Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Persistence of generalized density functions

Master Thesis

Maxim Mikhaylov

June 30, 2025

Supervisors: Prof. Dr. B. Gärtner, Dr. P. Schnider
Department of Computer Science, ETH Zürich

Abstract

Topological data analysis (TDA) often uses density-like functions defined on the ambient space $\mathbb{R}^d$ to infer the underlying topological structure of a dataset $P \subseteq \mathbb{R}^d$. The persistent homology of the sublevelset or superlevelset filtrations induced by these functions captures multi-scale topological features. A key desirable property is *stability*: small perturbations in the dataset result in similarly small changes in the persistence diagrams. A classic example is the nearest-neighbor distance function $f(P, x) := \min_{p \in P} d(x, p)$, whose sublevel sets are homotopy equivalent to the Čech complex [17], for which the bottleneck distance between persistence diagrams $\mathrm{Dgm}(f_P)$ and $\mathrm{Dgm}(f_Q)$ is bounded by the Hausdorff distance $d_H(P, Q)$ between datasets [7]:

$$d_b(\mathrm{Dgm}(f_P), \mathrm{Dgm}(f_Q)) \leq d_H(P, Q).$$

This thesis investigates *generalized density functions*, which are functions $f : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$, where $\mathcal{P}(\mathbb{R}^d)$ denotes the set of all finite subsets of $\mathbb{R}^d$, and the conditions under which they satisfy a stability property of the form

$$d_b(\mathrm{Dgm}(f_P), \mathrm{Dgm}(f_Q)) \leq c \cdot d_H(P, Q)$$

for some finite constant c.

The primary contributions of this work are stability theorems for several classes of generalized density functions. Specifically, we prove stability bounds for:

- Several generalizations of the nearest-neighbor distance function of the form $f(P, x) = \min_{p \in P} h(x, p)$.
- Functions that are Lipschitz continuous with respect to the Hausdorff distance on the space of point clouds.
- Morse functions that satisfy a Lipschitz-like condition.

Beyond these core stability results, we explore the properties of the space of stable functions and investigate how common operations, such as addition and taking minima, affect stability. We identify conditions under which stability is preserved under these operations and provide counterexamples demonstrating cases where it is not.

Our findings unify and extend existing stability results, offering practical guidance for the selection and design of generalized density functions for topological data analysis.

Contents

Chapter 1

Introduction

Topological Data Analysis (TDA) provides a robust framework for understanding the shape of data [5]. Among its tools is *persistent homology*, which extracts topological features across different scales [2]. Traditionally, persistent homology is applied to a finite point cloud $P \subseteq \mathbb{R}^d$ through filtrations of simplicial complexes built on top of P, such as the Vietoris–Rips or Čech complexes [8]. An alternative (though sometimes equivalent) approach involves defining a real-valued function $\mathbb{R}^d \to \mathbb{R}$ that encodes the geometry of the data, and then computing the persistent homology of its sublevel (or superlevel) set filtration [10]. The resulting persistent homology captures topological features such as connected components, loops, and voids, and encodes them in a persistence diagram, which summarizes their birth and death across the filtration [8].

A classical and widely studied example is the *nearest-neighbor distance function*,

$$f(P, x) = \min_{p \in P} d(x, p), \tag{1.1}$$

where d is a metric on $\mathbb{R}^d$. The sublevel sets of this function are homotopy equivalent to the Čech filtration [17], a fundamental construction in TDA. This function satisfies a *stability* property [7]:

$$d_b(\mathrm{Dgm}(f_P), \mathrm{Dgm}(f_Q)) \leq c \cdot d_H(P, Q), \tag{1.2}$$

where d_b is the bottleneck distance between persistence diagrams and d_H is the Hausdorff distance between point clouds. Stability is not merely a theoretical nicety; it is the foundation that provides reliability of TDA [8]. In real-world applications, where data is inevitably noisy or subject to measurement error, an unstable method could produce drastically different topological summaries from minor, inconsequential variations in the input. A stable GDF ensures that the extracted topological features are genuine reflections of the data's underlying structure, rather than artifacts of noise or sampling.

