



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Persistence of generalised 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-neighbour distance function $f_{\text{dist}}(P, x) := \min_{p \in P} d(x, p)$, whose sublevel sets are homotopy equivalent to the Čech complex [29], 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 *generalised density functions*, which are functions $f : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$, 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 generalised density functions. Specifically, we prove stability bounds for:

- Several generalisations of the nearest-neighbour 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 generalised density functions for topological data analysis.

Contents

Contents	iii
1 Introduction	1
2 Background	5
2.1 Tameness	6
2.2 Quantifying stability	7
2.3 Known stability results	8
2.3.1 Čech filtration	8
2.3.2 Weighted Čech filtration	8
2.3.3 Stability of L_∞ -norm-bounded functions	9
2.3.4 Stability of functions with interleaved persistence mod- ules	10
3 Stable Generalised Density Functions	15
3.1 0-stable generalised density functions	15
3.2 Unstable generalised density functions	16
3.3 Point cloud Lipschitz generalised density functions	16
3.4 Weighted Čech filtration revisited	18
3.5 Generalised Čech filtrations	19
3.5.1 Isotropic, monotone and Lipschitz kernels	19
3.5.2 Kernels Lipschitz in p	20
3.5.3 Kernels depending on the point cloud	22
3.6 Morse functions	24
4 Operations on stable functions	27
4.1 Affine transformations	27
4.2 Addition of stable functions	30
4.3 Minimum and maximum of stable functions	32
5 The space of stable functions	37

CONTENTS

5.1	The space of c -PC-Lipschitz functions	37
5.2	The space of c -stable functions	39
5.3	The relationship between \mathcal{S}_c and \mathcal{L}_c	40
6	Conclusion	41
6.1	Summary of findings	41
6.2	Open questions and future work	42
A	Weighted Čech complex can be represented as a GDF	45
B	The generalised Čech complex using the Mahalanobis distance kernel is stable	47
	Bibliography	51

Chapter 1

Introduction

Topological Data Analysis (TDA) provides a robust framework for understanding the shape of data [6]. Its primary tool, *persistent homology*, extracts topological features across different scales [2] and is traditionally 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 $\mathbb{R}^d \rightarrow \mathbb{R}$ that encodes the data’s geometry, and then computing the persistent homology of the function’s sublevel (or superlevel) set filtration [12]. The resulting persistent homology captures topological features such as connected components, loops, and voids, and encodes them in a *persistence diagram*, which summarises their birth and death across the filtration [9].

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

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

where d is a metric on \mathbb{R}^d . The sublevel sets of this function are homotopy equivalent to the Čech filtration [29], a fundamental construction in TDA. This function satisfies a *stability* property [8]:

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

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-neighbour function is stable, it is just one of many possible density-like functions for TDA [1, 17, 27]. In practice, alternative functions may offer computational advantages [15, 4], better capture intrinsic structure [1], or incorporate domain knowledge [19]. For example, the DTM filtration [1] modifies the nearest-neighbour function to make it more robust to noise and outliers. This motivates the study of *generalised density functions* (GDFs), which are functions of the form

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

For a given point cloud P , such a function defines a real-valued function $f_P : \mathbb{R}^d \rightarrow \mathbb{R}$, 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

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

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 $\text{Dgm}(f_P)$ 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-neighbour 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 $f_P : \mathbb{R}^d \rightarrow \mathbb{R}$ 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, recognising 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 generalised density functions. They also provide practical guidance for designing stable filtrations.

Chapter 2

Background

This chapter reviews the mathematical foundations for stability of generalised 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 [12].

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

Definition 2.1 (Generalised density function) *A generalised density function is a map $f : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$, 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 $f_P : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $f_P : \mathbb{R}^d \rightarrow \mathbb{R}$, its sublevel set at a value $a \in \mathbb{R}$ is*

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

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

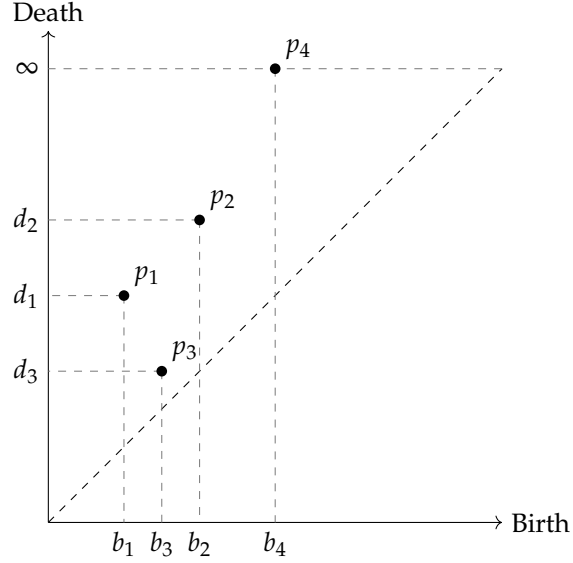


Figure 2.1: An example of a persistence diagram. A point p_i is born at b_i and dies at d_i . The point p_4 does not die, so its death time is ∞ . The dashed line indicates the diagonal Δ , where $b_i = d_i$.

the parameter a varies [12]. The *persistence diagram*, denoted $\text{Dgm}_k(f_P)$, 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

Many established stability results and the original findings of this thesis rely on the involved persistence modules being *q-tame*, a condition ensuring the finiteness of the rank for maps induced by filtrations.

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

q-tameness is a generalisation 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 $f : X \rightarrow \mathbb{R}$ 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 $f_P : \mathbb{R}^d \rightarrow \mathbb{R}$, 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 $f : X \rightarrow \mathbb{R}$ 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 [14]. 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-neighbour 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*

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

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 neighbour 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*

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

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 $\Delta = \{(x, x) \mid x \in \mathbb{R}\}$.

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 generalised density function $f : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$ 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*

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

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-neighbour distance function $f_{\text{dist}}(P, x) := \min_{p \in P} d(x, p)$. 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, 20], 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*

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

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

2.3.2 Weighted Čech filtration

The weighted Čech filtration is a generalisation 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 $g : P \rightarrow \mathbb{R}_{\geq 0}$. For every $p \in P, a \in \mathbb{R}^+$, we define the function $r_p(a)$ to be:*

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

For $q = \infty$, we also define

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

$g(p)$ 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

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

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

The weighted Čech filtration is a strict generalisation 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

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

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 $g : P \cup Q \rightarrow \mathbb{R}^+$ be a Lipschitz continuous function with Lipschitz constant c . Then for any dimension $k \geq 0$, we have*

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

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 $g, h : X \rightarrow \mathbb{R}$ be two continuous tame functions. Then for all $k \geq 0$, the persistence diagrams of g and h satisfy*

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

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 [24]. We can easily adapt this theorem to the case of GDFs:

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

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 neighbour 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-neighbour 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 $\check{C}(P)$ [29], which implies that the persistence modules corresponding to f_{dist} are also PFD, which implies q -tameness by Lemma 2.5. \square

Now we can prove the stability of the nearest-neighbour 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

$$d(x, q') \leq d(x, p) + d(p, q') \leq d(x, p) + d. \quad (2.13)$$

and thus:

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

We can show that $f_{\text{dist}, P}(x) \leq f_{\text{dist}, Q}(x) + d$ by the same reasoning. Thus, we have

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

which implies that f_{dist} is 1-stable by Theorem 2.14 and Lemma 2.16. \square

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 $\varphi_a : M_a \rightarrow N_{a+\varepsilon}$ and $\psi_a : N_a \rightarrow M_{a+\varepsilon}$ such that the following diagrams commute for all $a < a'$:

$$\begin{array}{ccc}
 M_a & \xrightarrow{m_a^{a'}} & M_{a'} \\
 \searrow \varphi_a & & \searrow \varphi_{a'} \\
 & N_{a+\varepsilon} & \xrightarrow{n_{a+\varepsilon}^{a'+\varepsilon}} N_{a'+\varepsilon}
 \end{array}
 \qquad
 \begin{array}{ccc}
 M_a & \xrightarrow{m_a^{a+2\varepsilon}} & M_{a+2\varepsilon} \\
 \searrow \varphi_a & & \nearrow \psi_{a+\varepsilon} \\
 & N_{a+\varepsilon} &
 \end{array}$$

$$\begin{array}{ccc}
 & M_{a+\varepsilon} & \xrightarrow{n_{a+\varepsilon}^{a'+\varepsilon}} M_{a'+\varepsilon} \\
 \nearrow \psi_a & & \nearrow \psi_{a'} \\
 N_a & \xrightarrow{n_a^{a'}} & N_{a'}
 \end{array}
 \qquad
 \begin{array}{ccc}
 & M_{a+\varepsilon} & \searrow \varphi_{a+\varepsilon} \\
 \nearrow \psi_a & & \\
 N_a & \xrightarrow{n_a^{a+2\varepsilon}} & N_{a+2\varepsilon}
 \end{array}$$

Note that \mathbb{M} and \mathbb{N} are 0-interleaved if and only if they are isomorphic [29]. 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:

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

Finally, the interleaving distance and the bottleneck distance are equal as long as the persistence modules are q -tame:

Theorem 2.19 (Isometry theorem, [7]) Let \mathbb{M}, \mathbb{N} be q -tame persistence modules. Then

$$d_i(\mathbb{M}, \mathbb{N}) = d_b(\text{Dgm}(\mathbb{M}), \text{Dgm}(\mathbb{N})). \quad (2.17)$$

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 $\varphi_a : F_a \rightarrow G_{a+\varepsilon}$ and $\psi_a : G_a \rightarrow F_{a+\varepsilon}$ 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 ([29]) *For any two filtrations \mathcal{F} and \mathcal{G} over \mathbb{R} , we have*

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

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

Proof ([29]) Let X be a metric 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 $\check{C}^r(P)$ at scale r is homotopy equivalent to the union of the balls centered at points of P , formally

$$\check{C}^r(P) \simeq \bigcup_{p \in P} B(p, r). \quad (2.19)$$

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

$$d(x, q) \leq d(x, p) + d(p, q) \leq r + d, \quad (2.20)$$

implying $x \in B(q, r + d)$. As x was arbitrary, this provides an inclusion

$$\bigcup_{p \in P} B(p, r) \subseteq \bigcup_{q \in Q} B(q, r + d). \quad (2.21)$$

