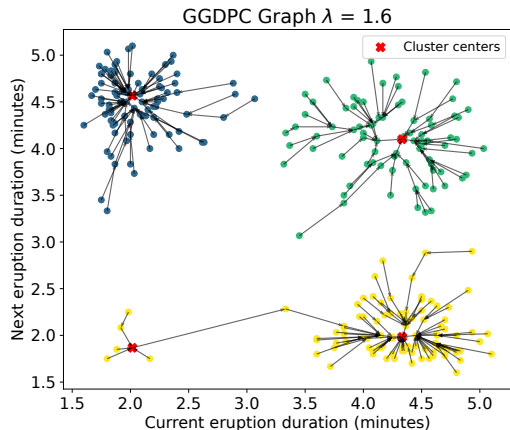
Hierarchy of the GGDPC Graph

G_λ : trim the outgoing edge from *every* X_i with $\hat{w}_n(X_i) > \lambda$ in the GGDPC graph G .



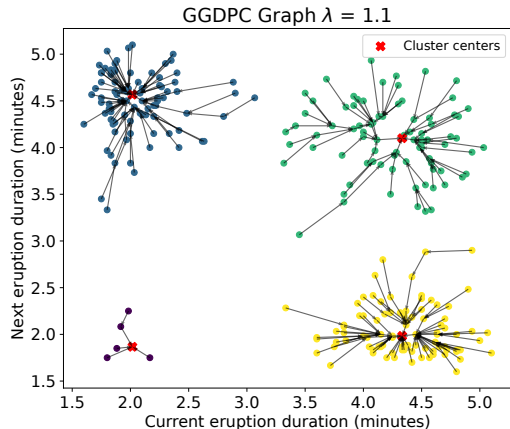
Hierarchy of the GGDPC Graph

G_λ : trim the outgoing edge from *every* X_i with $\hat{w}_n(X_i) > \lambda$ in the GGDPC graph G .



Hierarchy of the GGDPC Graph

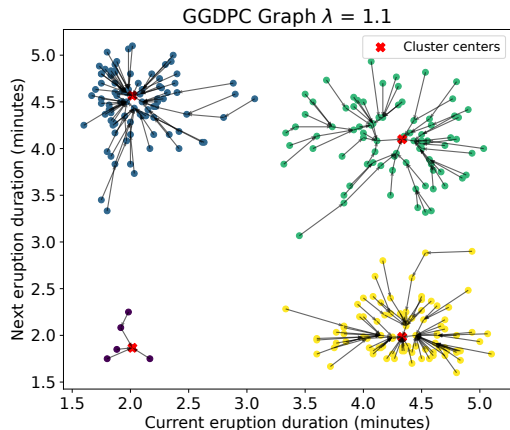
G_λ : trim the outgoing edge from *every* X_i with $\hat{w}_n(X_i) > \lambda$ in the GGDPC graph G .



- G_λ induces a partition of $\mathbb{X}_n$ at each $\lambda > 0$.

Hierarchy of the GGDPC Graph

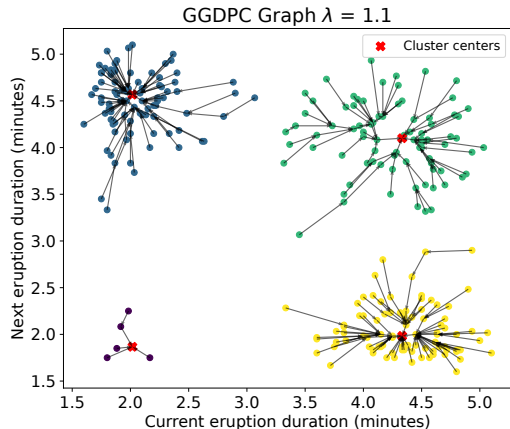
G_λ : trim the outgoing edge from *every* X_i with $\hat{w}_n(X_i) > \lambda$ in the GGDPC graph G .



- G_λ induces a partition of $\mathbb{X}_n$ at each $\lambda > 0$.
- Define the GGDPC dendrogram by $\mathcal{T}_G : [0, \infty) \rightarrow \mathcal{P}(\mathbb{X}_n)$.

Hierarchy of the GGDPC Graph

G_λ : trim the outgoing edge from *every* X_i with $\hat{w}_n(X_i) > \lambda$ in the GGDPC graph G .



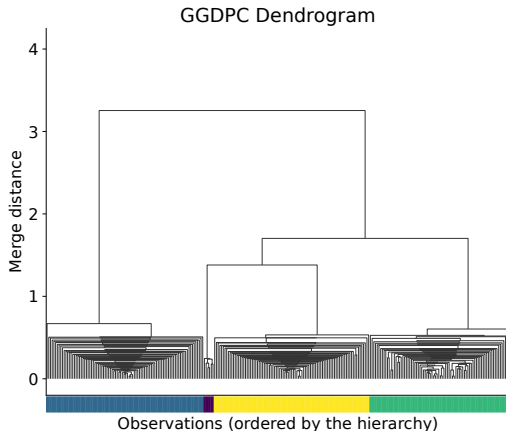
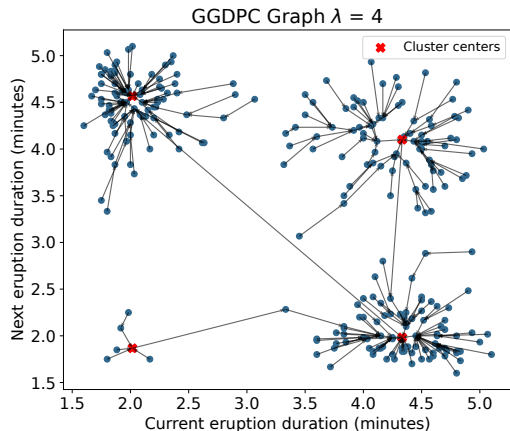
- G_λ induces a partition of $\mathbb{X}_n$ at each $\lambda > 0$.
- Define the GGDPC dendrogram by $\mathcal{T}_G : [0, \infty) \rightarrow \mathcal{P}(\mathbb{X}_n)$.

Hierarchy of $\mathcal{T}_G$:

- For any $0 \leq \lambda \leq \lambda' < \infty$, the partition $\mathcal{T}_G(\lambda)$ refines $\mathcal{T}_G(\lambda')$.

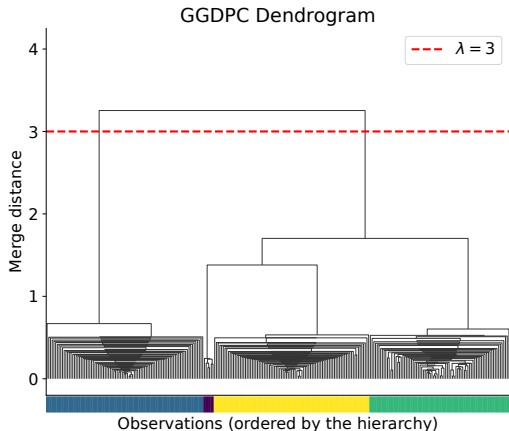
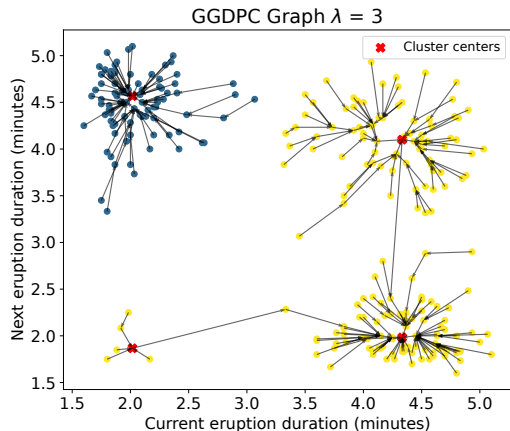
Computation of the GGDPC Dendrogram

Divisive Approach: Start from G at $\lambda \geq \max \hat{w}_n(X_i)$ and remove edges as $\lambda \downarrow 0$.