While the nearest-neighbor function is stable, it is just one of many possible density-like functions for TDA [1, 12, 16]. In practice, alternative functions may offer computational advantages [11, 4], better capture intrinsic structure [1], or incorporate domain knowledge [13]. For example, the DTM filtration [1] modifies the nearest-neighbor function to make it more robust to noise and outliers. This motivates the study of *generalized density functions* (GDFs), which are functions of the form

$$f(P,x) : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}. \tag{1.3}$$

For a given point cloud P, such a function defines a real-valued function $f_P : \mathbb{R}^d \to \mathbb{R}$, whose sublevel set filtration can be used to analyze the topological features of P. We say f is *c-stable* if for all finite point clouds $P, Q \subseteq \mathbb{R}$ we have

$$d_b(\mathrm{Dgm}(f_P), \mathrm{Dgm}(f_Q)) \leq c \cdot d_H(P,Q), \tag{1.4}$$

where the *stability constant* c measures the sensitivity of the persistence diagrams to perturbations in the data.

We limit ourselves to the case where P and Q are finite sets, as known stability results for GDFs often hold only in this case, because otherwise the functions f_P often are not tame, and, consequently, the persistence diagrams $\mathrm{Dgm}(f_P)$ are not q-tame.

We also only consider the space $\mathbb{R}^d$ instead of a general topological space for the sake of simplicity, as the various theorems used in this thesis impose different conditions on the underlying topological space. The euclidean space $\mathbb{R}^d$ was chosen as the lowest common denominator, although many results in this thesis hold for more general spaces. An inquisitive reader is invited to track down the most general conditions for the various novel theorems by keeping track of the theorems used in the proofs.

In this thesis, we primarily investigate which GDFs are stable. Existing stability results, such as those for the Čech filtration and its weighted variants, provide a starting point. This thesis aims to create a more comprehensive framework by identifying broader conditions that ensure stability, thereby unifying and extending these results. Concretely, we prove stability of the following classes of functions:

1. We begin by examining natural extensions of the nearest-neighbor function, specifically of the form $f(P,x) = \min_{p \in P} h(d(p,x))$, where h is a monotone Lipschitz continuous function. This function allows the balls of the Čech complex to grow at different rates, although the speed is still the same for all balls at any given r.

2. This is further extended to $f(P, x) = \min_{p \in P} h(p, x)$, where $h(p, x)$ is a Lipschitz function with respect to p. This allows for more general shapes, as well as different growth rates for different points p.

3. We then investigate a more direct condition: functions $f(P, x)$ that are Lipschitz with respect to point clouds *(PC-Lipschitz)*, where $\mathcal{P}(\mathbb{R}^d)$ is equipped with the Hausdorff distance, and the space of real-valued functions $f_P : \mathbb{R}^d \to \mathbb{R}$ is equipped with the L_∞-norm. This class includes the previous two classes as special cases, and allows for more general functions that are still stable, but may not be represented as a minimum over a set of functions.

4. Finally, we consider Morse functions whose restriction to the critical points is Lipschitz continuous, which includes some functions that are not Lipschitz continuous in the usual sense, but are still stable.

Beyond these classes, we also study when stability is preserved under common operations. For example, if two functions are each stable, under what conditions is their pointwise minimum or average also stable? We identify sufficient conditions and provide concrete counterexamples demonstrating cases where stability fails. This analysis shows how more complex stable functions can be constructed from simpler ones, and provides insight into the structure of the space of stable functions.

Finally, we examine the topological and algebraic properties of the space of c-stable functions itself. We analyze properties such as convexity, contractibility, and closedness, as well as establish a lattice structure on the set of c-PC-Lipschitz functions.

Taken together, our results contribute to the theoretical foundations of TDA by deepening the understanding of stability for generalized density functions. They also provide practical guidance for designing stable filtrations.

TODO: this is too similar to the abstract — add more context?

