ETH
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 $f_{\text{dist}}(P, x) := \min_{p \in P} d(x, p)$, whose sublevel sets are homotopy equivalent to the Čech complex [22], for which the bottleneck distance between persistence diagrams $\text{Dgm}(f_P)$ and $\text{Dgm}(f_Q)$ is bounded by the Hausdorff distance $d_H(P, Q)$ between datasets [8]:

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

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

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

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 [6]. 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 [9]. 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 [11]. 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 [9].

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 [22], a fundamental construction in TDA. This function satisfies a *stability* property [8]:

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

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 [9]. 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, 15, 21]. In practice, alternative functions may offer computational advantages [14, 4], better capture intrinsic structure [1], or incorporate domain knowledge [17]. 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 $P, Q \subseteq \mathbb{R}^d$ 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 is a modification of the Čech complex which allows the balls to grow at non-uniform rates, although it still requires the growth rate to be the same for all points $p \in P$.

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, recognizing that full Lipschitz continuity might be too strong, we consider Morse functions where Lipschitz-like conditions are imposed only on critical values. This aims to capture stability for functions whose may not be Lipschitz continuous globally, but are nonetheless 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.

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 [11].

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

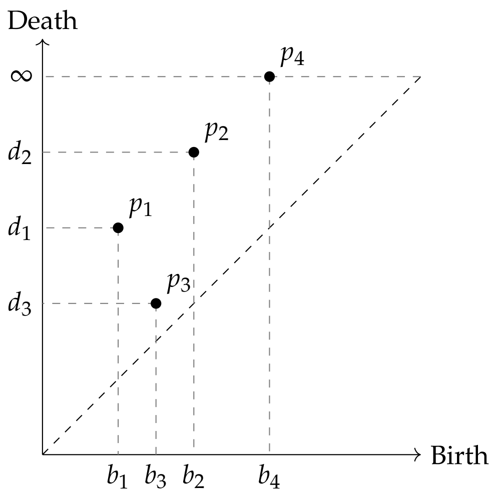

Figure 2.1: An example of a persistence diagram. A point p_i is born at b_i and dies at d_i. The point p_4 does not die, so its death time is ∞. The dashed line indicates the diagonal Δ, where $b_i = d_i$.

the parameter a varies [11]. 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 Tameness

TODO: wording in this section

Many known stability results outlined in Section 2.3 as well as the novel results in this thesis require the involved persistence modules to be q-*tame*.

Definition 2.3 (q-tame persistence module, [7]) *A persistence module* $\mathbb{M}$ *with vector spaces* M_a *and linear maps* $m_a^b : M_a \to M_b$ *is* q-tame *if* m_a^b *is of finite rank for all* $a < b$.

q-tameness is a generalization of the following notion:

Definition 2.4 (Pointwise finite-dimensional persistence module) *A persistence module* $\mathbb{M}$ *is* pointwise finite-dimensional (PFD) *if for every* $a \in \mathbb{R}$, *the vector space* M_a *is finite-dimensional.*

Lemma 2.5 ([7]) *If a persistence module is PFD, then it is* q*-tame.*

The following theorems can be used to prove q-tameness of persistence modules induced by sublevel set filtrations.

Theorem 2.6 ([5]) *Let X be a realisation of a finite complex, and let $f : X \to \mathbb{R}$ be a continuous function. Then the persistence module induced by sublevel sets of f is q-tame.*

While this theorem is powerful, it is of limited use in this thesis, as we consider functions $f_P : \mathbb{R}^d \to \mathbb{R}$, and no triangulation of $\mathbb{R}^d$ is finite, as $\mathbb{R}^d$ is not compact. This condition is relaxed in the following theorem:

Theorem 2.7 ([7], Corollary 3.34) *Let X be a realisation of a* locally *finite complex, and let $f : X \to \mathbb{R}$ be a continuous function which is bounded below and the preimage $f^{-1}(K)$ of every compact set $K \subseteq \mathbb{R}$ is compact. Then the persistence module induced by sublevel sets of f is q-tame.*

This theorem is applicable to $\mathbb{R}^d$, as it admits a locally finite triangulation [13]. The condition that the preimage of every compact set is compact, also known as *properness*, is required to remove pathological functions such as the sine function, whose sublevel sets $\sin^{-1}(-\infty, a]$ have infinite-dimensional 0-homology for all $a \in (-1, 1)$.