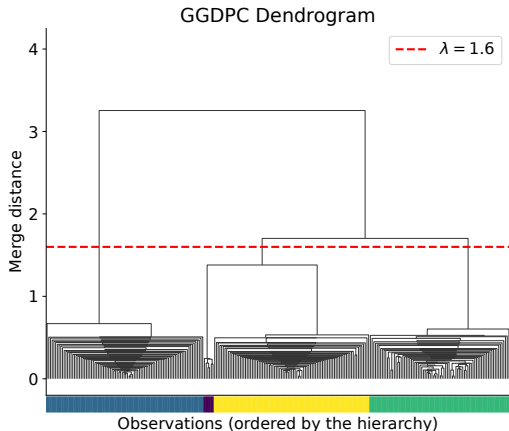
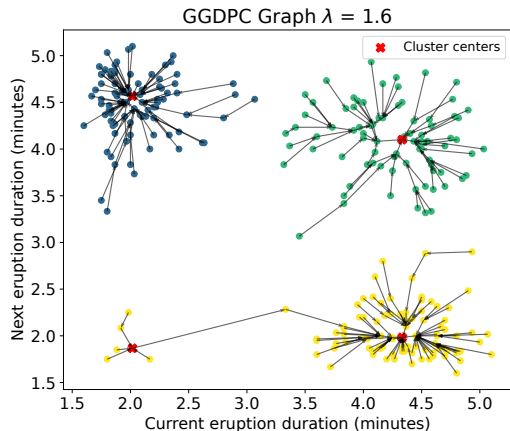
Computation of the GGDPC Dendrogram

Divisive Approach: Start from G at $\lambda \geq \max \hat{w}_n(X_i)$ and remove edges as $\lambda \downarrow 0$.



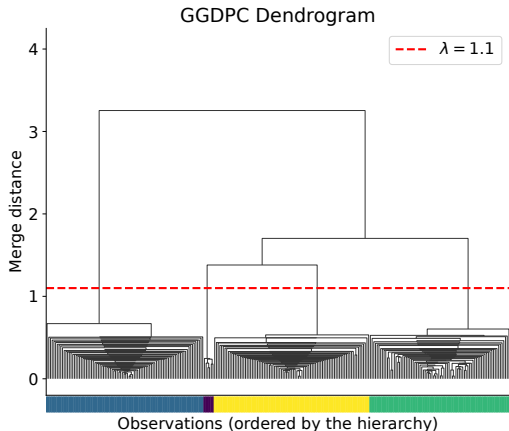
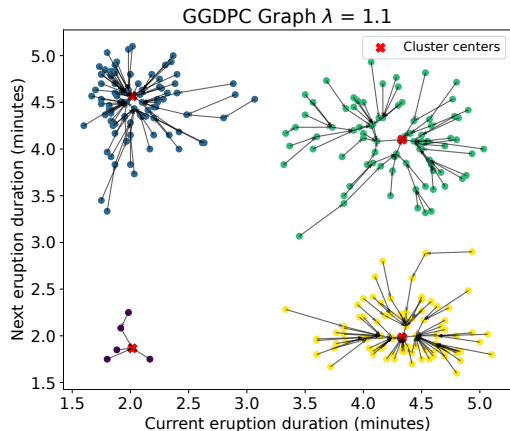
Computation of the GGDPC Dendrogram

Divisive Approach: Start from G at $\lambda \geq \max \hat{w}_n(X_i)$ and remove edges as $\lambda \downarrow 0$.



Computation of the GGDPC Dendrogram

Divisive Approach: Start from G at $\lambda \geq \max \hat{w}_n(X_i)$ and remove edges as $\lambda \downarrow 0$.



Divisive Approach: Start from G at $\lambda \geq \max \hat{w}_n(\mathbf{X}_i)$ and remove edges as $\lambda \downarrow 0$.

Agglomerative Approach: Choose $d_\star > d_{\max} = \max_{1 \leq i, j \leq n} \|\mathbf{X}_i - \mathbf{X}_j\|$ and define a *symmetric* matrix $D \in \mathbb{R}^{n \times n}$ with $D_{ii} = 0$ and

$$D_{ij} = D_{ji} = \begin{cases} \|\mathbf{X}_i - \mathbf{X}_j\| & \text{if } \mathbf{X}_i \text{ and } \mathbf{X}_j \text{ are adjacent in } G, \\ d_\star & \text{otherwise.} \end{cases}$$

Divisive Approach: Start from G at $\lambda \geq \max \widehat{w}_n(X_i)$ and remove edges as $\lambda \downarrow 0$.

Agglomerative Approach: Choose $d_\star > d_{\max} = \max_{1 \leq i, j \leq n} \|X_i - X_j\|$ and define a *symmetric* matrix $D \in \mathbb{R}^{n \times n}$ with $D_{ii} = 0$ and

$$D_{ij} = D_{ji} = \begin{cases} \|X_i - X_j\| & \text{if } X_i \text{ and } X_j \text{ are adjacent in } G, \\ d_\star & \text{otherwise.} \end{cases}$$

- Apply the single linkage clustering to D .

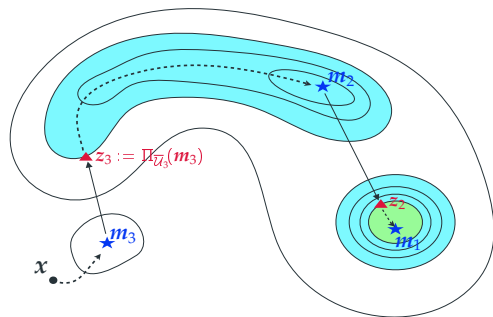
Proposition (Zhang and Chen 2026)

Let $\mathcal{T}_{\text{SL}}(\lambda)$ be the partition of $\mathbb{X}_n$ by cutting the single linkage dendrogram induced by D at height λ . Then,

$$\mathcal{T}_{\text{SL}}(\lambda) = \mathcal{T}_G(\lambda) \quad \text{for every } \lambda \geq 0.$$

Question: Does the GGDPC dendrogram converge to a population limit as $n \rightarrow \infty$?

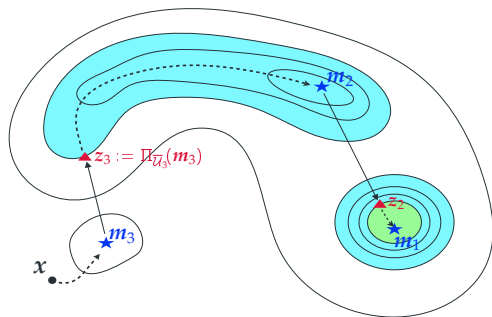
Question: Does the GGDPC dendrogram converge to a population limit as $n \rightarrow \infty$?



$$\bar{U}_j := \overline{\{x \in \mathbb{R}^d : p(x) > p(m_j)\}} \text{ and}$$

$$z_j := \arg \min_{x \in \bar{U}_j} w(m_j), \quad w(m_j) = d(m_j, \bar{U}_j).$$

Question: Does the GGDPC dendrogram converge to a population limit as $n \rightarrow \infty$?



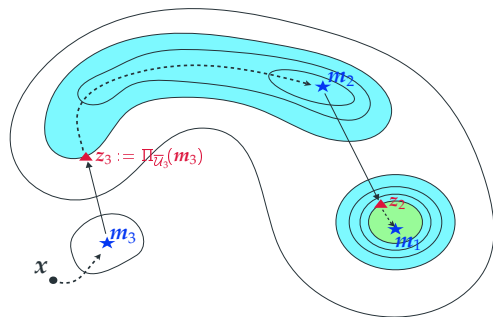
$$\bar{\mathcal{U}}_j := \overline{\{x \in \mathbb{R}^d : p(x) > p(m_j)\}} \text{ and}$$

$$z_j := \arg \min_{x \in \bar{\mathcal{U}}_j} w(m_j), \quad w(m_j) = d(m_j, \bar{\mathcal{U}}_j).$$

Define a weighted graph G_{mode} on $\mathcal{M}$:

- edges: $m_j \rightarrow m_{\pi(j)}$, $j = 2, \dots, M$ such that $z_j \in \mathcal{C}_{\pi(j)}$ (basin of attraction of $m_{\pi(j)}$);
- $m_j \rightarrow m_{\pi(j)}$ has weights $w(m_j)$.

Question: Does the GGDPC dendrogram converge to a population limit as $n \rightarrow \infty$?



$$\bar{\mathcal{U}}_j := \overline{\{x \in \mathbb{R}^d : p(x) > p(m_j)\}} \text{ and}$$

$$z_j := \arg \min_{x \in \bar{\mathcal{U}}_j} w(m_j), \quad w(m_j) = d(m_j, \bar{\mathcal{U}}_j).$$

Define a weighted graph G_{mode} on $\mathcal{M}$:

- edges: $m_j \rightarrow m_{\pi(j)}$, $j = 2, \dots, M$ such that $z_j \in \mathcal{C}_{\pi(j)}$ (basin of attraction of $m_{\pi(j)}$);
- $m_j \rightarrow m_{\pi(j)}$ has weights $w(m_j)$.