By symmetric reasoning, we have the same with the roles of P and Q reversed. With these three facts, we see that the following diagram commutes up to homotopy, as all maps are either inclusions or homotopies:

$$\begin{array}{ccccc}
 \check{C}^r(P) & \hookrightarrow & \check{C}^{r+d}(P) & \hookrightarrow & \check{C}^{r+2d}(P) \\
 \uparrow \simeq & & \uparrow \simeq & & \uparrow \simeq \\
 \bigcup_{p \in P} B(p, r) & \hookrightarrow & \bigcup_{p \in P} B(p, r + d) & \hookrightarrow & \bigcup_{p \in P} B(p, r + 2d) \\
 & \searrow & & \swarrow & \\
 & & \bigcup_{q \in Q} B(q, r) & \hookrightarrow & \bigcup_{q \in Q} B(q, r + d) & \hookrightarrow & \bigcup_{q \in Q} B(q, r + 2d) \\
 & \swarrow & & \searrow & \\
 \bigcup_{q \in Q} B(q, r) & \hookrightarrow & \bigcup_{q \in Q} B(q, r + d) & \hookrightarrow & \bigcup_{q \in Q} B(q, r + 2d) \\
 \uparrow \simeq & & \uparrow \simeq & & \uparrow \simeq \\
 \check{C}^r(Q) & \hookrightarrow & \check{C}^{r+d}(Q) & \hookrightarrow & \check{C}^{r+2d}(Q)
 \end{array} \quad (2.22)$$

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

$$d_i(\check{C}(P), \check{C}(Q)) \leq d. \quad (2.23)$$

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

$$d_b(\mathrm{Dgm}_k(\check{C}(P)), \mathrm{Dgm}_k(\check{C}(Q))) \leq d_i(H_k(\check{C}(P)), H_k(\check{C}(Q))) \quad (2.24)$$

$$\leq d_i(\check{C}(P), \check{C}(Q)) \quad (2.25)$$

$$\leq d, \quad (2.26)$$

which is precisely the statement of stability of the Čech filtration. \square

Chapter 3

Stable Generalised Density Functions

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 generalised 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 $\mathbb{R}^d \rightarrow \mathbb{R}$.*

Theorem 3.2 *A point cloud independent GDF $f : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$ 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 [12], we have

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

satisfying the condition for 0-stability. \square

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.

Example 3.3 Consider the GDF $f : \mathcal{P}(\mathbb{R}) \times \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(P, x) = x + g(P)$, where $g : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$ is an arbitrary function. The sublevel sets $f_P^{-1}(-\infty, a]$ are exactly the intervals $(-\infty, a - g(P)]$, 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 the PD is empty in higher dimensions. Thus, for any two finite point clouds $P, Q \subseteq \mathbb{R}$, the persistence diagrams $\text{Dgm}(f_P)$ and $\text{Dgm}(f_Q)$ are identical, and we have 0-stability without independence on the point cloud. This example also trivially generalises to \mathbb{R}^d where f defines a hyperplane.

3.2 Unstable generalised density functions

To form an intuition for stability, we construct a simple GDF that is not c -stable for any $c \in \mathbb{R}^+$, i.e. *unstable*.

Example 3.4 Consider the GDF $f : \mathcal{P}(\mathbb{R}) \times \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$f(P, x) = \begin{cases} 0, & \text{if } x \in P, \\ 1, & \text{otherwise.} \end{cases} \quad (3.2)$$

Consider two point clouds $P = \{\mathbf{0}\}$ and $Q = \{\mathbf{0}, \varepsilon \cdot \mathbf{1}\}$, where $\mathbf{0} \in \mathbb{R}^d$ denotes the vector of zeros and $\mathbf{1} \in \mathbb{R}^d$ is the vector of ones, and $\varepsilon > 0$ is a small positive number. The Hausdorff distance between P and Q is $d_H(P, Q) = d(\mathbf{0}, \varepsilon \cdot \mathbf{1}) = \varepsilon$. As shown in Figure 3.1, the persistence diagrams have bottleneck distance of at least $\frac{1}{\sqrt{2}}$. As $\varepsilon \rightarrow 0$, the bottleneck distance remains constant, while the Hausdorff distance tends to zero, which implies that f is not c -stable for any $c \in \mathbb{R}^+$.

While the example above is discontinuous, discontinuity is not required for instability. The following example has the same 0-dimensional persistence diagrams for $P = \{\mathbf{0}\}$ and $Q = \{\mathbf{0}, \varepsilon \cdot \mathbf{1}\}$ as the function in (3.2), but is continuous:

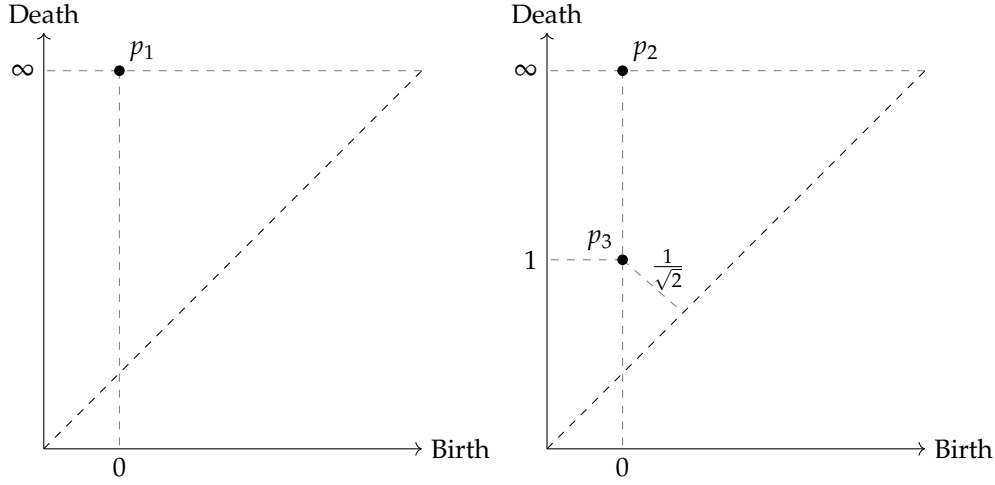
$$f(P, x) = 2 \cdot D(P) \cdot \min_{p \in P} d(x, p), \quad (3.3)$$

where $D(P)$ is the diameter of the point cloud P .

3.3 Point cloud Lipschitz generalised density functions

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

Definition 3.5 (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 $f_P : \mathbb{R}^d \rightarrow \mathbb{R}$, it is Lipschitz continuous with constant c , where the space of finite point clouds is equipped with the Hausdorff distance,



(a) The 0-dimensional PD of f_P . The point p_1 corresponds to the single connected component of the sublevel set $f_P^{-1}(-\infty, a]$, which includes only $\mathbf{0}$ for $a \in [0, 1)$ and \mathbb{R}^d for $a \geq 1$. (b) The 0-dimensional PD of f_Q . The point p_2 corresponds to a connected component of the sublevel set $f_Q^{-1}(-\infty, a]$, which includes only $\mathbf{0}$ for $a \in [0, 1)$ and \mathbb{R}^d for $a \geq 1$, while the point p_3 corresponds to the connected component that contains $\varepsilon \cdot \mathbf{1}$ for $a \in [0, 1)$.

Figure 3.1: Persistence diagrams in dimension 0 of f_P and f_Q for the unstable GDF example. When matching points between the diagrams, the optimal matching connects p_1 to p_2 with cost 0 and p_3 to the diagonal with cost $\frac{1}{\sqrt{2}}$.

and the space of real-valued functions $f_P : \mathbb{R}^d \rightarrow \mathbb{R}$ is equipped with the L_∞ -norm. Formally, for any non-empty finite point clouds $P, Q \subseteq \mathbb{R}^d$, we have

$$\|f_P - f_Q\|_\infty \leq c \cdot d_H(P, Q). \quad (3.4)$$

Lemma 3.6 *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

$$\|f_P - f_Q\|_\infty \leq c \cdot d_H(P, Q). \quad (3.5)$$

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:

$$d_b(\text{Dgm}(f_P), \text{Dgm}(f_Q)) \leq \|f_P - f_Q\|_\infty \leq c \cdot d_H(P, Q), \quad (3.6)$$

which shows that f is c -stable. \square

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.

3.4 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 3.7 *The GDF*

$$f_{g,q}(P, x) = \min_{p \in P} (d(x, p)^q + g(p)^q)^{1/q} \quad (3.7)$$

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 $g : \mathbb{R}^d \rightarrow \mathbb{R}^+$.