Chapter 2

Background

This chapter reviews the mathematical foundations for stability of generalized density functions. We briefly review the core concepts of how real-valued functions induce filtrations for persistent homology and then focus on established stability theorems that serve as a foundation and motivation for this work. We assume the reader has a working knowledge of topological data analysis, mainly persistent homology. Otherwise, a comprehensive introduction to the subject is provided in the textbook by Herbert Edelsbrunner and John Harer [10].

At the heart of this thesis lies the concept of a generalized density function (GDF).

Definition 2.1 (Generalized density function) *A generalized density function is a map $f : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$, where $\mathcal{P}(\mathbb{R}^d)$ denotes the collection of all finite subsets of $\mathbb{R}^d$. For a given point cloud $P \in \mathcal{P}(\mathbb{R}^d)$, a GDF f induces a real-valued function $f_P : \mathbb{R}^d \to \mathbb{R}$ defined by $f_P(x) := f(P, x)$.*

These functions f_P are used to construct *sublevel set filtrations*.

Definition 2.2 (Sublevel set filtration) *Given a function $f_P : \mathbb{R}^d \to \mathbb{R}$, its sublevel set at a value $a \in \mathbb{R}$ is*

$$f_P^{-1}(-\infty, a] = \{x \in \mathbb{R}^d \mid f_P(x) \leq a\}. \tag{2.1}$$

The sublevel set filtration of g is the nested family of sets $\{f_P^{-1}(-\infty, a]\}_{a \in \mathbb{R}}$.

Alternatively, superlevel set filtrations can be used, which are defined by $\{f_P^{-1}[a, +\infty)\}_{a \in \mathbb{R}}$. The choice between sublevel and superlevel set filtrations is a matter of convention, as the superlevel set filtration of f_P is exactly the sublevel set filtration of the function $-f_P$, and vice versa. For the sake of consistency, we will use sublevel set filtrations throughout this thesis.

Applying a homology functor H_k to a filtration $\{f_P^{-1}(-\infty, a]\}_{a \in \mathbb{R}}$ yields a persistence module, which tracks the evolution of topological features as

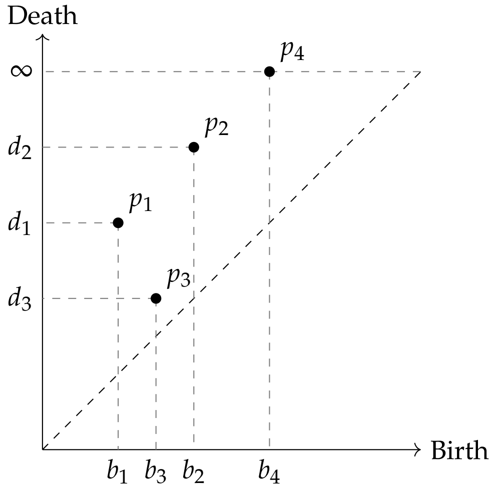

Figure 2.1: An example of a persistence diagram. A point p_i is born at b_i and dies at d_i. The point p_4 does not die, so its death time is ∞. The dashed line indicates the diagonal Δ, where $b_i = d_i$.

the parameter a varies [10]. The *persistence diagram*, denoted $\mathrm{Dgm}_k(f_P)$, is a multiset of points (b_i, d_i) in the extended plane $\overline{\mathbb{R}}^2$, where b_i represents the birth time and d_i the death time of a k-dimensional feature. An example of a persistence diagram is shown in Figure 2.1.

2.1 Quantifying stability

To formalize the notion of stability, we need metrics to compare both point clouds and persistence diagrams. A common metric for point clouds is the *Hausdorff distance*, which is also used in the definition of stability of the nearest-neighbor distance function.

Definition 2.3 (Hausdorff distance) *Given two non-empty sets $P, Q \subseteq \mathbb{R}^d$, where $\mathbb{R}^d$ is a metric space, the Hausdorff distance between P and Q is defined as*

$$d_H(P, Q) = \max \left\{ \sup_{p \in P} \inf_{q \in Q} d(p, q), \sup_{q \in Q} \inf_{p \in P} d(q, p) \right\}. \tag{2.2}$$

Intuitively, the Hausdorff distance measures the "maximum mismatch" between the two sets, capturing the largest distance between a point in one set and its closest neighbor in the other set.