Modal Distance Dendrogram: define a mapping $\mathcal{T}_{\text{mode}} : [0, \infty) \rightarrow \mathcal{P}(\mathcal{M})$.

- $\mathcal{P}(\mathcal{M})$: the collection of all partitions of $\mathcal{M}$.

Dendrograms $\iff$ Ultrametric¹ on $\mathbb{X}_n$ (Carlsson and Mémoli, 2010):

$$\begin{aligned} u(\mathbf{X}_i, \mathbf{X}_j) &= \min \left\{ \lambda \geq 0 : \mathbf{X}_i, \mathbf{X}_j \text{ belong to the same block of } \mathcal{T}_G(\lambda) \right\} \\ &= \max \left\{ \hat{w}_n(\mathbf{X}_k) : (\mathbf{X}_k \rightarrow \hat{\Phi}_n(\mathbf{X}_k)) \in \text{Path}_G(\mathbf{X}_i, \mathbf{X}_j) \right\}. \end{aligned}$$

- $\text{Path}_G(\mathbf{X}_i, \mathbf{X}_j)$: the edge set in G joining $\mathbf{X}_i$ and $\mathbf{X}_j$.

¹An ultrametric $u : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_+$ is a valid metric and satisfies $\max \{u(x, z), u(z, y)\} \geq u(x, y)$ for any $x, y, z \in \mathbb{X}$.

Dendrograms $\iff$ Ultrametric¹ on $\mathbb{X}_n$ (Carlsson and Mémoli, 2010):

$$\begin{aligned} u(\mathbf{X}_i, \mathbf{X}_j) &= \min \left\{ \lambda \geq 0 : \mathbf{X}_i, \mathbf{X}_j \text{ belong to the same block of } \mathcal{T}_G(\lambda) \right\} \\ &= \max \left\{ \hat{w}_n(\mathbf{X}_k) : (\mathbf{X}_k \rightarrow \hat{\Phi}_n(\mathbf{X}_k)) \in \text{Path}_G(\mathbf{X}_i, \mathbf{X}_j) \right\}. \end{aligned}$$

- $\text{Path}_G(\mathbf{X}_i, \mathbf{X}_j)$: the edge set in G joining $\mathbf{X}_i$ and $\mathbf{X}_j$.

Similarly, $u_{\mathcal{M}}(\mathbf{m}_j, \mathbf{m}_k) = \max \left\{ w(\mathbf{m}_\ell) : (\mathbf{m}_\ell \rightarrow \mathbf{m}_{\pi(\ell)}) \in \text{Path}_{G_{\mathcal{M}}}(\mathbf{m}_j, \mathbf{m}_k) \right\}.$

¹An ultrametric $u : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_+$ is a valid metric and satisfies $\max \{u(x, z), u(z, y)\} \geq u(x, y)$ for any $x, y, z \in \mathbb{X}$.

Dendrograms $\iff$ **Ultrametric**¹ on $\mathbb{X}_n$ (Carlsson and Mémoli, 2010):

$$\begin{aligned} u(\mathbf{X}_i, \mathbf{X}_j) &= \min \left\{ \lambda \geq 0 : \mathbf{X}_i, \mathbf{X}_j \text{ belong to the same block of } \mathcal{T}_G(\lambda) \right\} \\ &= \max \left\{ \hat{w}_n(\mathbf{X}_k) : (\mathbf{X}_k \rightarrow \hat{\Phi}_n(\mathbf{X}_k)) \in \text{Path}_G(\mathbf{X}_i, \mathbf{X}_j) \right\}. \end{aligned}$$

- $\text{Path}_G(\mathbf{X}_i, \mathbf{X}_j)$: the edge set in G joining $\mathbf{X}_i$ and $\mathbf{X}_j$.

Similarly, $u_{\mathcal{M}}(\mathbf{m}_j, \mathbf{m}_k) = \max \left\{ w(\mathbf{m}_\ell) : (\mathbf{m}_\ell \rightarrow \mathbf{m}_{\pi(\ell)}) \in \text{Path}_{G_{\mathcal{M}}}(\mathbf{m}_j, \mathbf{m}_k) \right\}$.

Theorem (Gromov-Hausdorff convergence; Zhang and Chen 2026)

- 1 Morse function p + smooth estimator $\hat{p}$ + *no flat tangential region at $\mathbf{z}_j$* .
- 2 $\eta_n + \frac{q_n a_n}{\eta_n} = o(1)$, $\|\hat{p} - p\|_\infty = o_P(1)$, $\|\hat{g} - \nabla p\|_\infty = o_P(1)$, $\|\nabla^2 \hat{p} - \nabla^2 p\|_\infty = o_P(1)$.

$$d_{\text{GH}}((\mathcal{T}_G, \mathbb{X}_n), (\mathcal{T}_{\text{mode}}, \mathcal{M})) = O_P \left(\eta_n + a_n \left[\frac{q_n}{\eta_n} + \|\hat{g} - \nabla p\|_\infty \right] \right) \quad \text{for any } a_n \rightarrow \infty.$$

¹An ultrametric $u : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_+$ is a valid metric and satisfies $\max \{u(x, z), u(z, y)\} \geq u(x, y)$ for any $x, y, z \in \mathbb{X}$.

Convergence Rate Comparisons

Let $r_n = \frac{\log n}{n}$, $q_n = r_n^{1/d}$, $e_n = \|\widehat{g} - \nabla p\|_\infty$, and choose any deterministic sequence $a_n \rightarrow \infty$ slowly so that $a_n q_n = o(1)$ and $a_n e_n = o_P(1)$.

Criterion	Step size	Error bound
Modal consistency (Hausdorff)	$\eta_n = o(1)$	$O_P(q_n + e_n)$
Clustering (1 - ARI)	$\eta_n \asymp \sqrt{q_n}$	$O_P(\sqrt{q_n} + e_n)$
Dendrogram (d_{GH})	$\eta_n \asymp \sqrt{q_n}$	$O_P(\sqrt{a_n q_n} + a_n e_n)$

Convergence Rate Comparisons

Let $r_n = \frac{\log n}{n}$, $q_n = r_n^{1/d}$, $e_n = \|\hat{g} - \nabla p\|_\infty$, and choose any deterministic sequence $a_n \rightarrow \infty$ slowly so that $a_n q_n = o(1)$ and $a_n e_n = o_P(1)$.

Criterion	Step size	Error bound
Modal consistency (Hausdorff)	$\eta_n = o(1)$	$O_P(q_n + e_n)$
Clustering (1 - ARI)	$\eta_n \asymp \sqrt{q_n}$	$O_P(\sqrt{q_n} + e_n)$
Dendrogram (d_{GH})	$\eta_n \asymp \sqrt{q_n}$	$O_P(\sqrt{a_n q_n} + a_n e_n)$

For (boundary-corrected) KDE $\hat{p}_n$ with $e_n = O_P(r_n^{2/(d+6)})$:

Criterion	Rate	Interpretation
Modal consistency (Hausdorff)	$O_P\left(r_n^{\min\left\{\frac{2}{d+6}, \frac{1}{d}\right\}}\right)$	Minimax up to logs for $d \leq 6$
Clustering (1 - ARI)	$O_P\left(r_n^{\min\left\{\frac{2}{d+6}, \frac{1}{2d}\right\}}\right)$	1NN error dominates for $d \geq 3$
Dendrogram (d_{GH})	$O_P\left((\log \log n) r_n^{\min\left\{\frac{2}{d+6}, \frac{1}{2d}\right\}}\right)$	Same polynomial rate as ARI

Stability of GGDPC Paths and Density Waterfall



GGDPC uphill path from $\hat{\mathbf{Y}}_0 = \mathbf{x} \in \mathcal{C}_j$ to $\hat{\mathbf{Y}}_{k+1} = \hat{\Phi}_n(\hat{\mathbf{Y}}_k), k = 1, \dots, \hat{T}_n$:

$$\hat{L}_n(\mathbf{x}) := \sum_{k=0}^{\hat{T}_n-1} \left\| \hat{\mathbf{Y}}_{k+1} - \hat{\mathbf{Y}}_k \right\|.$$