2.2 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 2.8 (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 2.9 (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 2.10 *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.*

2.3 Known stability results

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

2.3.1 Čech filtration

As mentioned in Chapter 1, the Čech filtration can be represented as the sublevel set filtration of the nearest-neighbor distance function $f_{\mathrm{dist}}(P, x) := \min_{p \in P} d(x, p)$. 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, 18], 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 2.11 (Stability of the Čech filtration, [8]) *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.

2.3.2 Weighted Čech filtration

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

Definition 2.12 (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 2.13 (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.

2.3.3 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 2.14 (Stability of persistence diagrams, [10]) *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*

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

Remark 2.15 *This theorem also holds when g and h are not necessarily tame, as long as the resulting persistence modules are q-tame [7].*

Importantly, this theorem is applicable to the case $X = \mathbb{R}^d$, as the space $\mathbb{R}^d$ is triangulable [20]. 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 nearest neighbor distance function using this theorem. To do so, we first need to show the following lemma:

Lemma 2.16 *The persistence modules induced by the nearest-neighbor distance function $f_{\text{dist}}(P, x)$ are q-tame.*

Proof The Čech filtration $\check{C}(P)$ is a filtration of a finite simplicial complex if P is finite, and its persistence modules $H_k(\check{C}(P))$ are thus pointwise finite-dimensional. The sublevel set filtration induced by f_{dist} is homotopy equivalent to $\check{C}(P)$ [22], which implies that the persistence modules corresponding to f_{dist} are also PFD, which implies q-tameness by Lemma 2.5. □

Now we can prove the stability of the nearest-neighbor distance function:

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:

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

We can show that $f_{\text{dist},P}(x) \leq f_{\text{dist},Q}(x) + d$ by the same reasoning. 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 2.14 and Lemma 2.16. □

2.3.4 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 2.17 (Interleaving of persistence modules) *Two persistence modules $\mathbb{M}$ and $\mathbb{N}$ with M_a, N_a as their respective vector spaces 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 2.18 (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 under a mild tameness condition:

Theorem 2.19 (Isometry theorem, [7]) *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 2.20 (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 2.17 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 2.21 ([22]) *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 ([22]) 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

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

By Lemma 2.16 and Theorem 2.19, as well as Lemma 2.21, we have

$$\begin{aligned}
d_b(\mathrm{Dgm}_k(\check{C}(P)), \mathrm{Dgm}_k(\check{C}(Q))) &\leq d_i(H_k(\check{C}(P)), H_k(\check{C}(Q))) && (2.23)\\
&\leq d_i(\check{C}(P), \check{C}(Q)) && (2.24)\\
&\leq d. && (2.25)
\end{aligned}$$

□

Chapter 3

Stable Generalized Density Functions

TODO: wording Chapter 2 reviewed foundational stability results for GDFs. This chapter examines the central question of this thesis: *for which GDFs $f(P, x)$ can we guarantee that the sublevel set filtration of f_P is stable?* We explore several classes of GDFs for which stability can be guaranteed, ranging from simple cases to more complex constructions. We also highlight examples of functions that are not stable.

3.1 0-stable generalized density functions

The simplest class of stable GDFs are those whose persistence diagrams are invariant under changes to the input point cloud P.