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

$$f_{g,q}(P, x) = (d(x, p)^q + g(p)^q)^{1/q}. \quad (3.8)$$

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

$$d(x, q') \leq d(x, p) + d(p, q') \leq d(x, p) + d_H(P, Q). \quad (3.9)$$

By monotonicity of (3.8) with respect to $d(x, p)$ and $g(p)$, we have

$$f_{g,q}(Q, x) = \min_{q'' \in Q} (d(x, q'')^q + g(q'')^q)^{1/q} \quad (3.10)$$

$$\leq (d(x, q')^q + g(q')^q)^{1/q} \quad (3.11)$$

$$\leq ((d(x, p) + d_H(P, Q))^q + (g(p) + c \cdot d_H(P, Q))^q)^{1/q} \quad (3.12)$$

$$= \|(d(x, p) + d_H(P, Q), g(p) + c \cdot d_H(P, Q))\|_q. \quad (3.13)$$

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

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

$$\leq \|(d(x, p) + d_H(P, Q) - d(x, p), g(p) + c \cdot d_H(P, Q) - g(p))\|_q \quad (3.15)$$

$$= \|(d_H(P, Q), c \cdot d_H(P, Q))\|_q, \quad (3.16)$$

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

$$f_{g,q}(P, x) - f_{g,q}(Q, x) \leq \|(d_H(P, Q), c \cdot d_H(P, Q))\|_q, \quad (3.17)$$

which combined gives us

$$\|f_{g,q}(P, \cdot) - f_{g,q}(Q, \cdot)\|_\infty = \sup_{x \in \mathbb{R}^d} |f_{g,q}(P, x) - f_{g,q}(Q, x)| \quad (3.18)$$

$$\leq \|(d_H(P, Q), c \cdot d_H(P, Q))\|_q \quad (3.19)$$

$$= d_H(P, Q) \cdot \|(1, c)\|_q \quad (3.20)$$

$$= d_H(P, Q) \cdot (1 + c^q)^{1/q}, \quad (3.21)$$

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

For finite point clouds P , the weighted Čech filtration induces pointwise finite-dimensional persistence modules ([1], Proposition 3.1), which are by Lemma 2.5 q -tame. By Lemma 3.6, the weighted Čech filtration is d -stable with $d = (1 + c^q)^{1/q}$. \square

3.5 Generalised Čech filtrations

The nearest-neighbour 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 generalise this to $f(P, x) = \min_{p \in P} h(x, p)$, where $h(x, p) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a more general “kernel” function. We refer to GDFs of this form as defining *generalised Čech filtrations*. The stability of such GDFs depends on the properties of the kernel h .

3.5.1 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 [13].

Definition 3.8 (Isotropic kernel) *A kernel $h : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is isotropic if there exists a function $k : \mathbb{R}_+ \rightarrow \mathbb{R}$ such that $h(x, p) = k(d(x, p))$ for all $x, p \in \mathbb{R}^d$.*

Such kernels correspond to generalised 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 generalised Čech filtration is stable if the kernel h is isotropic, monotone and Lipschitz continuous:

Theorem 3.9 *Let $h : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a kernel that is isotropic, such that k is monotone increasing and c -Lipschitz continuous. Then the generalised 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

$$d(x, q') \leq d(x, p) + d(p, q') \leq d(x, p) + d_H(P, Q). \quad (3.22)$$

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

$$f(Q, x) = \min_{q \in Q} k(d(x, q)) \quad (3.23)$$

$$\leq k(d(x, q')) \quad (3.24)$$

$$\leq k(d(x, p) + d_H(P, Q)) \quad (3.25)$$

$$\leq k(d(x, p)) + c \cdot d_H(P, Q) \quad (3.26)$$

$$= f(P, x) + c \cdot d_H(P, Q), \quad (3.27)$$

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

$$f(P, x) \leq f(Q, x) + c \cdot d_H(P, Q). \quad (3.28)$$

Combining these inequalities, we obtain

$$\|f(P, \cdot) - f(Q, \cdot)\|_\infty = \sup_{x \in \mathbb{R}^d} |f(P, x) - f(Q, x)| \quad (3.29)$$

$$\leq c \cdot d_H(P, Q). \quad (3.30)$$

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 3.6, the generalised Čech filtration using h is c -stable. \square

3.5.2 Kernels Lipschitz in p

We generalise 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 :

$$|h(x, p_1) - h(x, p_2)| \leq c \cdot d(p_1, p_2). \quad (3.31)$$

Such kernels correspond to generalised Čech filtrations where the standard process of growing balls around points $p \in P$ is modified significantly; the sublevel sets $f_P^{-1}(-\infty, a]$ 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. Somewhat surprisingly, even this weak condition coupled with q -tameness is sufficient for stability.

Theorem 3.10 *Let $h : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a kernel that is c -Lipschitz continuous with respect to its second argument. Then the generalised 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

$$|h(x, q') - h(x, p)| \leq c \cdot d(q', p) \leq c \cdot d_H(P, Q) \quad (3.32)$$

$$f(Q, x) \leq h(x, q') \leq c \cdot d_H(P, Q) + h(x, p), \quad (3.33)$$

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

$$f(Q, x) \leq c \cdot d_H(P, Q) + f(P, x) \quad (3.34)$$

$$f(Q, x) - f(P, x) \leq c \cdot d_H(P, Q). \quad (3.35)$$

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

$$|f(P, x) - f(Q, x)| \leq c \cdot d_H(P, Q) \quad (3.36)$$

for any x , which by Lemma 3.6 implies that f is c -stable. \square

This theorem generalises Theorem 3.9, as any isotropic kernel $h(x, p) = k(d(x, p))$ with k being c -Lipschitz automatically is c -Lipschitz with respect to p . This follows because the distance function is 1-Lipschitz, and composing it with a c -Lipschitz function yields a c -Lipschitz function [30]. Nonetheless, we included the less general case of Theorem 3.9, as many common kernels fit these conditions.

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 3.11 *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 $x_i \rightarrow \infty$, we have $g(x_i) \rightarrow \infty$ ([18], Proposition 2.17). As every $h_p(x)$ is proper, then $p \in P$, we have $h_p(x_i) \rightarrow \infty$. Since $f_P(x_i) = \min_p h_p(x_i)$, it follows that $f_P(x_i) \rightarrow \infty$ 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. \square

3.5.3 Kernels depending on the point cloud

We further generalise by allowing the kernel itself to depend on the point cloud P :

$$f(P, x) = \min_{p \in P} h(x, p, P). \quad (3.37)$$

Example 3.12 Consider $h(x, p, P)$ that is the pairwise Mahalanobis distance between x and p , defined by locally estimated covariance of the point cloud P [22]. Let $N_k(p, P)$ be the set of k nearest neighbours of p in P . The covariance estimate Σ_p using k nearest neighbours of p is given by

$$\Sigma_p = \frac{1}{k} \sum_{q \in N_k(p, P)} (q - p)(q - p)^\top. \quad (3.38)$$

We would like to invert Σ_p , but it may be numerically unstable or not invertible if k is too small or if some points in $N_k(p, P)$ are close to each other. To avoid this, we add a small positive regularisation term ϵI , where I is the identity matrix. The Mahalanobis distance is then defined as

$$h(x, p, P) = \sqrt{(x - p)^\top (\Sigma_p + \epsilon I)^{-1} (x - p)}. \quad (3.39)$$

Geometrically, this GDF defines a generalised Čech filtration where the sublevel sets $f_p^{-1}(-\infty, a]$ are unions of ellipsoids centered at p with each ellipsoid having axes aligned with the eigenvectors of $\Sigma_p + \epsilon I$ and lengths proportional to the square roots of the eigenvalues of $\Sigma_p + \epsilon I$. The regularisation term ensures that the ellipsoids are not degenerate. The standard Čech filtration can be seen as a special case of this GDF with $(\Sigma_p + \epsilon I)$ replaced by the identity matrix.

We prove two theorems that guarantee stability of such GDFs.

Theorem 3.13 Let $h : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$ 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 $\mathcal{P}(\mathbb{R}^d)$):

$$|h(x, p_1, P) - h(x, p_2, P)| \leq c \cdot d(p_1, p_2), \quad (3.40)$$

$$|h(x, p, P_1) - h(x, p, P_2)| \leq d \cdot d_H(P_1, P_2). \quad (3.41)$$

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

$$|h(x, p, P) - h(x, q', Q)| \quad (3.42)$$

$$\leq |h(x, p, P) - h(x, p, Q)| + |h(x, p, Q) - h(x, q', Q)| \quad (3.43)$$

$$\leq d \cdot d_H(P, Q) + c \cdot d(p, q'), \quad (3.44)$$

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

$$f(Q, x) - f(P, x) \leq h(x, q', Q) - f(P, x) \quad (3.45)$$

$$\leq h(x, q', Q) - h(x, p, P) \quad (3.46)$$

$$\leq (c + d) \cdot d_H(P, Q). \quad (3.47)$$

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

$$|f(P, x) - f(Q, x)| \leq (c + d) \cdot d_H(P, Q). \quad (3.48)$$

By Lemma 3.6, this implies that f is $(c + d)$ -PC Lipschitz. \square

Using this theorem, we show in Appendix B that the GDF from Example 3.12 is stable.

Theorem 3.14 *Let $h : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$ 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.:*

$$|h(x, p_1, P) - h(x, p_2, P)| \leq c \cdot d(p_1, p_2), \quad (3.49)$$

$$|h(x, p, P \cup \{p_1\}) - h(x, p, P \cup \{p_2\})| \leq d \cdot d(p_1, p_2) \quad (3.50)$$

$$|h(x, p, \{p_1\}) - h(x, p, \{p_2\})| \leq d \cdot d(p_1, p_2). \quad (3.51)$$

Then the generalised 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

$$|h(x, p, \{p_i\}_{i=1}^n) - h(x, p, \{q_i\}_{i=1}^n)| \quad (3.52)$$

$$= |h(x, p, \{p_i\}_{i=1}^{n-1} \cup \{p_n\}) - h(x, p, \{p_i\}_{i=1}^{n-1} \cup \{q_n\})| \quad (3.53)$$

$$\leq |h(x, p, \{p_i\}_{i=1}^{n-1}) - h(x, p, \{q_i\}_{i=1}^{n-1})| + d \cdot d_H(P, Q) \quad (3.54)$$

$$\leq |h(x, p, \{p_i\}_{i=1}^{n-2}) - h(x, p, \{q_i\}_{i=1}^{n-2})| + 2 \cdot d \cdot d_H(P, Q) \quad (3.55)$$

$$\leq \dots \quad (3.56)$$

$$\leq |P| \cdot d \cdot d_H(P, Q). \quad (3.57)$$

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

$$|h(x, p, P) - h(x, p, Q)| \leq d \cdot \max(|P|, |Q|) \cdot d_H(P, Q). \quad (3.58)$$

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

$$|h(x, p, P) - h(x, q', Q)| \quad (3.59)$$

$$\leq |h(x, p, P) - h(x, p, Q)| + |h(x, p, Q) - h(x, q', Q)| \quad (3.60)$$

$$\leq d \cdot \max(|P|, |Q|) \cdot d_H(P, Q) + c \cdot d_H(P, Q). \quad (3.61)$$

□

The conditions on the kernel for the induced persistence modules to be q -tame are exactly the same as the ones in Lemma 3.11.