- $\hat{\Phi}_n(\mathbf{x}) = \arg \min_{\mathbf{X}_i \in \mathbb{X}_n} \{ \|\mathbf{X}_i - [\mathbf{x} + \eta_n \hat{\mathbf{g}}(\mathbf{x})]\| : \hat{p}(\mathbf{X}_i) > \hat{p}(\mathbf{x}) \}.$
- $\hat{T}_n := \inf \left\{ k \geq 0 : \hat{\Phi}_n(\hat{\mathbf{Y}}_k) \text{ is undefined or does not belong to } \mathcal{C}_j \right\}.$

GGDPC uphill path from $\hat{\mathbf{Y}}_0 = \mathbf{x} \in \mathcal{C}_j$ to $\hat{\mathbf{Y}}_{k+1} = \hat{\Phi}_n(\hat{\mathbf{Y}}_k), k = 1, \dots, \hat{T}_n$:

$$\hat{L}_n(\mathbf{x}) := \sum_{k=0}^{\hat{T}_n-1} \left\| \hat{\mathbf{Y}}_{k+1} - \hat{\mathbf{Y}}_k \right\|.$$

- $\hat{\Phi}_n(\mathbf{x}) = \arg \min_{\mathbf{X}_i \in \mathbb{X}_n} \{ \|\mathbf{X}_i - [\mathbf{x} + \eta_n \hat{\mathbf{g}}(\mathbf{x})]\| : \hat{p}(\mathbf{X}_i) > \hat{p}(\mathbf{x}) \}.$
- $\hat{T}_n := \inf \left\{ k \geq 0 : \hat{\Phi}_n(\hat{\mathbf{Y}}_k) \text{ is undefined or does not belong to } \mathcal{C}_j \right\}.$

Gradient ascent flow of p from $\boldsymbol{\gamma}_x(0) = \mathbf{x}$ is $\boldsymbol{\gamma}'_x(t) = \nabla p(\boldsymbol{\gamma}_x(t))$:

$$L(\mathbf{x}) = \int_0^\infty \|\nabla p(\boldsymbol{\gamma}_x(t))\| dt.$$

GGDPC uphill path from $\hat{\mathbf{Y}}_0 = \mathbf{x} \in \mathcal{C}_j$ to $\hat{\mathbf{Y}}_{k+1} = \hat{\Phi}_n(\hat{\mathbf{Y}}_k), k = 1, \dots, \hat{T}_n$:

$$\hat{L}_n(\mathbf{x}) := \sum_{k=0}^{\hat{T}_n-1} \left\| \hat{\mathbf{Y}}_{k+1} - \hat{\mathbf{Y}}_k \right\|.$$

- $\hat{\Phi}_n(\mathbf{x}) = \arg \min_{\mathbf{X}_i \in \mathbb{X}_n} \{ \|\mathbf{X}_i - [\mathbf{x} + \eta_n \hat{g}(\mathbf{x})]\| : \hat{p}(\mathbf{X}_i) > \hat{p}(\mathbf{x}) \}.$
- $\hat{T}_n := \inf \left\{ k \geq 0 : \hat{\Phi}_n(\hat{\mathbf{Y}}_k) \text{ is undefined or does not belong to } \mathcal{C}_j \right\}.$

Gradient ascent flow of p from $\boldsymbol{\gamma}_x(0) = \mathbf{x}$ is $\boldsymbol{\gamma}'_x(t) = \nabla p(\boldsymbol{\gamma}_x(t))$:

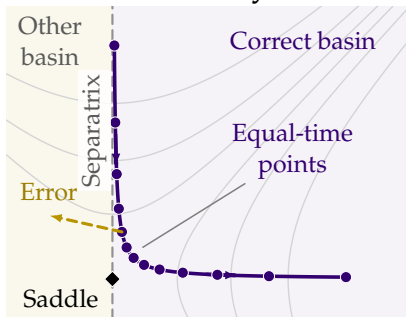
$$L(\mathbf{x}) = \int_0^\infty \|\nabla p(\boldsymbol{\gamma}_x(t))\| dt.$$

Section 5.4.2 in [Arias-Castro and Qiao \(2025\)](#): “... proving (or disproving) the consistency of quick shift over the entire basins of attraction *remains an open problem.*”

Question: Does $\sup_{\mathbf{x} \in \mathcal{C}_j} |\hat{L}_n(\mathbf{x}) - L(\mathbf{x})| \xrightarrow{P} 0$, and at what rate?

Challenges in Proving Path-Length Consistency

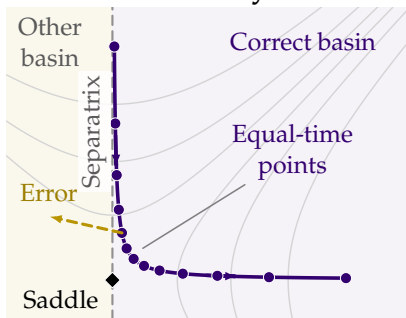
Near a boundary saddle



1NN and gradient errors may push GGDPC paths into the wrong basin.

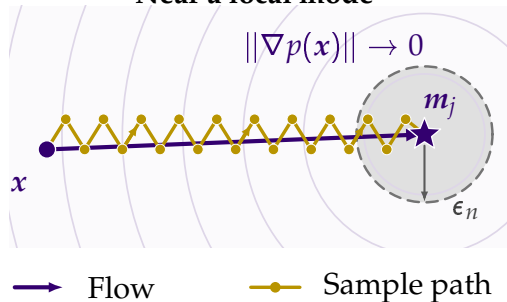
Challenges in Proving Path-Length Consistency

Near a boundary saddle



1NN and gradient errors may push GGDPC paths into the wrong basin.

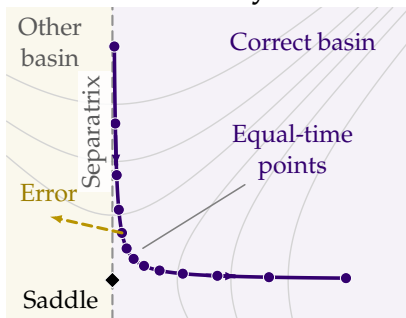
Near a local mode



1NN dominates gradient ascent, and sample paths can zigzag.

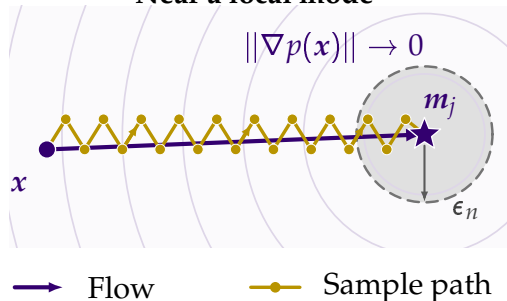
Challenges in Proving Path-Length Consistency

Near a boundary saddle



1NN and gradient errors may push GGDPC paths into the wrong basin.

Near a local mode



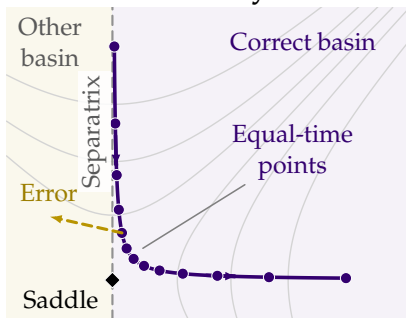
1NN dominates gradient ascent, and sample paths can zigzag.

- Besides normal repulsion, assume ∇p is C^1 -linearizable near boundary saddles s :

$$\boldsymbol{\gamma}'_x(t) = \nabla p(\boldsymbol{\gamma}_x(t)) \quad \xleftrightarrow{\text{smooth map } \Phi: \boldsymbol{y}(t) = \Phi(\boldsymbol{\gamma}_x(t))} \quad \boldsymbol{y}'(t) = A\boldsymbol{y}(t), \quad A = \nabla^2 p(s) \in \mathbb{R}^{d \times d}.$$

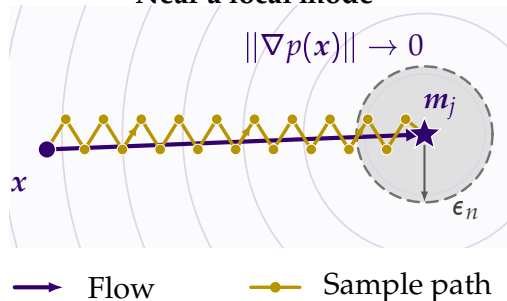
Challenges in Proving Path-Length Consistency

Near a boundary saddle



1NN and gradient errors may push GGDPC paths into the wrong basin.