Definition 3.1 (Point cloud independent GDF) *A GDF $f(P, x)$ is* point cloud independent *if for any two point clouds $P, Q \subseteq \mathbb{R}^d$, we have $f(P, x) = f(Q, x)$ for all $x \in \mathbb{R}^d$. Equivalently, $f_P = f_Q$ as functions $\mathbb{R}^d \to \mathbb{R}$.*

Theorem 3.2 *A point cloud independent GDF $f : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ is 0-stable.*

Proof If $f(P, x)$ is point cloud independent, then for any two finite point clouds $P, Q \subseteq \mathbb{R}^d$, we have identical sublevel set filtrations $\{f_P^{-1}(-\infty, a]\}_a$ and $\{f_Q^{-1}(-\infty, a]\}_a$. This implies that their persistence modules and persistence diagrams are also identical. As the bottleneck distance between identical persistence diagrams is zero [11], we have

$$d_b(\mathrm{Dgm}(f_P), \mathrm{Dgm}(f_Q)) = 0 \leq 0 \cdot d_H(P, Q), \tag{3.1}$$

satisfying the condition for 0-stability. □

However, point cloud independence is a sufficient, but not necessary, condition for 0-stability. Some GDFs whose values depend on P can still yield identical persistence diagrams for all P, thus being 0-stable. Consider the

GDF $f : \mathcal{P}(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ defined by $f(P, x) = x + |P|$. The sublevel sets $f_P^{-1}(-\infty, a]$ are exactly the intervals $(-\infty, a - |P|]$, which always have exactly one connected component that persists indefinitely. Thus, the dimension 0 persistence diagram of f_P contains a single point born at $-\infty$ that never dies, and is empty in higher dimensions. Thus, for any two finite point clouds $P, Q \subseteq \mathbb{R}$, the persistence diagrams $\mathrm{Dgm}(f_P)$ and $\mathrm{Dgm}(f_Q)$ are identical, and we have 0-stability without independence on the point cloud. This example also trivially generalizes to $\mathbb{R}^d$ with roots of f defining a hyperplane. TODO: reword

3.2 Point cloud Lipschitz generalized density functions

A powerful sufficient condition for stability is a form of Lipschitz continuity with respect to the point cloud:

Definition 3.3 (Point cloud Lipschitz (PC-Lipschitz) GDF) *A GDF f is c-*PC Lipschitz *if, when considered as a map that assigns to each finite non-empty point cloud $P \subseteq \mathbb{R}^d$ a function $f_P : \mathbb{R}^d \to \mathbb{R}$, it is Lipschitz continuous with constant c, where the space of finite point clouds 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. Formally, for any non-empty finite point clouds $P, Q \subseteq \mathbb{R}^d$, we have*

$$\|f_P - f_Q\|_\infty \leq c \cdot d_H(P, Q). \tag{3.2}$$

Lemma 3.4 *If f is a c-PC Lipschitz GDF and for all finite $P \subseteq \mathbb{R}^d$, the induced persistence modules are q-tame, then f is c-stable.*

Proof Let f be a c-PC Lipschitz GDF. By definition, for any non-empty two finite point clouds $P, Q \subseteq \mathbb{R}^d$, we have

$$\|f_P - f_Q\|_\infty \leq c \cdot d_H(P, Q). \tag{3.3}$$

Since the persistence modules induced by f_P and f_Q are q-tame, by Theorem 2.14, the bottleneck distance between their persistence diagrams is bounded by the L_∞-norm of the difference $f_P - f_Q$. Combining these two inequalities, we obtain:

$$d_b(\mathrm{Dgm}(f_P), \mathrm{Dgm}(f_Q)) \leq \|f_P - f_Q\|_\infty \leq c \cdot d_H(P, Q), \tag{3.4}$$

which shows that f is c-stable. □

While this lemma is a straightforward consequence of Theorem 2.14, it provides a versatile tool for establishing stability, as demonstrated in the following sections.

3.3 Weighted Čech filtration revisited

The stability of the weighted Čech filtration (and thus DTM-filtrations) was stated in Theorem 2.13. We provide a more direct proof of this result than in the original paper [1] by showing that the corresponding GDF is PC-Lipschitz.