3.6 Morse functions

The PC-Lipschitz condition is a strong global condition, which may not always hold for some stable GDFs.

Example 3.15 Consider a Čech-like GDF

$$f_P(x) = \min_{p \in P} d(x, p)^2, \quad (3.62)$$

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 P .

Example 3.16 Consider a GDF of the form

$$f_P(x) = \|x - g(P)\|, \quad (3.63)$$

where $g : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ is an arbitrary function of the point cloud P . This GDF is not Lipschitz continuous for many choices of g , but it is stable for any g , as it only has a single critical value at zero, which results in a single point $(0, \infty)$ in the 0-dimensional persistence diagram and no points in higher dimensions.

Motivated by the two examples above, we can relax the PC-Lipschitz condition by restricting ourselves to *Morse functions*, that is, smooth functions $f_P : \mathbb{R}^d \rightarrow \mathbb{R}$ whose critical points $x \in \text{Crit}(f_P)$ (where the gradient vanishes) are non-degenerate, i.e., the Hessian matrix $H(f_P)(x)$ is non-singular [21].

Theorem 3.17 Let $f : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$ 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, and all sublevel sets $f_P^{-1}(-\infty, a]$ and $f_Q^{-1}(-\infty, a]$ are compact. Furthermore, let there exist a bijection $\pi : \text{Crit}(f_P) \rightarrow \text{Crit}(f_Q)$ of critical points of f_P and f_Q such that the index of x is equal to the index of $\pi(x)$ for all $x \in \text{Crit}(f_P) \cup \text{Crit}(f_Q)$, and the following condition holds:

$$\forall x \in \text{Crit}(f_P) : |f_P(x) - f_Q(\pi(x))| \leq c \cdot d_H(P, Q). \quad (3.64)$$

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 $f_P^{-1}(-\infty, a]$ ([23], Theorem 3.5 and the remark following it). Let us denote such a CW-complex by $C_P(a)$. The previous statement can be written concisely as

$$\forall a \in \mathbb{R} \quad C_P(a) \simeq f_P^{-1}(-\infty, a]. \quad (3.65)$$

Let $d = c \cdot d_H(P, Q)$. We now show that $C_P(a)$ as a set is included in $C_Q(a + d)$. Consider a cell e^λ in $C_P(a)$. This cell corresponds to a critical point $x \in \text{Crit}(f_P)$ of index λ such that $f_P(x) \leq a$. By the condition (3.64), we have

$$f_Q(\pi(x)) \leq f_P(x) + d \leq a + d. \quad (3.66)$$

Thus, we have a critical point $\pi(x) \in \text{Crit}(f_Q)$ also of index λ such that $f_Q(\pi(x)) \leq a + d$. This means that the cell e^λ in $C_P(a)$ corresponds to a cell of the same dimensionality in $C_Q(a + d)$. Since this holds for any cell in $C_P(a)$, we have

$$C_P(a) \subseteq C_Q(a + d). \quad (3.67)$$

By the same reasoning, we can show that $C_Q(a)$ is included in $C_P(a + d)$. We also trivially have that $C_P(a) \subseteq C_P(a')$ for any $a' \geq a$ and the same for C_Q .

Similarly to the proof of stability of the Čech filtration using interleaving, the following diagram commutes up to homotopy, as all maps are either inclusions or homotopies:

$$\begin{array}{ccccc} f_P^{-1}(-\infty, a] & \hookrightarrow & f_P^{-1}(-\infty, a + d] & \hookrightarrow & f_P^{-1}(-\infty, a + 2d] \\ \uparrow \simeq & & \uparrow \simeq & & \uparrow \simeq \\ C_P(a) & \hookrightarrow & C_P(a + d) & \hookrightarrow & C_P(a + 2d) \\ & \searrow & \nearrow & \searrow & \nearrow \\ C_Q(a) & \hookrightarrow & C_Q(a + d) & \hookrightarrow & C_Q(a + 2d) \\ \simeq \downarrow & & \simeq \downarrow & & \simeq \downarrow \\ f_Q^{-1}(-\infty, a] & \hookrightarrow & f_Q^{-1}(-\infty, a + d] & \hookrightarrow & f_Q^{-1}(-\infty, a + 2d] \end{array} \quad (3.68)$$

Thus we have

$$d_i(\{f_P^{-1}(-\infty, a]\}_a, \{f_Q^{-1}(-\infty, a]\}_a) \leq d, \quad (3.69)$$

and similarly to the proof of stability of the Čech filtration, we have that the bottleneck distance between the persistence diagrams is bounded by d from above. \square

It should be noted that this theorem is not a strict generalisation of previous theorems, as there are GDFs that are PC-Lipschitz but which are not Morse.

For conditions that guarantee q -tameness of induced persistence modules of Morse functions, we refer the reader to Theorem 8.1.4 and Corollary 8.1.12 of Maximilian Schmahl's PhD thesis [28].

Chapter 4

Operations on stable functions

Having identified several classes of stable generalized density functions (GDFs) in Chapter 3, we now turn to natural subsequent questions: *how does stability behave under common operations? If we combine stable GDFs, for instance, by addition or by taking their minimum, is the resulting GDF also stable? If so, can we bound its stability constant in terms of the constants of the original functions?* This chapter explores these questions, identifying conditions under which stability is preserved and providing counterexamples where it is not. Understanding these properties is crucial for constructing more complex stable GDFs from simpler ones and for gaining deeper insight into the structure of the space of stable functions.

4.1 Affine transformations

We begin with the simplest operations: adding a constant to a stable GDF and multiplying it by a constant.

Theorem 4.1 (Stability under constant addition) *Let $f(P, x)$ be a c -stable GDF with q -tame induced persistence modules. Then for any constant $z \in \mathbb{R}$ the GDF $g(P, x) = f(P, x) + z$ is also c -stable with q -tame induced persistence modules.*

Proof First we relate the sublevel sets of g_P and f_P . For any $a \in \mathbb{R}$:

$$g_P^{-1}(-\infty, a] = \{x \in \mathbb{R}^d \mid g_P(x) \leq a\} \quad (4.1)$$

$$= \{x \in \mathbb{R}^d \mid f_P(x) + z \leq a\} \quad (4.2)$$

$$= \{x \in \mathbb{R}^d \mid f_P(x) \leq a - z\} \quad (4.3)$$

$$= f_P^{-1}(-\infty, a - z]. \quad (4.4)$$

This means the sublevel set filtration $\{g_P^{-1}(-\infty, a]\}_{a \in \mathbb{R}}$ is identical to the filtration $\{f_P^{-1}(-\infty, b]\}_{b \in \mathbb{R}}$ but with the indices shifted by z , i.e., $b = a - z$.

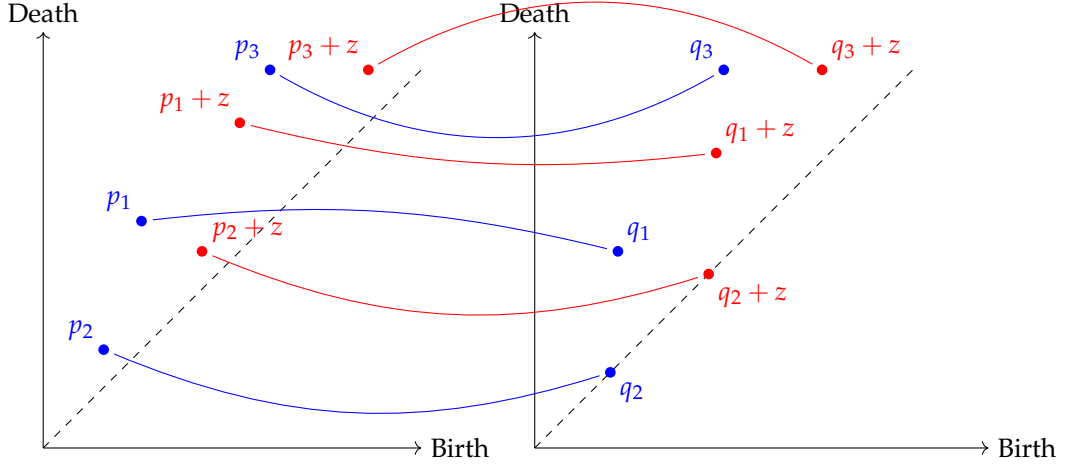


Figure 4.1: The persistence diagrams of f_P (left, blue), f_Q (right, blue), and g_P (left, red), g_Q (right, red) where $g_P = f_P + z$ and $g_Q = f_Q + z$. A matching between the diagrams of f_P and f_Q is shown in blue, and a matching between the diagrams of g_P and g_Q is shown in red.

Consequently, if a topological feature is born at b and dies at d in the filtration of f_P , a corresponding feature of the same dimensionality in the filtration of g_P will be born at $b + z$ and die at $d + z$, and vice versa. Thus, the persistence diagrams of f_P and g_P are related by a constant shift along both axes:

$$\text{Dgm}_k(g_P) = \{(b + z, d + z) \mid (b, d) \in \text{Dgm}_k(f_P)\}. \quad (4.5)$$

Next, we address q -tameness. Let $\mathbb{F} = (\{F_a\}, \{f_a^{a'}\})$ be a persistence module induced by f_P and $\mathbb{G} = (\{G_a\}, \{g_a^{a'}\})$ be that induced by g_P . From the relationship between the sublevel sets, we have $G_a = F_{a-z}$ and $g_a^{a'} = f_{a-z}^{a'-z}$. Since \mathbb{F} is q -tame by assumption, $f_a^{a'}$ is of finite rank for $a < a'$. This directly implies that $g_a^{a'}$ is also of finite rank for all $a < a'$, so \mathbb{G} is also q -tame.