Definition 2.4 (Bottleneck distance) *Given two persistence diagrams D_1 and D_2, the bottleneck distance between them is defined as*

$$d_b(D_1, D_2) = \inf_{\pi} \sup_{(b,d) \in D_1} \|(b, d) - \pi(b, d)\|_\infty, \tag{2.3}$$

where the infimum is taken over all bijections π between D_1 and D_2, allowing for the possibility of matching points to the diagonal $\Delta = \{(x, x) \mid x \in \overline{\mathbb{R}}\}$.

The bottleneck distance measures the cost of the "least expensive" matching between the points of the two diagrams, where the cost is defined as the longest edge in the matching.

With these metrics defined, we can formally define stability for a GDF.

Definition 2.5 *A generalized density function $f : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ is c-*stable *if for all non-empty finite point clouds $P, Q \subseteq \mathbb{R}^d$ and all homology dimensions $k \geq 0$, we have*

$$d_b(\mathrm{Dgm}_k(f_P), \mathrm{Dgm}_k(f_Q)) \leq c \cdot d_H(P, Q). \tag{2.4}$$

If no such finite c exists, we say that f is unstable *or ∞-stable.*

2.2 Known stability results

This section reviews the known stability results for GDFs and simplicial filtrations that can be represented as GDFs.

2.2.1 Čech filtration

As mentioned in Chapter 1, the Čech filtration can be represented as the sublevel set filtration of the nearest-neighbor distance function $f(P, x) := \min_{p \in P} d(x, p)$. The sublevel sets $f_P^{-1}(-\infty, a]$ are precisely $\bigcup_{p \in P} B(p, a)$, the union of closed balls of radius a centered at the points of P. The nerve of this union of balls defines the Čech complex. By the Nerve theorem [3, 14], if P is finite, the Čech complex is homotopy equivalent to the union of balls.

The stability of this construction is a well-known result:

Theorem 2.6 (Stability of the Čech filtration, [7]) *Let $f_P(x) := \min_{p \in P} d(x, p)$. Then for all finite $P, Q \subseteq \mathcal{P}(\mathbb{R}^d)$ and any dimension $k \geq 0$, we have*

$$d_b(\mathrm{Dgm}_k(f_P), \mathrm{Dgm}_k(f_Q)) \leq d_H(P, Q). \tag{2.5}$$

Using the terminology of this thesis, we say the nearest-neighbor distance function is 1-stable.

2.2.2 Weighted Čech filtration

The weighted Čech filtration is a generalization of the Čech filtration used to define the DTM-filtration [1].

Definition 2.7 (Weighted Čech filtration) *Let $q \in [1, \infty)$, $P \subseteq \mathbb{R}^d$ and $g : P \to \mathbb{R}_{\geq 0}$. For every $p \in P, a \in \mathbb{R}^+$, we define the function $r_p(a)$ to be:*

$$r_p^{(q)}(a) = \begin{cases} -\infty, & \text{if } a < g(p), \\ (a^q - g(p)^q)^{1/q}, & \text{otherwise.} \end{cases} \tag{2.6}$$

For $q = \infty$, we also define

$$r_p^{(q)}(a) = \begin{cases} -\infty, & \text{if } a < g(p), \\ a, & \text{otherwise.} \end{cases} \tag{2.7}$$

$g(p)$ acts as a weight for each point $p \in P$, influencing the radius $r_p^{(q)}(a)$ of the ball $B(p, r_p^{(q)}(a))$. Concretely, the weighted Čech filtration is defined as

$$V_q^a[P, g] = \bigcup_{p \in P} B(p, r_p^{(q)}(a)), \tag{2.8}$$

where the balls are closed, and the ball of radius $-\infty$ is the empty set.

The weighted Čech filtration is a strict generalization of the Čech filtration, as the Čech filtration is the special case where $g(p) = 0$ for all $p \in P$.