Theorem 3.5 *The GDF*

$$f_{g,q}(P,x) = \min_{p \in P} (d(x,p)^q + g(p)^q)^{1/q} \tag{3.5}$$

for the weighted Čech filtration is d-stable with $d = (1+c^q)^{1/q}$*, where c is the Lipschitz constant of the weight function* $g : \mathbb{R}^d \to \mathbb{R}^+$*.*

Proof Let $P, Q \subseteq \mathbb{R}^d$ be two finite point clouds with Hausdorff distance $d_H(P,Q)$. Similarly to the proof of Čech stability using Theorem 2.14, fix $x \in \mathbb{R}^d$ and let $p \in P$ be the point where the minimum for $(d(x,p)^q + g(p)^q)^{1/q}$ is attained. This implies that

$$f_{g,q}(P,x) = (d(x,p)^q + g(p)^q)^{1/q}. \tag{3.6}$$

By the definition of Hausdorff distance, there exists a point $q' \in Q$ such that $d(p,q') \leq d_H(P,Q)$, which implies that

$$d(x,q') \leq d(x,p) + d(p,q') \leq d(x,p) + d_H(P,Q). \tag{3.7}$$

By monotonicity of (3.6) with respect to $d(x,p)$ and $g(p)$, we have

$$\begin{aligned}
f_{g,q}(Q,x) &= \min_{q'' \in Q} (d(x,q'')^q + g(q'')^q)^{1/q} && (3.8)\\
&\leq (d(x,q')^q + g(q')^q)^{1/q} && (3.9)\\
&\leq ((d(x,p) + d_H(P,Q))^q + (g(p) + c \cdot d_H(P,Q))^q)^{1/q} && (3.10)\\
&= \|(d(x,p) + d_H(P,Q), g(p) + c \cdot d_H(P,Q))\|_q. && (3.11)
\end{aligned}$$

Comparing the values of $f_{g,q}(P,x)$ and $f_{g,q}(Q,x)$, we have

$$\begin{aligned}
&f_{g,q}(Q,x) - f_{g,q}(P,x)\\
&\leq \|(d(x,p) + d_H(P,Q), g(p) + c \cdot d_H(P,Q))\|_q - \|(d(x,p), g(p))\|_q && (3.12)\\
&\leq \|(d(x,p) + d_H(P,Q) - d(x,p), g(p) + c \cdot d_H(P,Q) - g(p))\|_q && (3.13)\\
&= \|(d_H(P,Q), c \cdot d_H(P,Q))\|_q, && (3.14)
\end{aligned}$$

and by the same reasoning with P and Q swapped, we have

$$f_{g,q}(P,x) - f_{g,q}(Q,x) \leq \|(d_H(P,Q), c \cdot d_H(P,Q))\|_q, \tag{3.15}$$

which combined gives us

$$\begin{aligned}\|f_{g,q}(P,\cdot) - f_{g,q}(Q,\cdot)\|_\infty &= \sup_{x\in\mathbb{R}^d} |f_{g,q}(P,x) - f_{g,q}(Q,x)| && (3.16)\\ &\leq \|(d_H(P,Q), c\cdot d_H(P,Q))\|_q && (3.17)\\ &= d_H(P,Q)\cdot \|(1,c)\|_q && (3.18)\\ &= d_H(P,Q)\cdot (1+c^q)^{1/q}, && (3.19)\end{aligned}$$

which is exactly the condition for d-PC Lipschitz continuity with $d = (1 + c^q)^{1/q}$.