Finally, we prove c -stability for g . Fix point clouds P and Q . By stability of f , we have that

$$d_b(\text{Dgm}_k(f_P), \text{Dgm}_k(f_Q)) \leq c \cdot d_H(P, Q). \quad (4.6)$$

Let π_i be a sequence of matchings between $\text{Dgm}_k(f_P) \cup \Delta$ and $\text{Dgm}_k(f_Q) \cup \Delta$ that achieves this bound. For each i , we can construct a matching π'_i between $\text{Dgm}_k(g_P) \cup \Delta$ and $\text{Dgm}_k(g_Q) \cup \Delta$ by shifting the points in π_i by z :

$$\pi'_i(b + z, d + z) = (b' + z, d' + z) \text{ where } (b', d') = \pi_i(b, d). \quad (4.7)$$

The cost of this matching is the same as that of π_i :

$$\|(b + z, d + z) - \pi'_i(b + z, d + z)\|_\infty = \|(b, d) - (b', d')\|_\infty, \quad (4.8)$$

thus π'_i is a matching between $\text{Dgm}_k(g_P)$ and $\text{Dgm}_k(g_Q)$ with the same cost as π_i . An example of this is shown in Figure 4.1. Therefore, we have:

$$d_b(\text{Dgm}_k(g_P), \text{Dgm}_k(g_Q)) \leq c \cdot d_H(P, Q). \quad (4.9)$$

Since this holds for arbitrary P and Q , g is c -stable. \square

Theorem 4.2 (Stability under positive constant multiplication) *Let $f(P, x)$ be a c -stable GDF with q -tame induced persistence modules. Then for any constant $\alpha \geq 0$, the function $g(P, x) = \alpha f(P, x)$ is a $(c\alpha)$ -stable GDF with q -tame induced persistence modules.*

Proof We first consider the case when $\alpha = 0$. In this case, $g(P, x) = 0$ for all $x \in \mathbb{R}^d$, which is 0-stable by Theorem 3.2. Its induced persistence modules are pointwise finite dimensional, hence q -tame.

Now, let $\alpha > 0$. The argument mirrors that of Theorem 4.1. The sublevel sets of g and of f are related by:

$$g_P^{-1}(-\infty, a] = \{x \in \mathbb{R}^d \mid \alpha f_P(x) \leq a\} = f_P^{-1}(-\infty, a/\alpha]. \quad (4.10)$$

This implies that the persistence module G induced by g_P is the same as the persistence module F induced by f_P up to reindexing where a is replaced with $a \cdot \alpha$, i.e. $F_a = G_{a \cdot \alpha}$ and the maps $f_a^{a'}$ and $g_{a \cdot \alpha}^{a' \cdot \alpha}$ are identical. The q -tameness of F implies the q -tameness of G .

A point $(b, d) \in \text{Dgm}_k(f_P)$ corresponds to $(\alpha b, \alpha d) \in \text{Dgm}_k(g_P)$ and vice versa. By the same argument as in the proof of Theorem 4.1, we can construct a matching $\pi'_i(b \cdot \alpha, d \cdot \alpha) = (b' \cdot \alpha, d' \cdot \alpha)$ where $(b', d') = \pi_i(b, d)$, which gives a bound on the bottleneck distance between the persistence diagrams of g_P and g_Q :

$$d_b(\text{Dgm}_k(g_P), \text{Dgm}_k(g_Q)) \leq c \cdot \alpha \cdot d_H(P, Q). \quad (4.11)$$

\square

We can extend this theorem to negative constants using the following lemma:

Lemma 4.3 (Stability under negation) *Let $f(P, x)$ be a c -stable GDF with q -tame induced persistence modules. Then the function $g(P, x) = -f(P, x)$ is also a c -stable GDF with q -tame induced persistence modules.*

Proof This result follows directly from the symmetry theorem by David Cohen-Steiner, Herbert Edelsbrunner, and John Harer [11]. The theorem states that for any real-valued function f on a d -dimensional manifold, the persistence diagram of f in dimension k is related to the persistence diagram of $-f$ in dimension $d - k$ via reflection across the minor diagonal (the line $b + d = 0$). More precisely, for every point $(b, d) \in \text{Dgm}_k(f)$, there corresponds a point $(-d, -b) \in \text{Dgm}_{d-k}(-f)$, and this correspondence is bijective.

Given that \mathbb{R}^d is a d -manifold, we can apply this theorem. Now let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be finite point clouds and f be a c -stable GDF with q -tame induced persistence modules. Let the bottleneck distance $d_b(\text{Dgm}_k(f_P), \text{Dgm}_k(f_Q))$ be achieved by a sequence of matchings π_i between the diagrams. We can construct a sequence of matchings π'_i between the diagrams $\text{Dgm}_{d-k}(f_P)$ and $\text{Dgm}_{d-k}(f_Q)$ by reflecting the points in π_i across the minor diagonal:

$$\pi'_i(-d, -b) = (-d', -b') \text{ where } \pi_i(b, d') = (b', d'). \quad (4.12)$$

By the symmetry theorem, this is a valid sequence of matchings and the cost of π'_i is the same as that of π_i , thus $-f$ is also c -stable.

q -tameness follows from the fact that the induced persistence modules of f and $-f$ are identical up to reindexing with the sign change. \square

Combining Theorem 4.2 and Lemma 4.3, we obtain the following corollary:

Corollary 4.4 (Stability under constant multiplication) *Let $f(P, x)$ be a c -stable GDF with q -tame induced persistence modules. Then for any constant $\alpha \in \mathbb{R}$, the function $g(P, x) = \alpha f(P, x)$ is also a $c|\alpha|$ -stable GDF with q -tame induced persistence modules.*

4.2 Addition of stable functions

Having examined affine transformations, we now explore the effect of adding two stable GDFs. Let f and g be GDFs with stability constants c_f and c_g . What can be said about the stability constant of their sum $h(P, x) = f(P, x) + g(P, x)$? A simple case offers initial intuition: if $g = f$, then $h = f + f = 2 \cdot f$. By Theorem 4.1, h is $(2 \cdot c_f)$ -stable. One would naturally expect that the stability constant of $h = f + g$ is then bounded by $c_f + c_g$, but this is not always the case, as the following example demonstrates.

Example 4.5 (Sum of stable GDFs may not be stable) *Consider GDFs on \mathbb{R} ($d = 1$). Fix a constant $a > 0$, let*

$$f(P, x) = x + a \cdot \min P \quad \text{and} \quad g(P, x) = -x + a \cdot \min P. \quad (4.13)$$

As described in Example 3.3, both GDFs are 0-stable.

The sum of f and g is:

$$h(P, x) = f(P, x) + g(P, x) = 2a \cdot \min P. \quad (4.14)$$

The function h_P is constant for a fixed P . Its 0-dimensional persistence diagram has exactly one point $(2a \cdot \min P, \infty)$, representing the single connected component \mathbb{R} itself, born at $2a \cdot \min P$ and never dying, as can be seen in Figure 4.2.

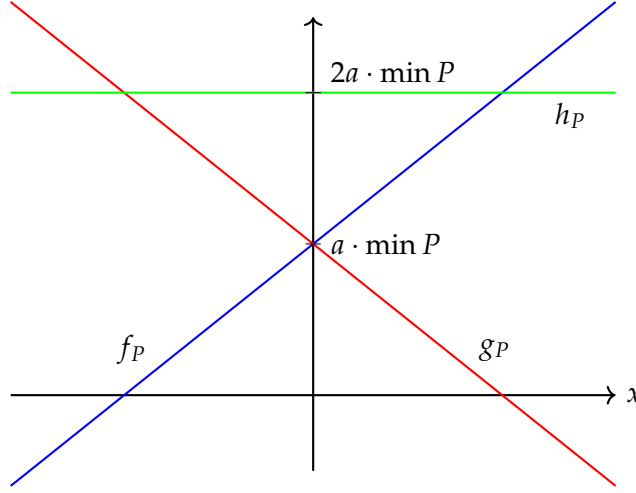


Figure 4.2: The GDFs f_P (red), g_P (blue) and h_P (green) for an arbitrary point cloud P .

Let $P = \{1\}, Q = \{2\}$ with $d_H(P, Q) = 1$. The 0-dimensional persistence diagrams of h are $\text{Dgm}_0(h_P) = \{(2a, \infty)\}$ and $\text{Dgm}_0(h_Q) = \{(4a, \infty)\}$. In the optimal matching π between these diagrams, the point $(2a, \infty)$ must be matched to $(4a, \infty)$, as matching either of them to the diagonal Δ would result in an infinite cost. The cost of π is thus:

$$\|(2a, \infty) - (4a, \infty)\|_\infty = 2a. \quad (4.15)$$

Therefore, the bottleneck distance between the diagrams is $2a$, and for h to be c -stable, we need:

$$2a \leq c \cdot d_H(P, Q) = c. \quad (4.16)$$

This means that for any $c \in [0, \infty)$, we can choose $a = c$ such that the stability constant of h is larger than c . Thus, the sum of two 0-stable GDFs is not necessarily c -stable for any $c \in [0, \infty)$.

Despite this counterexample for general stable GDFs, stability under addition holds for the more restrictive class of PC-Lipschitz GDFs.

Theorem 4.6 (Sum of PC-Lipschitz GDFs) *Let $f(P, x)$ and $g(P, x)$ be a c_f -PC-Lipschitz GDF and a c_g -PC-Lipschitz GDF. Then their sum $h(P, x) = f(P, x) + g(P, x)$ is a $(c_f + c_g)$ -PC-Lipschitz GDF and is thus $(c_f + c_g)$ -stable if the induced persistence modules of h are q -tame.*

Proof Let $P, Q \subseteq \mathbb{R}^d$ be finite point clouds. By subadditivity of the ∞ -norm

and the definition of PC-Lipschitz continuity, we have:

$$\|h_P - h_Q\|_\infty = \|(f_P + g_P) - (f_Q + g_Q)\|_\infty \quad (4.17)$$

$$\leq \|f_P - f_Q\|_\infty + \|g_P - g_Q\|_\infty \quad (4.18)$$

$$\leq c_f \cdot d_H(P, Q) + c_g \cdot d_H(P, Q) \quad (4.19)$$

$$= (c_f + c_g) \cdot d_H(P, Q), \quad (4.20)$$

which shows that h is $(c_f + c_g)$ PC-Lipschitz. By Lemma 3.6, h is $(c_f + c_g)$ -stable if its induced persistence modules are q -tame. \square

4.3 Minimum and maximum of stable functions

Similar to addition, the pointwise minimum of two PC-Lipschitz GDFs is also PC-Lipschitz, and thus stable.