Near a local mode



1NN dominates gradient ascent, and sample paths can zigzag.

- Besides normal repulsion, assume ∇p is C^1 -linearizable near boundary saddles s :

$$\gamma'_x(t) = \nabla p(\gamma_x(t)) \quad \xleftrightarrow{\text{smooth map } \Phi: y(t) = \Phi(\gamma_x(t))} \quad y'(t) = Ay(t), \quad A = \nabla^2 p(s) \in \mathbb{R}^{d \times d}.$$

- In a shrinking neighborhood $B(m_j, \epsilon_n)$, the total path length is $O_P(n\epsilon_n^d \cdot \epsilon_n)$.

Theorem (Zhang and Chen 2026)

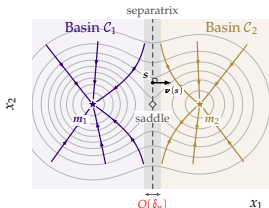
- ① Morse function p + normal repulsion + smooth $\hat{p}$ + C^1 -linearization on boundary saddles.
- ② For a deterministic sequence $\delta_n \downarrow 0$, assume
 - $\eta_n + \frac{q_n}{\eta_n} = o(\delta_n)$, $\delta_n \left| \log \left(\frac{q_n}{\eta_n} \right) \right| = O(1)$, $\frac{q_n \log n}{\eta_n^{d+1}} = o(1)$ with $q_n := \left(\frac{\log n}{n} \right)^{1/d}$;
 - $\|\hat{p} - p\|_\infty = o_P(1)$, $\|\hat{g} - \nabla p\|_\infty = o_P(\delta_n)$, $\|\nabla^2 \hat{p} - \nabla^2 p\| = o_P(1)$.

Theorem (Zhang and Chen 2026)

- ① Morse function p + normal repulsion + smooth $\hat{p}$ + C^1 -linearization on boundary saddles.
- ② For a deterministic sequence $\delta_n \downarrow 0$, assume
 - $\eta_n + \frac{q_n}{\eta_n} = o(\delta_n)$, $\delta_n \left| \log \left(\frac{q_n}{\eta_n} \right) \right| = O(1)$, $\frac{q_n \log n}{\eta_n^{d+1}} = o(1)$ with $q_n := \left(\frac{\log n}{n} \right)^{1/d}$;
 - $\|\hat{p} - p\|_\infty = o_P(1)$, $\|\hat{g} - \nabla p\|_\infty = o_P(\delta_n)$, $\|\nabla^2 \hat{p} - \nabla^2 p\| = o_P(1)$.

Then, for some constant $c > 0$,

$$\sup_{x \in \mathcal{C}_j: d(x, \partial \mathcal{C}_j) > c \delta_n} \left| \hat{L}_n(x) - L(x) \right| = O_P \left(\frac{q_n}{\eta_n} \left| \log \left(\frac{q_n}{\eta_n} \right) \right| + \frac{q_n}{\eta_n \delta_n} + \frac{\eta_n}{\delta_n} + \frac{q_n \log n}{\eta_n^{d+1}} + \frac{\|\hat{g} - \nabla p\|_\infty}{\delta_n} \right).$$



Theorem (Zhang and Chen 2026)

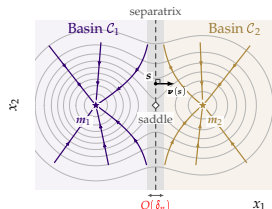
- 1 Morse function p + normal repulsion + smooth $\hat{p}$ + C^1 -linearization on boundary saddles.
- 2 For a deterministic sequence $\delta_n \downarrow 0$, assume

- $\eta_n + \frac{q_n}{\eta_n} = o(\delta_n)$, $\delta_n \left| \log \left(\frac{q_n}{\eta_n} \right) \right| = O(1)$, $\frac{q_n \log n}{\eta_n^{d+1}} = o(1)$ with $q_n := \left(\frac{\log n}{n} \right)^{1/d}$;
- $\|\hat{p} - p\|_\infty = o_P(1)$, $\|\hat{g} - \nabla p\|_\infty = o_P(\delta_n)$, $\|\nabla^2 \hat{p} - \nabla^2 p\| = o_P(1)$.

Then, for some constant $c > 0$,

$$\sup_{x \in \mathcal{C}_j: d(x, \partial \mathcal{C}_j) > c \delta_n} \left| \hat{L}_n(x) - L(x) \right| = O_P \left(\frac{q_n}{\eta_n} \left| \log \left(\frac{q_n}{\eta_n} \right) \right| + \frac{q_n}{\eta_n \delta_n} + \frac{\eta_n}{\delta_n} + \frac{q_n \log n}{\eta_n^{d+1}} + \frac{\|\hat{g} - \nabla p\|_\infty}{\delta_n} \right).$$

- **Cyan:** error rate away from local modes and boundary saddles.
- **Blue:** error rate near boundary saddles.
- **Orange:** Choose $\epsilon_n \asymp \frac{q_n}{\eta_n}$ so that $n \epsilon_n^{d+1} \asymp \frac{q_n \log n}{\eta_n^{d+1}}$.
- **Green:** gradient estimation error (near boundary saddles).



Convergence Rate Comparisons

Let $r_n = \frac{\log n}{n}$, $q_n = r_n^{1/d}$, $e_n = \|\widehat{g} - \nabla p\|_\infty$, and choose any deterministic sequence $a_n \rightarrow \infty$ slowly so that $a_n q_n = o(1)$ and $a_n e_n = o_P(1)$.

Path-Length Consistency: Choose $\eta_n \asymp (q_n \log n)^{\frac{1}{d+2}}$ and $\delta_n \asymp \sqrt{\eta_n + e_n}$,

$$\sup_{x \in \mathcal{C}_j: d(x, \partial \mathcal{C}_j) > c \delta_n} \left| \widehat{L}_n(x) - L(x) \right| = O_P \left(\sqrt{(q_n \log n)^{\frac{1}{d+2}} + e_n} \right).$$

Convergence Rate Comparisons

Let $r_n = \frac{\log n}{n}$, $q_n = r_n^{1/d}$, $e_n = \|\widehat{g} - \nabla p\|_\infty$, and choose any deterministic sequence $a_n \rightarrow \infty$ slowly so that $a_n q_n = o(1)$ and $a_n e_n = o_P(1)$.

Path-Length Consistency: Choose $\eta_n \asymp (q_n \log n)^{\frac{1}{d+2}}$ and $\delta_n \asymp \sqrt{\eta_n + e_n}$,

$$\sup_{x \in \mathcal{C}_j: d(x, \partial \mathcal{C}_j) > c \delta_n} |\widehat{L}_n(x) - L(x)| = O_P \left(\sqrt{(q_n \log n)^{\frac{1}{d+2}} + e_n} \right).$$

For (boundary-corrected) KDE $\widehat{p}_n$ with $e_n = O_P(r_n^{2/(d+6)})$:

Criterion	Rate	Interpretation
Mode set (Hausdorff)	$O_P \left(r_n^{\min\{\frac{2}{d+6}, \frac{1}{d}\}} \right)$	1NN error dominates for $d \geq 7$
Clustering (1 - ARI)	$O_P \left(r_n^{\min\{\frac{2}{d+6}, \frac{1}{2d}\}} \right)$	1NN error dominates for $d \geq 3$
Dendrogram (d_{GH})	$O_P \left((\log \log n) r_n^{\min\{\frac{2}{d+6}, \frac{1}{2d}\}} \right)$	Same polynomial rate as ARI
Path length	$O_P \left(\left[(\log n)^{\frac{1}{d+2}} r_n^{\frac{1}{d(d+2)}} + r_n^{\frac{2}{d+6}} \right]^{\frac{1}{2}} \right)$	1NN error dominates for $d \geq 2$

Density Waterfall Plot

- $\mathbf{X}^* = \arg \max_{\mathbf{X}_j \in \mathbb{X}_n} \hat{p}(\mathbf{X}_j)$ is the sample global mode.
- $\mathbf{X}_i \equiv \mathbf{X}_{i_1} \rightarrow \mathbf{X}_{i_2} \rightarrow \cdots \rightarrow \mathbf{X}_{i_{T_i}} \equiv \mathbf{X}^*$: unique directed path in the GGDPC graph G .
- $\hat{d}_G(\mathbf{X}_i) = \sum_{\ell=1}^{T_i-1} \|\mathbf{X}_{i_{\ell+1}} - \mathbf{X}_{i_\ell}\|$ is the GGDPC graph distance from $\mathbf{X}_i$ to $\mathbf{X}^*$.