The induced persistence modules of the weighted Čech filtration are pointwise finite dimensional for finite point clouds P ([1], Proposition 3.1), which implies that they are q-tame. By Lemma 3.4, the weighted Čech filtration is d-stable with $d = (1 + c^q)^{1/q}$. □

3.4 Generalized Čech filtrations

The nearest-neighbor distance function $f_{\text{dist}}(P,x) = \min_{p\in P} d(x,p)$ can be seen as taking a minimum over a set of functions $h_p(x) = d(x,p)$, each centered around a point $p \in P$. We can generalize this to $f(P,x) = \min_{p\in P} h(x,p)$, where $h(x,p) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a more general "kernel" function. We refer to GDFs of this form as defining *generalized Čech filtrations*. The stability of such GDFs depends on the properties of the kernel h.

3.4.1 Isotropic, monotone and Lipschitz kernels

Borrowing terminology from kernel methods, we call a kernel *isotropic* if it depends only on the distance between its arguments [12].

Definition 3.6 (Isotropic kernel) *A kernel $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is* isotropic *if there exists a function $k : \mathbb{R}_+ \to \mathbb{R}$ such that $h(x,p) = k(d(x,p))$ for all $x, p \in \mathbb{R}^d$.*

Such kernels correspond to generalized Čech filtrations where the radius of the balls grows with the same rate for each point $p \in P$, but that rate may not be uniform.

A generalized Čech filtration is stable if the kernel h is isotropic, monotone and Lipschitz continuous:

Theorem 3.7 *Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a kernel that is isotropic, such that k is monotone increasing and c-Lipschitz continuous. Then the generalized Čech filtration using h is c-stable.*

Proof We want to show that the corresponding GDF f is c-PC Lipschitz. Let $P, Q \subseteq \mathbb{R}^d$ be two finite point clouds. Fix a point $x \in \mathbb{R}^d$ and $p \in P$ be such that $f(P,x) = k(d(x,p))$. There exists $q' \in Q$ such that $d(p,q') \leq d_H(P,Q)$ and

$$d(x,q') \leq d(x,p) + d(p,q') \leq d(x,p) + d_H(P,Q). \tag{3.20}$$

Since k is monotone and c-Lipschitz, we have:

$$\begin{aligned} f(Q,x) &= \min_{q \in Q} k(d(x,q)) && (3.21)\\ &\leq k(d(x,q')) && (3.22)\\ &\leq k(d(x,p) + d_H(P,Q)) && (3.23)\\ &\leq k(d(x,p)) + c \cdot d_H(P,Q) && (3.24)\\ &= f(P,x) + c \cdot d_H(P,Q), && (3.25) \end{aligned}$$

and by the same reasoning with P and Q swapped, we have

$$f(P,x) \leq f(Q,x) + c \cdot d_H(P,Q). \tag{3.26}$$

Combining these inequalities, we obtain

$$\begin{aligned} \|f(P,\cdot) - f(Q,\cdot)\|_\infty &= \sup_{x \in \mathbb{R}^d} |f(P,x) - f(Q,x)| && (3.27)\\ &\leq c \cdot d_H(P,Q). && (3.28) \end{aligned}$$

The induced persistence modules of $f(P,x)$ are q-tame by the same reasoning as in the proof of Lemma 2.5. Therefore, by Lemma 3.4, the generalized Čech filtration using h is c-stable. $\square$

3.4.2 Kernels Lipschitz in p

We can generalize further by considering $f(P,x) = \min_{p \in P} h(x,p)$ where the kernel $h(x,p)$ is c-Lipschitz with respect to its second argument p, for any fixed x:

$$|h(x,p_1) - h(x,p_2)| \leq c \cdot d(p_1,p_2). \tag{3.29}$$

Such kernels correspond to generalized Čech filtrations where the standard process of growing balls around points $p \in P$ is modified significantly; the growing shapes may not be symmetric, connected, grow at the same rate for different $p \in P$ or even include the point p itself. The only restriction imposed by Lipschitzness of the kernel is that as a point p moves, the shape changes in a controlled manner. Surprisingly, even this weak condition coupled with q-tameness is sufficient for stability. TODO: rephrase this paragraph, explain more formally that "shapes" are just the sublevel sets