Theorem 4.7 (Minimum of PC-Lipschitz GDFs) *Let $f(P, x)$ and $g(P, x)$ be a c_f -PC-Lipschitz GDF and a c_g -PC-Lipschitz GDF. Then the function $h(P, x) = \min\{f(P, x), g(P, x)\}$ is a $\max\{c_f, c_g\}$ -PC-Lipschitz GDF and is thus $\max\{c_f, c_g\}$ -stable if the induced persistence modules of h are q -tame.*

Proof This result follows directly from a standard property of Lipschitz functions: the pointwise minimum (or maximum) of two Lipschitz functions $F_1, F_2 : X \rightarrow Y$ with Lipschitz constants L_1, L_2 is itself Lipschitz with constant $\max\{L_1, L_2\}$ ([30], Proposition 1.5.5). \square

However, this property does not hold for general stable GDFs, as shown in the following counterexample.

Example 4.8 (Minimum of 0-stable GDFs can be unstable) *Consider GDFs on \mathbb{R} ($d = 1$). Fix a constant $a \in \mathbb{R}$ and let*

$$f(P, x) = a \cdot |x - \min p| \quad \text{and} \quad g(P, x) = a \cdot |x - \max p|. \quad (4.21)$$

The 0-dimensional persistence diagrams for these GDFs are $\text{Dgm}_0(f_P) = \{(0, \infty)\}$ and $\text{Dgm}_0(g_P) = \{(0, \infty)\}$ for any P , and thus are both 0-stable.

Consider $P = \{-1, 1\}$ and $Q = \{0\}$ with $d_H(P, Q) = 1$. As we can see in Figure 4.3, the 0-dimensional persistence diagrams of h_P and h_Q are

$$\text{Dgm}_0(h_P) = \{(0, a), (0, \infty)\}, \quad (4.22)$$

$$\text{Dgm}_0(h_Q) = \{(0, \infty)\}. \quad (4.23)$$

The cost of matching $(0, a)$ to Δ is $\frac{a}{2}$, so the bottleneck distance between the diagrams is $\frac{a}{2}$. Thus, for every $c \in [0, \infty)$ we can choose $a = 2c$ such that the stability constant of $h(P, x) = f(P, x) + g(P, x)$ is c . Thus, the minimum of two 0-stable GDFs is not necessarily c -stable for any $c \in [0, \infty)$.

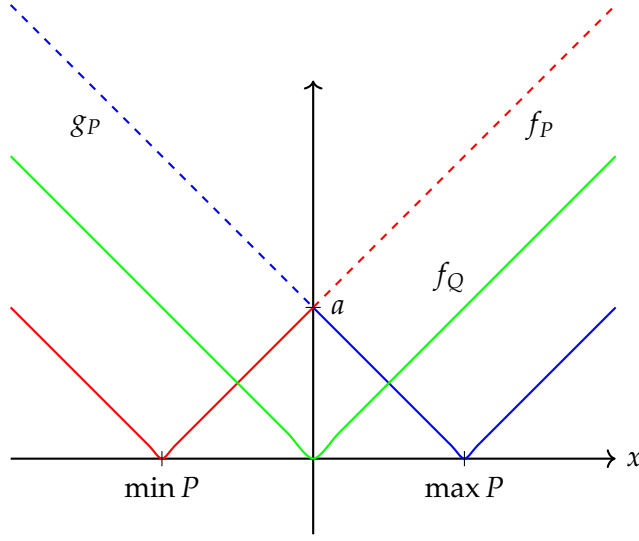


Figure 4.3: The GDFs f_P (red), g_P (blue) and f_Q (green) for point clouds $P = \{-1, 1\}$ and $Q = \{0\}$. The minimum h_P of f_P and g_P is shown in solid red and green lines.

This example shows that stability alone is not sufficient to guarantee that the minimum of two GDFs is stable. The issue arose because the topological changes induced by perturbing the point cloud were different for the two functions, leading to an unstable interaction in their minimum. This suggests that for stability to be preserved, the interleaving maps of the two functions must be compatible. Building on this intuition, we can establish conditions for stability that are not reliant on the strong PC-Lipschitz property.

Theorem 4.9 *Let f and g be c_f -stable and c_g -stable GDFs, respectively. Let the sublevel sets of f and g be denoted as $f(P)_a = f_P^{-1}(-\infty, a]$ and $g(P)_a = g_P^{-1}(-\infty, a]$.*

Assume that for any finite point clouds $P, Q \in \mathcal{P}(X)$:

1. *The induced persistence modules \mathbb{F}_P and \mathbb{G}_P are q -tame.*
2. *The functions f_P, g_P are continuous.*
3. *The stability of f and g is realised by interleavings of their sublevel set filtrations, i.e.:*

$$c_f \cdot d_H(P, Q) = d_i(\mathbb{F}_P, \mathbb{F}_Q) = d_i(\{f(P)_a\}, \{f(Q)_a\}) \quad (4.24)$$

$$c_g \cdot d_H(P, Q) = d_i(\mathbb{G}_P, \mathbb{G}_Q) = d_i(\{g(P)_a\}, \{g(Q)_a\}). \quad (4.25)$$

4. OPERATIONS ON STABLE FUNCTIONS

Let these interleavings be realised by maps

$$\varphi_a^f : f(P)_a \rightarrow f(Q)_{a+c_f \cdot d_H(P,Q)} \quad (4.26)$$

$$\psi_a^f : f(Q)_a \rightarrow f(P)_{a+c_f \cdot d_H(P,Q)} \quad (4.27)$$

$$\varphi_a^g : g(P)_a \rightarrow g(Q)_{a+c_g \cdot d_H(P,Q)} \quad (4.28)$$

$$\psi_a^g : g(Q)_a \rightarrow g(P)_{a+c_g \cdot d_H(P,Q)}. \quad (4.29)$$

4. For all $x \in \mathbb{R}^d$ such that $f(P, x) = g(P, x)$, we also have that $\varphi_a^f(x) = \varphi_a^g(x)$ and $\psi_a^f(x) = \psi_a^g(x)$.

5. For all $x \in \mathbb{R}^d$ such that $f(P, x) \leq g(P, x)$, we also have that $f(Q, x) \leq g(Q, x)$.

Then the function $h(P, x) = \min\{f(P, x), g(P, x)\}$ is a $\max\{c_f, c_g\}$ -stable GDF if the induced persistence modules of h are q -tame.

Proof Fix $P, Q \in \mathcal{P}(X)$. Let $\varepsilon_f = c_f \cdot d_H(P, Q)$ be the interleaving distance between the persistence modules \mathbb{F}_P and \mathbb{F}_Q , and similarly $\varepsilon_g = c_g \cdot d_H(P, Q)$. We want to define ζ_a and η_a such that the following diagrams commute, where $\varepsilon_h := \max\{c_f, c_g\} \cdot d_H(P, Q)$:

$$\begin{array}{ccc} h(P)_a & \xrightarrow{\quad} & h(P)_{a'} \\ & \searrow \zeta_a & \searrow \zeta_{a'} \\ & h(Q)_{a+\varepsilon} & \xrightarrow{\quad} h(Q)_{a'+\varepsilon} \end{array} \quad (4.30)$$

$$\begin{array}{ccc} & h(P)_{a+\varepsilon} & \xrightarrow{\quad} h(P)_{a'+\varepsilon} \\ \nearrow \eta_a & & \nearrow \eta_{a'} \\ h(Q)_a & \xrightarrow{\quad} & h(Q)_{a'} \end{array} \quad (4.31)$$

$$\begin{array}{ccc} h(P)_a & \xrightarrow{\quad} & h(P)_{a+2\varepsilon} \\ & \searrow \zeta_a & \nearrow \eta_{a+\varepsilon} \\ & h(Q)_{a+\varepsilon} & \end{array} \quad (4.32)$$

$$\begin{array}{ccc} & h(P)_{a+\varepsilon} & \\ \nearrow \eta_a & & \searrow \zeta_{a+\varepsilon} \\ h(Q)_a & \xrightarrow{\quad} & h(Q)_{a+2\varepsilon} \end{array} \quad (4.33)$$

Let us define the maps ζ_a and η_a as follows:

$$\zeta_a(x) = \begin{cases} \varphi_a^f(x) & \text{if } f(P, x) \leq g(P, x) \\ \varphi_a^g(x) & \text{otherwise} \end{cases} \quad (4.34)$$

$$\eta_a(x) = \begin{cases} \psi_a^f(x) & \text{if } f(Q, x) \leq g(Q, x) \\ \psi_a^g(x) & \text{otherwise} \end{cases} \quad (4.35)$$

These maps are well-defined, as for any $x \in H(P)_a$, either $f(P, x) \leq g(P, x)$ or $f(P, x) > g(P, x)$. In the first case, x is also in $f(P)_a$ and thus $\varphi_a^f(x)$ is well-defined. Similarly, in the second case, x is in $g(P)_a$ and $\varphi_a^g(x)$ is well-defined. The same holds for η_a .

For these maps to induce homomorphisms of homology groups, we need to ensure they are continuous [16].

Let

$$A = \{f(P, x) \leq g(P, x) \mid x \in h(P)_a\} \quad (4.36)$$

and

$$B = \{f(P, x) \geq g(P, x) \mid x \in h(P)_a\}. \quad (4.37)$$

Both of these sets are closed in $h(P)_a$ endowed with the subspace topology because we can define

$$k(x) := g(P, x) - f(P, x), \quad (4.38)$$

which is continuous as the difference of continuous functions. We then have:

$$A = h(P)_a \cap k^{-1}((-\infty, 0]) \quad (4.39)$$

$$B = h(P)_a \cap k^{-1}([0, \infty)), \quad (4.40)$$

and both preimages are closed sets in \mathbb{R}^d as preimages of closed sets under a continuous function.