Similarly to the Čech filtration, this filtration can be represented as a GDF of the form

$$f_{g,q}(P, x) = \min_{p \in P} (d(x, p)^q + g(p)^q)^{1/q}, \tag{2.9}$$

where similarly to the L_∞-norm, when $q = \infty$, the term $(d(x, p)^q + g(p)^q)^{1/q}$ is replaced by $\max(d(x, p), g(p))$. A proof of the equivalence of this GDF and the weighted Čech filtration is given in Appendix A.

If g is Lipschitz continuous, then the weighted Čech filtration is guaranteed to be stable:

Theorem 2.8 (Stability of the weighted Čech filtration, [1]) *Let $P, Q \subseteq \mathbb{R}^d$ be compact and $g : P \cup Q \to \mathbb{R}^+$ be a Lipschitz continuous function with Lipschitz constant c. Then for any dimension $k \geq 0$, we have*

$$d_b(\mathrm{Dgm}_k(V_q^a[P, g]), \mathrm{Dgm}_k(V_q^a[Q, g])) \leq (1 + c^q)^{1/q} \cdot d_H(P, Q), \tag{2.10}$$

This theorem immediately implies that the corresponding GDF is $(1 + c^q)^{1/q}$-stable.

2.2.3 Stability of L_∞-norm-bounded functions

This fundamental result relates the bottleneck distance between persistence diagrams of two functions to the L_∞-distance between the functions themselves.

Theorem 2.9 (Stability of persistence diagrams, [9]) *Let X be a triangulable space, and $g, h : X \to \mathbb{R}$ be two continuous tame functions. Then for all $k \geq 0$, the persistence diagrams of g and h satisfy*

$$d_b(\mathrm{Dgm}_k(g), \mathrm{Dgm}_k(h)) \leq \|g - h\|_\infty. \tag{2.11}$$

Importantly, this theorem is applicable to the case $X = \mathbb{R}^d$, as the space $\mathbb{R}^d$ is triangulable [15]. We can easily adapt this theorem to the case of GDFs:

$$d_b(\mathrm{Dgm}_k(f_P), \mathrm{Dgm}_k(f_Q)) \leq \|f_P - f_Q\|_\infty. \tag{2.12}$$

This theorem directly implies that if a GDF f satisfies $\|f_P - f_Q\|_\infty \leq c \cdot d_H(P, Q)$ for some constant c, then f is c-stable. This provides a powerful tool for proving stability of GDFs. As an example, we prove 1-stability of the Čech filtration using this theorem.

Proof Let $P, Q \subseteq \mathbb{R}^d$ be two finite point clouds with Hausdorff distance $d = d_H(P, Q)$. Let x be an arbitrary point in $\mathbb{R}^d$ with the nearest point in P denoted by p. This implies that $f_{\text{dist},P}(x) = d(x, p)$. By the definition of the Hausdorff distance, there exists a point $q' \in Q$ such that $d(p, q') \leq d$. By the triangle inequality, we have

$$d(x, q') \leq d(x, p) + d(p, q') \leq d(x, p) + d. \tag{2.13}$$

and thus:

$$f_{\text{dist},Q}(x) = \min_{q \in Q} d(x, q) \leq d(x, q') \leq d(x, p) + d = f_{\text{dist},P}(x) + d. \tag{2.14}$$

Symmetrically, we can show that $f_{\text{dist},P}(x) \leq f_{\text{dist},Q}(x) + d$. Thus, we have

$$\|f_{\text{dist},P} - f_{\text{dist},Q}\|_\infty \leq d, \tag{2.15}$$

which implies that f_{dist} is 1-stable by Theorem 2.9. $\square$

2.2.4 Stability of functions with interleaved persistence modules

An alternative approach to show stability is to use the *interleaving distance*, which bounds the bottleneck distance not only from above, but also from below.