Density Waterfall Plot

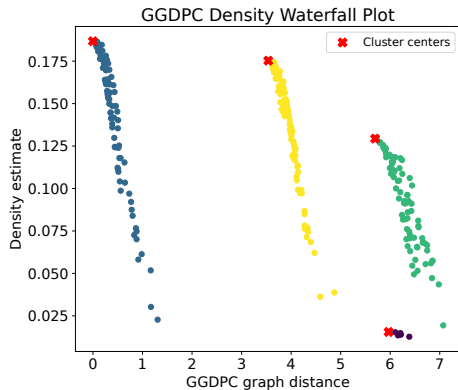
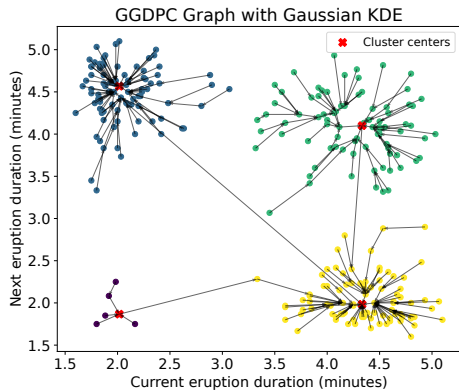
- $X^* = \arg \max_{X_j \in \mathbb{X}_n} \hat{p}(X_j)$ is the sample global mode.
- $X_i \equiv X_{i_1} \rightarrow X_{i_2} \rightarrow \cdots \rightarrow X_{i_{T_i}} \equiv X^*$: unique directed path in the GGDPC graph G .
- $\hat{d}_G(X_i) = \sum_{\ell=1}^{T_i-1} \|X_{i_{\ell+1}} - X_{i_\ell}\|$ is the GGDPC graph distance from X_i to X^* .

Density Waterfall Plot: $\left\{ \left(\hat{d}_G(X_i), \hat{p}(X_i) \right) : i = 1, \dots, n \right\}$.

Density Waterfall Plot

- $X^* = \arg \max_{X_j \in \mathbb{X}_n} \hat{p}(X_j)$ is the sample global mode.
- $X_i \equiv X_{i_1} \rightarrow X_{i_2} \rightarrow \dots \rightarrow X_{i_{T_i}} \equiv X^*$: unique directed path in the GGDPC graph G .
- $\hat{d}_G(X_i) = \sum_{\ell=1}^{T_i-1} \|X_{i_{\ell+1}} - X_{i_\ell}\|$ is the GGDPC graph distance from X_i to X^* .

Density Waterfall Plot: $\left\{ \left(\hat{d}_G(X_i), \hat{p}(X_i) \right) : i = 1, \dots, n \right\}$.

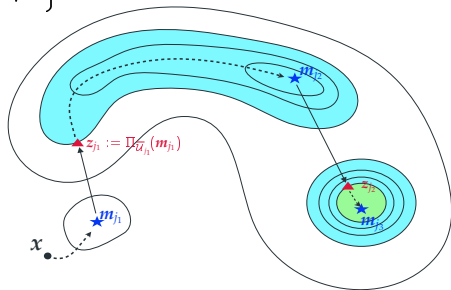


Wasserstein-1 Convergence of the Density Waterfall

Density Waterfall Plot: $\left\{ \left(\widehat{d}_G(\mathbf{X}_i), \widehat{p}(\mathbf{X}_i) \right) : i = 1, \dots, n \right\}$.

- Empirical density waterfall measure:
 $\forall A \in \text{Borel}(\mathbb{R}^2),$

$$\widehat{Q}_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{1} \left\{ \left(\widehat{d}_G(\mathbf{X}_i), \widehat{p}(\mathbf{X}_i) \right) \in A \right\}.$$



Density Waterfall Plot: $\left\{ \left(\widehat{d}_G(\mathbf{X}_i), \widehat{p}(\mathbf{X}_i) \right) : i = 1, \dots, n \right\}.$

- Empirical density waterfall measure:

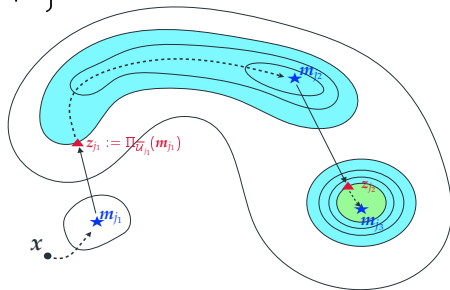
$$\forall A \in \text{Borel}(\mathbb{R}^2),$$

$$\widehat{Q}_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{1} \left\{ \left(\widehat{d}_G(\mathbf{X}_i), \widehat{p}(\mathbf{X}_i) \right) \in A \right\}.$$

- Population GGDPC graph distance:

$$\bar{d}_G(x) = L(x) + \sum_{\ell=1}^{\bar{T}_x-1} \left[\left\| \mathbf{m}_{j_\ell} - \Pi_{\bar{\mathcal{U}}_{j_\ell}}(\mathbf{m}_{j_\ell}) \right\| + L \left(\Pi_{\bar{\mathcal{U}}_{j_\ell}}(\mathbf{m}_{j_\ell}) \right) \right].$$

- Population density waterfall measure: $Q(A) = \mathbb{P} \left(\left(\bar{d}_G(\mathbf{X}), p(\mathbf{X}) \right) \in A \right).$



Wasserstein-1 Convergence of the Density Waterfall

Density Waterfall Plot: $\left\{ \left(\widehat{d}_G(\mathbf{X}_i), \widehat{p}(\mathbf{X}_i) \right) : i = 1, \dots, n \right\}$.

- Empirical density waterfall measure:
 $\forall A \in \text{Borel}(\mathbb{R}^2),$

$$\widehat{Q}_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{1} \left\{ \left(\widehat{d}_G(\mathbf{X}_i), \widehat{p}(\mathbf{X}_i) \right) \in A \right\}.$$

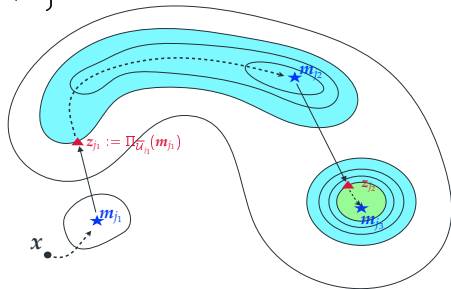
- Population GGDPC graph distance:

$$\bar{d}_G(x) = L(x) + \sum_{\ell=1}^{\bar{T}_x-1} \left[\left\| \mathbf{m}_{j_\ell} - \Pi_{\bar{\mathcal{U}}_{j_\ell}}(\mathbf{m}_{j_\ell}) \right\| + L \left(\Pi_{\bar{\mathcal{U}}_{j_\ell}}(\mathbf{m}_{j_\ell}) \right) \right].$$

- Population density waterfall measure: $Q(A) = \mathbb{P} \left(\left(\bar{d}_G(\mathbf{X}), p(\mathbf{X}) \right) \in A \right).$

$$\text{Wass}_1(\widehat{Q}_n, Q) = O_P \left(\text{"Path-Length Consistency Rate"} + \delta_n + \|\widehat{p} - p\|_\infty + \frac{\log n}{\sqrt{n}} \right).$$

δ_n : observations fell into the boundary tube; $\|\widehat{p} - p\|_\infty$: density estimation error; $\log n / \sqrt{n}$: Wasserstein-1 rate.