Theorem 3.8 *Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a kernel that is c-Lipschitz continuous with respect to its second argument. Then the generalized Čech filtration using h is c-stable if the induced persistence modules are q-tame.*

Proof Fix $x \in \mathbb{R}^d$ and $p \in P$. Then we have $q' \in Q$ such that $d(p, q') \leq d_H(P, Q)$. By Lipschitzness of h, we have

$$|h(x, q') - h(x, p)| \leq c \cdot d(q', p) \leq c \cdot d_H(P, Q) \tag{3.30}$$
$$f(Q, x) \leq h(x, q') \leq c \cdot d_H(P, Q) + h(x, p), \tag{3.31}$$

and as this holds for any $p \in P$, we have

$$f(Q, x) \leq c \cdot d_H(P, Q) + f(P, x) \tag{3.32}$$
$$f(Q, x) - f(P, x) \leq c \cdot d_H(P, Q). \tag{3.33}$$

By the same reasoning, we have the same with P and Q swapped, and combining the two inequalities, we obtain

$$|f(P, x) - f(Q, x)| \leq c \cdot d_H(P, Q) \tag{3.34}$$

for any x, which by Lemma 3.4 implies that f is c-stable. □

This theorem is a generalization of Theorem 3.7. An isotropic kernel $h = d \circ k$ with k being c-Lipschitz is also c-Lipschitz with respect to p. This follows because the distance function is 1-Lipschitz, and a composition of a c-Lipschitz function with a 1-Lipschitz function is c-Lipschitz [23].

The q-tameness condition for the induced persistence modules, while crucial, is not always straightforward to verify directly for a given kernel h. In the following lemma, we provide sufficient conditions for this.

Lemma 3.9 *Let $f(P, x) = \min_{p \in P} h(x, p)$. If, for every $p \in P$, the function $h_p(x) := h(x, p)$ is continuous in x, bounded below and proper (i.e., $h_p^{-1}(K)$ is compact for every compact $K \subseteq \mathbb{R}$), then $f_P(x)$ is continuous, bounded below and proper. Consequently, by Theorem 2.7, the induced persistence modules are q-tame.*

Proof To use Theorem 2.7, we need to show for f_P:

1. Continuity: $f_P(x)$ is the minimum of a finite set of continuous functions $h_p(x)$, thus f_P is continuous.
2. Bounded below: If each $h_p(x) \geq M_p$ for some M_p, then $f_P(x) = \min_p h_p(x) \geq \min_p M_p$, so f_P is bounded below.
3. Properness: A function g is proper if and only if for every sequence $\{x_i\}$ such that $x_i \to \infty$, we have $g(x_i) \to \infty$ ([16], Proposition 2.17). As every $h_p(x)$ is proper, then $p \in P$, we have $h_p(x_i) \to \infty$. Since $f_P(x_i) = \min_p h_p(x_i)$, it follows that $f_P(x_i) \to \infty$ as well. Thus, f_P is proper.

With f_P continuous, bounded below, and proper, Theorem 2.7 guarantees that the induced persistence modules are q-tame. □

3.4.3 Kernels depending on the point cloud

A further generalization allows the kernel itself to depend on the point cloud P:

$$f(P,x) = \min_{p \in P} h(x,p,P). \tag{3.35}$$

An example could be a distance metric defined by locally estimated covariance, leading to growing ellipses whose shape adapts to the local density of P. We prove two theorems that guarantee stability of such GDFs.

Theorem 3.10 *Let $h : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ be a kernel that is c-Lipschitz continuous with respect to its second argument (the point P) and d-Lipschitz continuous with respect to its third argument (the point cloud P, using d_H on $\mathcal{P}(\mathbb{R}^d)$):*

$$|h(x,p_1,P) - h(x,p_2,P)| \leq c \cdot d(p_1,p_2), \tag{3.36}$$

$$|h(x,p,P_1) - h(x,p,P_2)| \leq d \cdot d_H(P_1,P_2). \tag{3.37}$$

Then the GDF $f(P,x) = \min_{p \in P} h(x,p,P)$ is $(c+d)$-stable if the induced persistence modules are q-tame.