Definition 2.10 (Interleaving of persistence modules) *Two persistence modules $\mathbb{M}$ and $\mathbb{N}$ with M_a, N_a as their respective groups and $m_a^{a'}, n_a^{a'}$ as their homomorphisms are* ε-interleaved *for $\varepsilon \geq 0$ if there exist families of maps $\varphi_a : M_a \to N_{a+\varepsilon}$ and $\psi_a : N_a \to N_{a+\varepsilon}$ such that the following diagrams commute for all $a < a'$:*

$$\begin{tikzcd} M_a \arrow[rr, "m_a^{a'}"] \arrow[dr, "\varphi_a"'] & & M_{a'} \arrow[dr, "\varphi_{a'}"] & \\ & N_{a+\varepsilon} \arrow[rr, "n_{a+\varepsilon}^{a'+\varepsilon}"'] & & N_{a'+\varepsilon} \end{tikzcd} \qquad \begin{tikzcd} M_a \arrow[rr, "m_a^{a+2\varepsilon}"] \arrow[dr, "\varphi_a"'] & & M_{a+2\varepsilon} \\ & N_{a+\varepsilon} \arrow[ur, "\psi_{a+\varepsilon}"'] & \end{tikzcd}$$

$$\begin{tikzcd}
 & M_{a+\varepsilon} \arrow[rr, "n_{a+\varepsilon}^{a'+\varepsilon}"] & & M_{a'+\varepsilon} \\
N_a \arrow[ur, "\psi_a"] \arrow[rr, "n_a^{a'}"'] & & N_{a'} \arrow[ur, "\psi_{a'}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & M_{a+\varepsilon} \arrow[dr, "\varphi_{a+\varepsilon}"] & \\
N_a \arrow[ur, "\psi_a"] \arrow[rr, "n_a^{a+2\varepsilon}"'] & & N_{a+2\varepsilon}
\end{tikzcd}$$

Note that $\mathbb{M}$ and $\mathbb{N}$ are 0-interleaved if and only if they are isomorphic. The triangular diagrams on the right collapse to diagrams that show that φ_t and ψ_t are inverses of each other, while the trapezoidal diagrams on the left ensure they commute with the linear maps m and n.

Definition 2.11 (Interleaving distance) *The interleaving distance between two persistence modules $\mathbb{M}$ and $\mathbb{N}$ is given by:*

$$d_i(\mathbb{M}, \mathbb{N}) = \inf\{\varepsilon \mid \mathbb{M} \textit{ and } \mathbb{N} \textit{ are } \varepsilon\textit{-interleaved}\}. \tag{2.16}$$

Finally, the interleaving distance and the bottleneck distance are equal:

Theorem 2.12 (Isometry theorem, [6]) *Let $\mathbb{M}, \mathbb{N}$ be q-tame persistence modules. Then*

$$d_i(\mathbb{M}, \mathbb{N}) = d_b(\mathrm{Dgm}(\mathbb{M}), \mathrm{Dgm}(\mathbb{N})) \tag{2.17}$$

Additionally, we can define interleaving for filtrations:

Definition 2.13 (Interleaving of filtrations) *Let $\mathcal{F}, \mathcal{G}$ be filtrations over $\mathbb{R}$, that is, $\mathbb{R}$-indexed sequence of nested subspaces of some topological space X. $\mathcal{F}$ and $\mathcal{G}$ are* ε-interleaved *if there exist families of maps $\varphi_a : F_a \to G_{a+\varepsilon}$ and $\psi_a : G_a \to F_{a+\varepsilon}$ such that the same diagrams as in Definition 2.10 commute up to homotopy.*

The interleaving distance between two filtrations is defined the same way as for persistence modules. The following lemma shows that the former bounds the latter from above:

Lemma 2.14 ([17]) *For any two filtrations $\mathcal{F}$ and $\mathcal{G}$ over $\mathbb{R}$, we have*

$$d_i(H_p\mathcal{F}, H_p\mathcal{G}) \leq d_i(\mathcal{F}, \mathcal{G}). \tag{2.18}$$

With these results in hand, we can apply the interleaving approach to once again prove 1-stability of the Čech filtration.