Empirical Studies and Conclusion



Comparisons With Other DPC-Type Methods on a Two-Gaussian Mixture

Sample size	500	1000	2000	5000	8000
Adjusted Rand index					
DPC (Rodriguez and Laio, 2014)	.898	.920	.935	.946	.953
DPC-KNN-PCA (Du et al., 2016)	.935	.939	.943	.945	.948
SNN-DPC (Liu et al., 2018)	.881	.871	.872	.874	.878
DPC-CE (Guo et al., 2022)	.840	.886	.911	.928	.941
DPC-DLP (Seyedi et al., 2019)	.341	.494	.632	.772	.808
DPC-MDNN (Wang et al., 2025)	.211	.077	.010	.000	.000
GGDPC (Ours)	.916	.938	.953	.963	.967
Running time (seconds)					
DPC (Rodriguez and Laio, 2014)	.039	.100	.347	1.858	4.627
DPC-KNN-PCA (Du et al., 2016)	.020	.075	.313	2.239	6.096
SNN-DPC (Liu et al., 2018)	.167	1.111	9.922	159.913	632.394
DPC-CE (Guo et al., 2022)	.728	1.802	4.896	20.686	45.763
DPC-DLP (Seyedi et al., 2019)	.074	.337	1.634	14.996	41.864
DPC-MDNN (Wang et al., 2025)	.262	1.128	5.770	56.383	180.123
GGDPC (Ours)	.028	.067	.203	1.061	2.601

Average over 1,000 replications. GGDPC times exclude the optional dendrogram.

Main Takeaways of Today's Talk

Method Propose a gradient-guided improvement of DPC for better stability.

Main Takeaways of Today's Talk

Method Propose a gradient-guided improvement of DPC for better stability.

Theory Five criteria for its consistency:

Local Mode Recovery	Adjusted Rand Index	Dendrogram	Path Length	Waterfall Measure
Local	Geometric concepts			Global
				
Fast	Convergence Rate			Slow

Main Takeaways of Today's Talk

Method Propose a gradient-guided improvement of DPC for better stability.

Theory Five criteria for its consistency:

Local Mode Recovery | Adjusted Rand Index | Dendrogram | Path Length | Waterfall Measure

Local

Geometric concepts

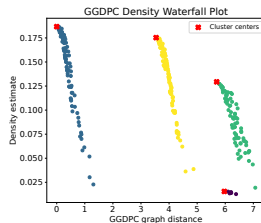
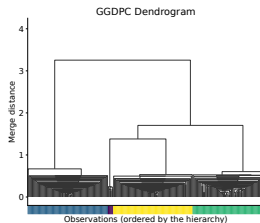
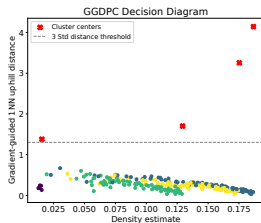
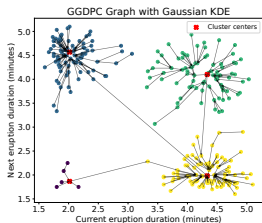
Global

Fast

Convergence Rate

Slow

Infographic GGDPC graph, decision diagram, modal distance dendrogram, density waterfall.



- 1 **Improved convergence rate** via direct mapping within the modal region $B(\hat{\mathbf{m}}_j, \epsilon_n)$,

$$\text{"path-length consistency"} = O_P \left(\left\{ \left[\frac{(\log n)^{d+1}}{n} \right]^{\frac{1}{d(d+2)}} + \left(\frac{\log n}{n} \right)^{\frac{2}{d+6}} \right\}^{\frac{1}{2}} \right) \Rightarrow O_P \left(\left(\frac{\log n}{n} \right)^{\frac{1}{3d}} \log n \right).$$

- 2 **Local signatures via density waterfall plots** for mean shift or k -means clustering.

- 1 **Improved convergence rate** via direct mapping within the modal region $B(\hat{\mathbf{m}}_j, \epsilon_n)$,

$$\text{"path-length consistency"} = O_P \left(\left\{ \left[\frac{(\log n)^{d+1}}{n} \right]^{\frac{1}{d(d+2)}} + \left(\frac{\log n}{n} \right)^{\frac{2}{d+6}} \right\}^{\frac{1}{2}} \right) \Rightarrow O_P \left(\left(\frac{\log n}{n} \right)^{\frac{1}{3d}} \log n \right).$$

- 2 **Local signatures via density waterfall plots** for mean shift or k -means clustering.

Thank you!



Paper: <https://arxiv.org/abs/2610.01050>.

References

- E. Arias-Castro and W. Qiao. Clustering by hill-climbing: Consistency results. *The Annals of Statistics*, 53(6): 2536–2562, 2025.
- E. Arias-Castro, W. Qiao, and L. Zheng. Estimation of the global mode of a density: Maximality, adaptation, and computational complexity. *Electronic Journal of Statistics*, 16(1):2774–2795, 2022.
- A. Azzalini and A. W. Bowman. A look at some data on the old faithful geyser. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 39(3):357–365, 1990.
- G. E. Carlsson and F. Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11(47):1425–1470, 2010.
- D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.
- M. Du, S. Ding, and H. Jia. Study on density peaks clustering based on k-nearest neighbors and principal component analysis. *Knowledge-Based Systems*, 99:135–145, 2016.
- L. Fu and E. Medico. Flame, a novel fuzzy clustering method for the analysis of dna microarray data. *BMC Bioinformatics*, 8(1):3, 2007.
- K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, 1975.
- W. Guo, W. Wang, S. Zhao, Y. Niu, Z. Zhang, and X. Liu. Density peak clustering with connectivity estimation. *Knowledge-Based Systems*, 243:108501, 2022.
- L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985.

- H. Jiang. On the consistency of quick shift. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- R. Liu, H. Wang, and X. Yu. Shared-nearest-neighbor-based clustering by fast search and find of density peaks. *information sciences*, 450:200–226, 2018.
- S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.
- J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297. University of California Press, 1967.
- A. Rodriguez and A. Laio. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496, 2014.
- J. P. Romano. On weak convergence and optimality of kernel density estimates of the mode. *The Annals of Statistics*, 16(2):629 – 647, 1988.
- S. A. Seyed, A. Lotfi, P. Moradi, and N. N. Qader. Dynamic graph-based label propagation for density peaks clustering. *Expert Systems with Applications*, 115:314–328, 2019.
- A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In *European Conference on Computer Vision*, pages 705–718, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.
- I. Verdinelli and L. Wasserman. Analysis of a mode clustering diagram. *Electronic Journal of Statistics*, 12(2):4288 – 4312, 2018.
- U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

References

- H. Wang, J. Zhang, Y. Shen, S. Wang, B. Deng, and W. Zhao. Improved density peak clustering with a flexible manifold distance and natural nearest neighbors for network intrusion detection. *Scientific Reports*, 15(1):8510, 2025.
- J. Xie, H. Gao, W. Xie, X. Liu, and P. W. Grant. Robust clustering by detecting density peaks and assigning points based on fuzzy weighted k-nearest neighbors. *Information Sciences*, 354:19–40, 2016.
- Y. Zhang and Y.-C. Chen. Gradient-guided density peak clustering. *arXiv preprint arXiv:2610.01050*, 2026.

Precise Regularity Conditions

- A1: Density.** (a) p has a compact support $\mathcal{C}$ with $\inf_{x \in \mathcal{C}} p(x) \geq p_{\min} > 0$;
(b) $\exists C_{\mathcal{C}}, \epsilon_0 > 0$ s.t. $\text{Leb}(\mathcal{C} \cap B(x, \epsilon)) \geq C_{\mathcal{C}} \cdot \text{Leb}(B(x, \epsilon))$, $\forall x \in \mathcal{C}$ and $0 < \epsilon \leq \epsilon_0$;
(c) p has a C^3 extension to a neighborhood with bounded derivatives;
(d) Its critical points are finite and non-degenerate; distinct modal heights;
(e) $\mathcal{C}$ is forward-invariant for $\gamma'_x(t) = \nabla p(\gamma_x(t))$.

A2: Estimation. On a fixed neighborhood of $\mathcal{C}$, $\hat{p} \in C^2$, $\hat{g} = \nabla \hat{p}$, and derivatives through order two are bounded in probability.