Proof Fix $x \in \mathbb{R}^d$ and $p \in P$. Then we have $q' \in Q$ such that $d(p,q') \leq d_H(P,Q)$. By Lipschitzness of h, we have

$$\begin{aligned} &|h(x,p,P) - h(x,q',Q)| && (3.38)\\ &\leq |h(x,p,P) - h(x,p,Q)| + |h(x,p,Q) - h(x,q',Q)| && (3.39)\\ &\leq d \cdot d_H(P,Q) + c \cdot d(p,q'), && (3.40)\end{aligned}$$

and as this holds for any $p \in P$, we have

$$\begin{aligned} f(Q,x) - f(P,x) &\leq h(x,q',Q) - f(P,x) && (3.41)\\ &\leq h(x,q',Q) - h(x,p,P) && (3.42)\\ &\leq (c+d) \cdot d_H(P,Q). && (3.43)\end{aligned}$$

By the same reasoning, we have the same inequality with P and Q swapped, and combining the two inequalities, we obtain

$$|f(P,x) - f(Q,x)| \leq (c+d) \cdot d_H(P,Q). \tag{3.44}$$

By Lemma 3.4, this implies that f is $(c+d)$-PC Lipschitz. □

Theorem 3.11 *Let $h : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ be a kernel that is c-Lipschitz continuous with respect to its second argument and d-Lipschitz continuous with respect to every element in P, i.e.:*

$$|h(x,p_1,P) - h(x,p_2,P)| \leq c \cdot d(p_1,p_2), \tag{3.45}$$

$$|h(x,p,P \cup \{p_1\}) - h(x,p,P \cup \{p_2\})| \leq d \cdot d(p_1,p_2) \tag{3.46}$$

$$|h(x,p,\{p_1\}) - h(x,p,\{p_2\})| \leq d \cdot d(p_1,p_2). \tag{3.47}$$

Then the generalized Čech filtration using h is $(c + d \cdot \max(|P|,|Q|))$-stable if the induced persistence modules are q-tame.

Proof If $|P| = |Q|$, we can match every point $p_i \in P$ with a point $q_i \in Q$ such that $d(p_i, q_i) \leq d_H(P, Q)$, and thus we have

$$
\begin{aligned}
&|h(x, p, \{p_i\}_{i=1}^{n}) - h(x, p, \{q_i\}_{i=1}^{n})| && (3.48)\\
&= |h(x, p, \{p_i\}_{i=1}^{n-1} \cup \{p_n\}) - h(x, p, \{p_i\}_{i=1}^{n-1} \cup \{q_n\})| && (3.49)\\
&\leq |h(x, p, \{p_i\}_{i=1}^{n-1}) - h(x, p, \{q_i\}_{i=1}^{n-1})| + d \cdot d_H(P, Q) && (3.50)\\
&\leq |h(x, p, \{p_i\}_{i=1}^{n-2}) - h(x, p, \{q_i\}_{i=1}^{n-2})| + 2 \cdot d \cdot d_H(P, Q) && (3.51)\\
&\leq \dots && (3.52)\\
&\leq |P| \cdot d \cdot d_H(P, Q). && (3.53)
\end{aligned}
$$

If $|P| < |Q|$, we can add points to P that are arbitrarily close to existing points in P such that the Hausdorff distance $d_H(P, Q)$ changes arbitrarily little, and $f(P, x)$ also changes arbitrarily little. The same is true with P and Q swapped. Thus, we can assume that $|P| = |Q|$ without loss of generality. This means that

$$|h(x, p, P) - h(x, p, Q)| \leq d \cdot \max(|P|, |Q|) \cdot d_H(P, Q). \tag{3.54}$$

Following the same reasoning as in the previous proof, we have

$$
\begin{aligned}
&|h(x, p, P) - h(x, q', Q)| && (3.55)\\
&\leq |h(x, p, P) - h(x, p, Q)| + |h(x, p, Q) - h(x, q', Q)| && (3.56)\\
&\leq d \cdot \max(|P|, |Q|) \cdot d_H(P, Q) + c \cdot d_H(P, Q). && (3.57)
\end{aligned}
$$

□

TODO: maybe add the growing ellipses example here?