Now we consider the restriction $\zeta_a|_A$. As $f(P, x) \leq g(P, x)$ for all $x \in A$, we have $\zeta_a(x) = \varphi_a^f(x)$ for all $x \in A$, so $h_P|_A = \varphi_a^f|_A$.

The case of $\zeta_a|_B$ is slightly different, as we need to separately consider the case of x being such that $f(P, x) = g(P, x)$. By condition 4, we have that $\varphi_a^f(x) = \varphi_a^g(x)$ for such x , and thus we have $\zeta_a|_B = \varphi_a^g|_B$.

Consequently, by the pasting lemma ([25], Theorem 18.3) ζ_a is continuous on $A \cup B = h(P)_a$. Applying the same reasoning to η_a , we conclude that η_a is continuous on $h(Q)_a$.

Now we prove commutativity of diagrams (4.30)-(4.33). We first examine the diagram (4.30). Let $x \in H(P)_a$. Suppose $f(P, x) \leq g(P, x)$ and thus

$\zeta_a(x) = \varphi_a^f(x)$. Then $\zeta_{a'}(x)$ is defined as $\varphi_{a'}^f(x)$, which is equal to $\varphi_a^f(x)$. The same follows when $f(P, x) > g(P, x)$. Thus, the diagram commutes and by the same reasoning, the diagram (4.31) also commutes.

Next, we consider the triangle diagrams (4.32) and (4.33). Let $x \in H(P)_a$. If $f(P, x) \leq g(P, x)$, then $\zeta_a(x) = \varphi_a^f(x)$. By condition 5, we also have that $f(Q, x) \leq g(Q, x)$, so $\eta_{a+\varepsilon}(x) = \psi_{a+\varepsilon}^f(x)$. As φ_a^f and $\psi_{a+\varepsilon}^f$ are interleaving maps for f ([29], Theorem 6.17), they commute with the inclusion map, and thus the triangle (4.32) commutes for the case when $f(P, x) \leq g(P, x)$.

The same reasoning applies when $f(P, x) > g(P, x)$, and thus the triangle (4.32) commutes in all cases. By the same reasoning, the triangle (4.33) also commutes.

As all the diagrams commute, we can conclude that

$$d_i(\{h(P)_a\}, \{h(Q)_a\}) \leq \max\{c_f, c_g\} \cdot d_H(P, Q), \quad (4.41)$$

which by Lemma 2.21 implies that

$$d_i(\mathbb{H}_P, \mathbb{H}_Q) \leq d_i(\{h(P)_a\}, \{h(Q)_a\}) \leq \max\{c_f, c_g\} \cdot d_H(P, Q), \quad (4.42)$$

and as we assume \mathbb{H}_P and \mathbb{H}_Q to be q -tame, we have that \mathbb{H}_P is $\max\{c_f, c_g\}$ -stable. \square

We also identify conditions under which the pointwise minimum of two GDFs induces q -tame persistence modules.

Lemma 4.10 *If for every $P \in \mathcal{P}(\mathbb{R}^d)$ the functions f_P and g_P are continuous, bounded below and proper, then their pointwise minimum h_P is also continuous, bounded below and proper. Consequently, by Theorem 2.7, the induced persistence modules of the pointwise minimum h_P are q -tame.*

Proof Same argument as in the proof of Lemma 3.11. \square

The results in this section concerning the pointwise minimum can be extended to the pointwise maximum. This is because $\max(A, B) = -\min(-A, -B)$, and negation preserves stability by Lemma 4.3.

Chapter 5

The space of stable functions

In the previous chapters, we identified several classes of stable generalized density functions and explored how they can be combined to form new stable functions. This naturally leads to considering the collections of these functions as algebraic structures and topological spaces. Let \mathcal{S}_c denote the set of all c -stable GDFs and \mathcal{L}_c denote the set of all c -PC-Lipschitz GDFs. We have established that $\mathcal{L}_c \subseteq \mathcal{S}_c$ (Lemma 3.6) and that this inclusion is proper (Example 3.3).

This chapter investigates the algebraic and topological structure of these sets. We will first explore \mathcal{L}_c , the more structured of the two spaces. We will then turn our attention to the broader space \mathcal{S}_c and conclude by examining the relationship between them.

5.1 The space of c -PC-Lipschitz functions

The set \mathcal{L}_c exhibits several structural properties.

Theorem 5.1 *The set \mathcal{L}_c forms a distributive lattice with respect to the operations of pointwise minimum and pointwise maximum, that is, the minimum and maximum of two c -stable GDFs is also a c -stable GDF and the following laws hold:*

$$\max(f, \min(f, g)) = f \quad (5.1)$$

$$\min(f, \max(f, g)) = f \quad (5.2)$$

$$\max(f, \min(g, h)) = \min(\max(f, g), \max(f, h)) \quad (5.3)$$

$$\min(f, \max(g, h)) = \max(\min(f, g), \min(f, h)). \quad (5.4)$$

This lattice is bounded neither from above nor from below.

Proof 1. Both minimum and maximum are closed in \mathcal{L}_c by Theorem 4.7. The absorption laws follow from the fact that they hold for (\mathbb{R}, \min, \max) .

2. Distributivity follows from the fact that it holds for (\mathbb{R}, \min, \max) .
3. To show that the lattice is not bounded, suppose f is a top element. $f + 1$ is also in \mathcal{L}_c by Theorem 4.1 and $\max(f, f + 1) = f + 1 \neq f$, thus f cannot be a top element.
4. We can similarly show that there is no bottom element by considering $f - 1$. \square

While \mathcal{L}_c has a lattice structure, it does not form a vector space. As shown in Theorem 4.6, the sum of two functions in \mathcal{L}_c lies in \mathcal{L}_{2c} , but not necessarily in \mathcal{L}_c . This lack of closure under addition prevents \mathcal{L}_c from being a vector space. Using an alternative operation such as averaging, fails to satisfy the axioms for scalar multiplication.

On the other hand, the space of all PC-Lipschitz GDFs $\mathcal{L} = \bigcup_{c \in \mathbb{R}^+} \mathcal{L}_c$ does form a vector space. However, this space is of less interest for the purposes of TDA, as practical applications do not concern themselves with merely stable functions, but rather with c -stable functions for some fixed c . For a comprehensive treatment of general Lipschitz function spaces, we refer the interested reader to the work of N. Weaver [30].

We now turn to topological properties of \mathcal{L}_c .

Theorem 5.2 *The set \mathcal{L}_c is convex.*

Proof Let $f, g \in \mathcal{L}_c$ and let $\alpha \in [0, 1]$. We need to show that the convex combination $\alpha f + (1 - \alpha)g$ is also in \mathcal{L}_c .

Since $\alpha \geq 0$, by Theorem 4.2, we have $\alpha f \in \mathcal{L}_{\alpha c}$. Similarly, as $(1 - \alpha) \geq 0$, $(1 - \alpha)g \in \mathcal{L}_{(1-\alpha)c}$.

Finally, by Theorem 4.6,

$$(\alpha f + (1 - \alpha)g) \in \mathcal{L}_{\alpha c + (1-\alpha)c} = \mathcal{L}_c. \quad (5.5)$$

\square

Corollary 5.3 *Let \mathcal{F} be a topological vector space of functions containing \mathcal{L}_c . Then \mathcal{L}_c is a contractible subset of \mathcal{F} [25].*

Theorem 5.4 *The set \mathcal{L}_c is closed in the topology of pointwise convergence.*

Proof Let $\{f^n\}_{n=1}^\infty$ be a sequence of functions in \mathcal{L}_c that converges pointwise to a function f . We need to show that f is also in \mathcal{L}_c . We have:

$$\|f_P - f_Q\|_\infty \leq \|f_P - f_P^n\|_\infty + \|f_P^n - f_Q^n\|_\infty + \|f_Q^n - f_Q\|_\infty \quad (5.6)$$

$$\leq \|f_P - f_P^n\|_\infty + c \cdot d_H(P, Q) + \|f_Q^n - f_Q\|_\infty, \quad (5.7)$$

which converges to $c \cdot d_H(P, Q)$ and thus f is c -PC-Lipschitz. \square

Corollary 5.5 *Since uniform convergence and compact convergence are stronger topologies than pointwise convergence, \mathcal{L}_c is also closed in these topologies.*

5.2 The space of c -stable functions

We now turn our attention to the larger space of all c -stable GDFs \mathcal{S}_c . As shown in various examples in Chapter 4, \mathcal{S}_c lacks the well-behaved structure of \mathcal{L}_c : it is not closed under addition, pointwise maximum or pointwise minimum.

Despite these limitations, \mathcal{S}_c still possesses important topological properties.

Theorem 5.6 *The singleton set of the function $g(P, x) = 0$ is a strong deformation retract of \mathcal{S}_c in any topology on \mathcal{F} where scalar multiplication is continuous.*

Proof Let $f \in \mathcal{S}_c$. We can define a homotopy $H_t(f)$ for $t \in [0, 1]$ as

$$H_t(f)(P, x) := (1 - t)f(P, x) \quad (5.8)$$

Then $H_0(f) = f$ and $H_1(f) = g$. As $H_t(g) = g$, this deformation retraction is strong. \square

Corollary 5.7 *The set $\mathcal{S}_c \subseteq \mathcal{F}$ is contractible in any topology on \mathcal{F} where scalar multiplication is continuous.*

The path constructed above shows that we can connect any function in \mathcal{S}_c to the zero function by a continuous path, provided scalar multiplication is continuous. We can join two such paths to connect any two functions in \mathcal{S}_c :

Theorem 5.8 *The set $\mathcal{S}_c \subseteq \mathcal{F}$ is path-connected in any topology on \mathcal{F} where scalar multiplication is continuous.*

Proof Let $f, g \in \mathcal{S}_c$. Let $h_t(P, x)$ be defined for $t \in [0, 1]$ as

$$h_t(P, x) := \begin{cases} (1 - 2t)f(P, x), & t \in [0, 0.5] \\ (2t - 1)g(P, x), & t \in (0.5, 1] \end{cases} \quad (5.9)$$

This path starts at $h(0) = f$, passes through the zero function at $h(0.5) = 0$, and ends at $h(1) = g$.