- A3: Separatrix.** (a) Each $\partial \mathcal{C}_a$ is a finite union of stable manifolds of saddles with one positive and $d - 1$ negative Hessian eigenvalues;
(b) It is a C^2 hypersurface away from finitely many critical points. Write $\mathcal{S}_{a,\text{reg}}$ for its noncritical C^2 part. For some $r_{\mathcal{S}}, \rho_{\mathcal{S}} > 0$:

$$x \in \mathcal{C}_a, \quad 0 < d(x, \partial \mathcal{C}_a) < r_{\mathcal{S}}, \quad \Pi_{\partial \mathcal{C}_a}(x) \cap \mathcal{S}_{a,\text{reg}} \neq \emptyset \implies |\Pi_{\partial \mathcal{C}_a}(x)| = 1,$$

$$\nu(s)^T \nabla^2 p(s) \nu(s) \geq \rho_{\mathcal{S}} \quad \text{on } \mathcal{S}_{a,\text{reg}} \text{ outside fixed saddle neighborhoods.}$$

- A4: Projections.** For every non-global mode $\mathbf{m}_j$ ($j \geq 2$), the projection $\mathbf{z}_j = \Pi_{\bar{\mathcal{U}}_j}(\mathbf{m}_j)$ is unique, $\mathbf{z}_j \in \mathcal{C}_{\pi(j)}$, $\nabla p(\mathbf{z}_j) \neq 0$, and $d(\mathbf{z}_j, \partial \mathcal{C}_{\pi(j)}) > 0$. For some $\kappa_j > 0$,

$$\mathbf{v}^T [I_d - \mu_j \nabla^2 p(\mathbf{z}_j)] \mathbf{v} \geq \kappa_j \quad \text{if } \|\mathbf{v}\| = 1, \mathbf{v} \perp \nabla p(\mathbf{z}_j), \quad \mathbf{z}_j - \mathbf{m}_j = \mu_j \nabla p(\mathbf{z}_j), \quad \mu_j > 0.$$

- A5: Saddles.** ∇p is C^1 -linearizable near every boundary saddle.

Stable and unstable manifolds: Let s be a critical point of p , and $\gamma'_x(t) = \nabla p(\gamma_x(t))$:

$$W^s(s) = \{x \in \mathcal{C} : \lim_{t \rightarrow +\infty} \gamma_x(t) = s\}, \quad W^u(s) = \{x \in \mathcal{C} : \lim_{t \rightarrow -\infty} \gamma_x(t) = s\}.$$

- By the stable manifold theorem, $W^s(s)$, $W^u(s)$ are locally C^2 manifolds in $\mathbb{R}^d$ near s .

C^1 -linearization: There exists a local C^1 -diffeomorphism Φ with $\Phi(0) = 0$ satisfying

$$\Phi(\gamma_x(t)) = e^{tA}\Phi(x), \quad A = \nabla^2 p(s),$$

while trajectories stay in the local charts. In these coordinates, $y'(t) = Ay(t)$.

Sternberg non-resonance condition:

For eigenvalues $\rho_1, \dots, \rho_d$ of $\nabla^2 p(s)$,
require at every order $N \geq 2$,

$$\rho_i - \sum_{j=1}^d k_j \rho_j \neq 0, \quad i = 1, \dots, d,$$

$$k_j \in \mathbb{N}, \quad 2 \leq \sum_{j=1}^d k_j \leq N.$$

Hartman spectral condition:

The eigenvalues of $\nabla^2 p(s)$ lie in

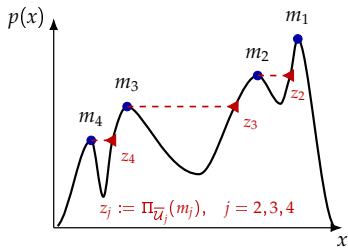
$$[-\alpha_L, -\alpha_R], \quad [\beta_L, \beta_R], \quad \text{with } \alpha_R, \alpha_L, \beta_L, \beta_R > 0.$$

Require $\alpha_L - \alpha_R < \mu\beta_L$ and $\beta_R - \beta_L < \nu\alpha_R$,

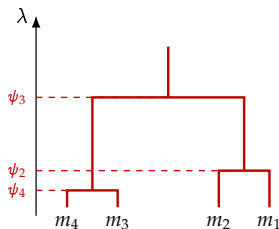
- $\mu, \nu > 0$ are the Hölder spectral exponents of the stable/unstable linearized flow.

Distinctions Between Modal Distance Dendrogram and Density Cluster Tree

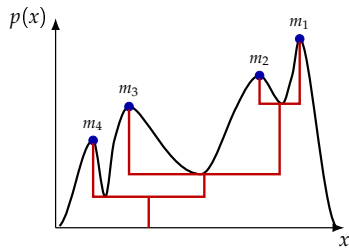
The modal distance dendrogram $\mathcal{T}_{\text{mode}}$ and the density cluster tree $\mathcal{T}_p$ can have different merge topologies.



(a) Density p and modal projections



(b) Modal-distance dendrogram $\mathcal{T}_{\text{mode}}$



(c) Density cluster tree $\mathcal{T}_p$

- Modal distance tree: $w(\mathbf{m}_4) < w(\mathbf{m}_2) \ll w(\mathbf{m}_3)$, so close pairs merge first.
- Density tree: merges follow the ordered density levels.

Gromov-Hausdorff Distance

An *ultrametric* on a set $\mathbb{X}$ is a function $u : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_+$ satisfying, for any $x, x', x'' \in \mathbb{X}$,

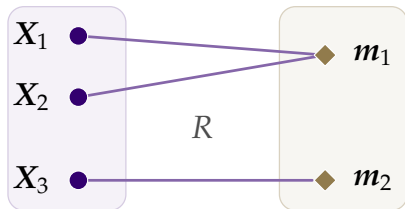
(a) $u(x, x') = 0 \Leftrightarrow x = x'$; (b) $u(x, x') = u(x', x)$; (c) $u(x, x') \leq \max \{u(x, x''), u(x'', x')\}$.

- $u_n(\mathbf{X}_i, \mathbf{X}_j) = \max \left\{ \widehat{w}_n(\mathbf{X}_k) : (\mathbf{X}_k \rightarrow \widehat{\Phi}_n(\mathbf{X}_k)) \in \text{Path}_G(\mathbf{X}_i, \mathbf{X}_j) \right\}$, where $\text{Path}_G(\mathbf{X}_i, \mathbf{X}_j)$ is the edge set in G joining $\mathbf{X}_i$ and $\mathbf{X}_j$.
- $u_{\mathcal{M}}(\mathbf{m}_j, \mathbf{m}_k) = \max \left\{ w(\mathbf{m}_\ell) : (\mathbf{m}_\ell \rightarrow \mathbf{m}_{\pi(\ell)}) \in \text{Path}_{G_{\mathcal{M}}}(\mathbf{m}_j, \mathbf{m}_k) \right\}$.

A correspondence $R \subseteq \mathbb{X}_n \times \mathcal{M}$ satisfies both coordinate projections are surjective.

Sample set $\mathbb{X}_n$

Mode set $\mathcal{M}$



The **Gromov-Hausdorff distance** between $(\mathbb{X}_n, u_n)$ and $(\mathcal{M}, u_{\mathcal{M}})$ is defined as

$$d_{\text{GH}}((\mathbb{X}_n, u_n), (\mathcal{M}, u_{\mathcal{M}})) = \frac{1}{2} \inf_R \sup_{(x, m), (y, m') \in R} |u_n(x, y) - u_{\mathcal{M}}(m, m')|.$$

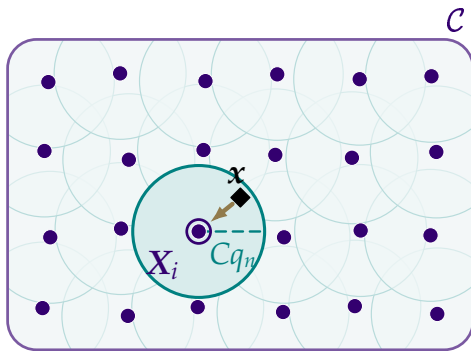
Uniform Sample Coverage Lemma

Lemma (Lemma C.1 in [Zhang and Chen 2026](#))

Suppose that Assumption A1(a) holds, and the density p is bounded on its support $\mathcal{C}$. Then, almost surely,

$$\sup_{x \in \mathcal{C}} \min_{i=1, \dots, n} \|x - X_i\| = O\left(\left(\frac{\log n}{n}\right)^{\frac{1}{d}}\right).$$

The same bound holds over any closed subset of $\mathcal{C}$.



Purple dots are samples; the gold arrow joins x to its nearest sample.

As $n \rightarrow \infty$, almost surely, sample neighborhoods cover the support: $\mathcal{C} \subseteq \bigcup_{i=1}^n B(X_i, Cq_n)$.