Proof ([17]) Let X be a topological space, and $P, Q \subseteq X$ be two finite point clouds with Hausdorff distance $d = d_H(P, Q)$. Recall that the Čech complex $\check{C}^r(P)$ is homotopy equivalent to the union of the balls $\bigcup_{p \in P} B(p, r)$. Let $x \in B(p, r)$. By the definition of the Hausdorff distance, there exists a point $q \in Q$ such that $d(p, q) \leq d$. Therefore, by the triangle inequality we have

$$d(x, q) \leq d(x, p) + d(p, q) \leq r + d, \tag{2.19}$$

which implies that $x \in B(q, r+d)$. This provides an inclusion

$$\bigcup_{p \in P} B(p,r) \subseteq \bigcup_{q \in Q} B(q, r+d). \tag{2.20}$$

With these two facts, we see that the following diagram commutes up to homotopy, as all maps are either inclusions or homotopies:

$$\begin{tikzcd}
\check{C}^r(P) \arrow[r, hook] \arrow[d, leftrightarrow, "\simeq"] & \check{C}^{r+d}(P) \arrow[r, hook] \arrow[d, leftrightarrow, "\simeq"] & \check{C}^{r+2d}(P) \arrow[d, leftrightarrow, "\simeq"] \\
\bigcup_{p\in P} B(p,r) \arrow[r, hook] \arrow[dr, hook] & \bigcup_{p\in P} B(p,r+d) \arrow[r, hook] \arrow[dr, hook] & \bigcup_{p\in P} B(p,r+2d) \\
\bigcup_{q\in Q} B(q,r) \arrow[r, hook] \arrow[ur, hook] & \bigcup_{q\in Q} B(q,r+d) \arrow[r, hook] \arrow[ur, hook] & \bigcup_{q\in Q} B(q,r+2d) \\
\check{C}^r(Q) \arrow[r, hook] \arrow[u, leftrightarrow, "\simeq"] & \check{C}^{r+d}(Q) \arrow[r, hook] \arrow[u, leftrightarrow, "\simeq"] & \check{C}^{r+2d}(Q) \arrow[u, leftrightarrow, "\simeq"]
\end{tikzcd} \tag{2.21}$$

All diagrams that are needed to show that $\check{C}^r$ and $\check{C}^r(Q)$ are d-interleaved are included in this diagram, and thus we have

$$d_i(\check{C}^r(P), \check{C}^r(Q)) \leq d. \tag{2.22}$$

By Lemma 2.14 and the isometry theorem, we have

$$d_b(\mathrm{Dgm}_k(\check{C}^r(P)), \mathrm{Dgm}_k(\check{C}^r(Q))) \leq d_i(H_k(\check{C}^r(P)), H_k(\check{C}^r(Q))) \tag{2.23}$$

$$\leq d_i(\check{C}^r(P), \check{C}^r(Q)) \leq d. \tag{2.24}$$

□

Chapter 3

What functions are stable?

Chapter 4

Operations on stable functions

Chapter 5

The space of stable functions

Chapter 6

Conclusion

Appendix A

Weighted Čech complex can be represented as a GDF

Lemma A.1 *For any point cloud $P \subseteq \mathbb{R}^d$, a function $g : P \to \mathbb{R}_{\geq 0}$, and a parameter $q \in [1, \infty)$, the weighted Čech filtration $V_q^a[P, g]$ is equal to the sublevel set of the generalized density function*

$$f_{g,q}(P, x) = \min_{p \in P} (d(x, p)^q + g(p)^q)^{1/q}. \tag{A.1}$$

Proof Fix P, g, q, and a. A point $x \in \mathbb{R}^d$ being included in $V_q^a[P, g]$ is equivalent to the condition

$$\exists p \in P : d(x, p) \leq r_p^{(q)}(a). \tag{A.2}$$

Suppose $a \geq g(p)$. Then, we have two cases:

1. $q = \infty$. Then, the condition is equivalent to

$$\exists p \in P : d(x, p) \leq a. \tag{A.3}$$

The proposed GDF takes value

$$f_{g,q}(P, x) = \min_{p \in P} \max(d(x, p), g(p)), \tag{A.4}$$

and the sublevel set $f_{g,q,P}^{-1}(-\infty, a]$ is precisely

$$\{x \in \mathbb{R}^d \mid \exists p \in P : \max(d(x, p), g(p)) \leq a\}, \tag{A.5}$$

which, due to $a \geq g(p)$, is the set

$$\{x \in \mathbb{R}^d \mid \exists p \in P : d(x, p) \leq a\}, \tag{A.6}$$

which is the same as $V_\infty^a[P, g]$.