3.5 Morse functions

The PC-Lipschitz condition is a strong global condition, which may not always hold for some useful GDFs. For example, we can consider a Čech-like GDF

$$f_P(x) = \min_{p \in P} d(x, p)^2, \tag{3.58}$$

where the balls grow faster the larger they are. While this function is not Lipschitz continuous on the whole space $\mathbb{R}^d$, it is Lipschitz continuous on some subset of $\mathbb{R}^d$ that covers the point cloud P. We can intuit that Lipschitz continuity everywhere is too strong of a condition, as it does not matter how the function behaves far away from some region of interest. TODO: terrible wording, what is this “region of interest”? TODO: one more example - function with no critical points, but not Lipschitz with the constant that we want

We can relax the PC-Lipschitz condition by restricting ourselves to *Morse functions*, that is, smooth functions $f_P : \mathbb{R}^d \to \mathbb{R}$ whose critical points $x \in \mathrm{Crit}(f_P)$ (where the gradient vanishes) are non-degenerate, i.e., the Hessian matrix $H(f_P)(x)$ is non-singular.

Theorem 3.12 *Let $f : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ be a GDF such that for any finite point clouds $P, Q \subseteq \mathbb{R}^d$, the functions f_P and f_Q are Morse, all sublevel sets $f_P^{-1}(-\infty, a]$ and $f_Q^{-1}(-\infty, a]$ are compact and the following condition holds:*

$$\forall x \in \mathrm{Crit}(f_P) \cup \mathrm{Crit}(f_Q) : |f_P(x) - f_Q(x)| \leq c \cdot d_H(P, Q), \tag{3.59}$$

or more concisely, the functions $f_P|_{\mathrm{Crit}(f_P)\cup\mathrm{Crit}(f_Q)}$ and $f_Q|_{\mathrm{Crit}(f_P)\cup\mathrm{Crit}(f_Q)}$ are c-PC-Lipschitz continuous. Then the GDF f is c-stable, if the induced persistence modules are q-tame.

Proof As f_P is Morse and its sublevel sets are compact, each sublevel set $f_P^{-1}(-\infty, a]$ has the homotopy type of a finite CW-complex, with one cell of dimension λ for each critical point of index λ in $f_P^{-1}(-\infty, a]$ ([19], Theorem 3.5 and the remark following it). Let's denote such a CW-complex by $C_P(a)$. The previous statement can be written concisely as

$$C_P(a) \simeq f_P^{-1}(-\infty, a], \tag{3.60}$$

for all $k \in \mathbb{N}$. TODO: □

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.□

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] Francesca Cagliari and Claudia Landi. Finiteness of rank invariants of multidimensional persistent homology groups, 2010.

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

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

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

[9] 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.

[10] 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.

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

[12] Marc G Genton. Classes of kernels for machine learning: a statistics perspective. *Journal of machine learning research*, 2(Dec):299–312, 2001.

[13] Loukas Grafakos et al. *Classical fourier analysis*, volume 2. Springer, 2008.

[14] 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.

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

[16] John M Lee and John M Lee. *Smooth manifolds*. Springer, 2003.

[17] 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.

[18] 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.

[19] John Willard Milnor. *Morse theory*. Number 51. Princeton university press, 1963.

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

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

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

[23] Nik Weaver. *Lipschitz algebras*. World Scientific, 2018.

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.