2. $q < \infty$. Then, the condition is equivalent to

$$\exists p \in P : d(x, p) \leq (a^q - g(p)^q)^{1/q} \tag{A.7}$$

The proposed GDF takes value

$$f_{g,q}(P, x) = \min_{p \in P} (d(x, p)^q + g(p)^q)^{1/q}, \tag{A.8}$$

and the sublevel set $f^{-1}_{g,q,P}(-\infty, a]$ is precisely

$$\{x \in \mathbb{R}^d \mid \exists p \in P : (d(x, p)^q + g(p)^q)^{1/q} \leq a\}, \tag{A.9}$$

which can be seen to be the same with trivial algebraic manipulations:

$$(d(x, p)^q + g(p)^q)^{1/q} \leq a \tag{A.10}$$

$$d(x, p)^q + g(p)^q \leq a^q \tag{A.11}$$

$$d(x, p)^q \leq a^q - g(p)^q \tag{A.12}$$

$$d(x, p) \leq (a^q - g(p)^q)^{1/q} \tag{A.13}$$

Now suppose $a < g(p)$. Then $r_p^{(q)}(a) = -\infty$, and thus the balls contain no points. As $(d(x, p)^q + g(p)^q)^{1/q} \geq g(p) > a$, the sublevel set is also empty.$\square$

Bibliography

[1] Hirokazu Anai, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hiroya Inakoshi, Raphaël Tinarrage, and Yuhei Umeda. Dtm-based filtrations. In *Topological data analysis: the abel symposium 2018*, pages 33–66. Springer, 2020.

[2] Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and topological inference*, volume 57. Cambridge University Press, 2018.

[3] Karol Borsuk. On the imbedding of systems of compacta in simplicial complexes. *Fundamenta Mathematicae*, 35(1):217–234, 1948.

[4] Mickael Buchet, Frederic Chazal, Steve Y. Oudot, and Donald R. Sheehy. Efficient and robust persistent homology for measures, 2014.

[5] Gunnar Carlsson and Mikael Vejdemo-Johansson. *Topological data analysis with applications*. Cambridge University Press, 2021.

[6] Frederic Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules, 2013.

[7] Frederic Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes, 2013.

[8] Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4:667963, 2021.

[9] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *Proceedings of the twenty-first annual symposium on Computational geometry*, pages 263–271, 2005.

[10] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. Applied Mathematics. American Mathematical Society, 2010.

[11] Leonidas J Guibas, Quentin Mérigot, and Dmitriy Morozov. Witnessed k-distance. In *Proceedings of the twenty-seventh annual symposium on Computational geometry*, pages 57–64, 2011.

[12] Pepijn Roos Hoefgeest and Lucas Slot. The christoffel-darboux kernel for topological data analysis, 2022.

[13] Minhyeok Lee and Soyeon Lee. Persistent homology analysis of ai-generated fractal patterns: A mathematical framework for evaluating geometric authenticity. *Fractal and Fractional*, 8(12), 2024.

[14] Jean Leray. Sur la forme des espaces topologiques et sur les points fixes des représentations. *Journal de Mathématiques Pures et Appliquées*, 24:95–167, 1945.

[15] Edwin E Moise. *Geometric topology in dimensions 2 and 3*, volume 47. Springer Science & Business Media, 2013.

[16] Jeff M. Phillips, Bei Wang, and Yan Zheng. Geometric inference on kernel density estimates, 2015.

[17] Patrick Schnider and Simon Weber. Introduction to topological data analysis lecture notes FS 2024. `https://ti.inf.ethz.ch/ew/courses/TDA24/Script.pdf`.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.