We now verify that $h_t(P, x) \in \mathcal{S}_c$ for all $t \in [0, 1]$. For $t \in [0, 0.5]$, the scaling factor is $1 - 2t$, which lies in $[0, 1]$, and by Theorem 4.2, h_t is in \mathcal{S}_c for $t \in [0, 0.5]$. Similarly, for $t \in (0.5, 1]$, the scaling factor is $2t - 1$, which also lies in $[0, 1]$ and thus the whole path h_t is in \mathcal{S}_c . \square

Theorem 5.9 *The set \mathcal{S}_c is closed in the topology of uniform convergence.*

Proof Let $\{f^n\}_{n=1}^\infty$ be a sequence of functions in \mathcal{S}_c that converges uniformly to a function f , that is, $\|f - f^n\|_\infty \rightarrow 0$. We need to show that f is also in \mathcal{S}_c .

For any finite sets $P, Q \subseteq \mathbb{R}$, by the triangle inequality, we have:

$$\begin{aligned} d_b(\text{Dgm}_k(f_P), \text{Dgm}_k(f_Q)) &\leq d_b(\text{Dgm}_k(f_P), \text{Dgm}_k(f_P^n)) + \\ &\quad d_b(\text{Dgm}_k(f_P^n), \text{Dgm}_k(f_Q^n)) + \\ &\quad d_b(\text{Dgm}_k(f_Q^n), \text{Dgm}_k(f_Q)) \end{aligned} \tag{5.10}$$

$$\leq \|f_P - f_P^n\|_\infty + c \cdot d_H(P, Q) + \|f_Q^n - f_Q\|_\infty, \tag{5.11}$$

which converges to $c \cdot d_H(P, Q)$ as $n \rightarrow \infty$. Thus, f is c -stable. \square

5.3 The relationship between \mathcal{S}_c and \mathcal{L}_c

As \mathcal{L}_c is much more well-behaved than \mathcal{S}_c , it would be desirable to have \mathcal{L}_c dense in \mathcal{S}_c . However, as \mathcal{L}_c is closed in the topologies of uniform convergence, of compact convergence and of pointwise convergence, \mathcal{L}_c is unfortunately not dense in \mathcal{S}_c in these topologies.

A possible direction for further investigation is to consider the notion of *shyness* [26], which generalises the concept of measure zero to infinite-dimensional spaces. This concept allows us to ask: is the “usual” stable function also PC-Lipschitz?

Conclusion

In this thesis, we explored a fundamental question in topological data analysis: *when are the results we get reliable?* Specifically, we looked at the stability of generalised density functions. If a small change in our input data leads to a large change in the resulting persistence diagram, we can't be confident in our analysis. Our main goal was to find and describe classes of GDFs that don't have this problem, functions that are stable against perturbations of the input point cloud.

6.1 Summary of findings

Our main contributions fall into three main areas:

1. *Identifying stable functions:* We have found several families of stable GDFs.
 - a) *PC-Lipschitz GDFs:* We showed that Lipschitz continuity with respect to the point cloud is a powerful sufficient condition for stability. This class of functions is well-behaved and easy to work with.
 - b) *Generalised Čech Filtrations:* We extended known stability for the nearest-neighbour distance function to more flexible versions, including:
 - i. Functions of the form $f(P, x) = \min_{p \in P} h(d(x, p))$ for monotone Lipschitz h .
 - ii. More general kernels $f(P, x) = \min_{p \in P} h(x, p)$ where h is Lipschitz in its second argument.
 - iii. Point cloud dependent kernels $f(P, x) = \min_{p \in P} h(x, p, P)$, as long as they satisfy certain Lipschitz conditions.

- c) *Morse Functions*: For Morse functions, we found that stability doesn't require a global property like being PC-Lipschitz. Instead, it depends on how the function's critical points behave.
- 2. *Operations on stable functions*: We investigated how stability is preserved under common operations:
 - a) For PC-Lipschitz GDFs, stability is preserved by affine transformations, addition, and taking the pointwise minimum or maximum.
 - b) However, for the broader class of all stable GDFs, we showed this isn't always true, although stability is still preserved by affine transformations. We provided examples where combining two stable functions results in an unstable one.
- 3. *The structure of the function spaces*: We studied the spaces of stable and PC-Lipschitz functions, showing that:
 - a) PC-Lipschitz GDFs form an unbounded distributive lattice under pointwise minimum and maximum.
 - b) The space of PC-Lipschitz GDFs is convex and closed in various function space topologies.
 - c) The space of all stable GDFs is path-connected and contractible, as well as closed in the topology of pointwise convergence.
 - d) Interestingly, the PC-Lipschitz functions are not dense within the space of all stable functions, meaning there are stable functions that cannot be approximated by these nicer PC-Lipschitz ones.

Together, these results help build a more solid theoretical foundation for using GDFs in practice, connecting and extending previous research.

6.2 Open questions and future work

This work also opens up several interesting questions for future research:

1. *What makes a function stable?* We found several conditions that are sufficient for stability, but we still lack necessary conditions. Fully characterizing the space of all stable functions remains an open problem.
2. *How common are PC-Lipschitz functions?*: While the PC-Lipschitz condition is useful, how restrictive is it? We could use tools like prevalence or shyness to understand how common PC-Lipschitz functions are among all stable functions.
3. *Non-Euclidean spaces*: Our work focused on data in standard Euclidean space, but TDA is applicable for other kinds of data as well. Extending stability results to those settings is an important next step.

4. *Wasserstein stability*: We measured stability using the bottleneck distance. It would be valuable to see how these results change when using other metrics between persistence diagrams, like the Wasserstein distance.

By mapping out the properties of stable GDFs, this work offers both a practical guide for researchers applying TDA and a theoretical foundation for those developing new methods.

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 $g : P \rightarrow \mathbb{R}_{\geq 0}$, and a parameter $q \in [1, \infty)$, the weighted Čech filtration $V_q^a[P, g]$ is equal to the sublevel set of the generalised density function*

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

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

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

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

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

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

The proposed GDF takes value

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

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

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

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

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

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

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

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

The proposed GDF takes value

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

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

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

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

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

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

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

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

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

Appendix B

The generalised Čech complex using the Mahalanobis distance kernel is stable

Lemma B.1 *The generalised Čech complex using the Mahalanobis distance kernel*

$$h(x, p, P) = \sqrt{(x - p)^\top (\Sigma(p, P) + \varepsilon I)^{-1} (x - p)} = \|A(p, P)(x - p)\|_2, \quad (\text{B.1})$$

where $A(p, P) = (\Sigma(p, P) + \varepsilon I)^{-1/2}$ is stable.

Proof To use Theorem 3.13, we need to bound $|h(x, p, P) - h(x, q, P)|$ and $|h(x, p, P) - h(x, p, Q)|$. Bounding the first term, we have:

$$\begin{aligned} & |h(x, p, P) - h(x, q, P)| \\ &= \left| \|A(p, P)(x - p)\|_2 - \|A(q, P)(x - q)\|_2 \right| \end{aligned} \quad (\text{B.2})$$

$$\leq \|A(p, P)(x - p) - A(q, P)(x - q)\|_2 \quad (\text{B.3})$$

$$= \|A(p, P)(x - p) - A(q, P)(x - q) + A(p, P)(x - q) - A(p, P)(x - q)\|_2 \quad (\text{B.4})$$

$$= \|A(p, P)(q - p) + (A(p, P) - A(q, P))(x - q)\|_2 \quad (\text{B.5})$$

$$\leq \|A(p, P)(q - p)\|_2 + \|(A(p, P) - A(q, P))(x - q)\|_2 \quad (\text{B.6})$$

$$\leq \|A(p, P)\|_{\text{op}} \cdot \|q - p\|_2 + \|A(p, P) - A(q, P)\|_{\text{op}} \cdot \|x - q\|_2. \quad (\text{B.7})$$

The operator norm $\|A(p, P)\|_{\text{op}} = \sup_{v: \|v\|_2=1} \|A(p, P)v\|_2$ is precisely the largest eigenvalue of the matrix $A(p, P)$. As $\Sigma(p, P)$ is positive semi-definite, adding it to εI cannot decrease the eigenvalues, and thus $\|\Sigma(p, P) + \varepsilon I\|_{\text{op}} \geq \varepsilon$, which implies

$$\|A(p, P)\|_{\text{op}} \leq \varepsilon^{-1/2}. \quad (\text{B.8})$$

We now bound the second term, $\|A(p, P) - A(q, P)\|_{\text{op}}$. First we bound $\|A(p, P) - A(q, P)\|_{\text{op}}$ in terms of $\|\Sigma(p, P) - \Sigma(q, P)\|_{\text{op}}$. As $(\Sigma(p, P) + \varepsilon I)$ is

positive definite for $\varepsilon > 0$, we can represent $A = (\Sigma + \varepsilon I)^{-1/2}$ as an integral:

$$(\Sigma + \varepsilon I)^{-1/2} = \frac{1}{\pi} \int_0^\infty t^{-1/2} (\Sigma + (\varepsilon + t)I)^{-1} dt. \quad (\text{B.9})$$

Using the identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$, we get:

$$\begin{aligned} & A(p, P) - A(q, P) \\ &= \frac{1}{\pi} \int_0^\infty t^{-1/2} \frac{\Sigma(p, P) - \Sigma(q, P)}{(\Sigma(p, P) + (\varepsilon + t)I)(\Sigma(q, P) + (\varepsilon + t)I)} dt. \end{aligned} \quad (\text{B.10})$$

Taking the operator norm and using $\|\Sigma + (\varepsilon + t)I\|_{\text{op}} \geq \varepsilon + t$, we have:

$$\begin{aligned} & \|A(p, P) - A(q, P)\|_{\text{op}} \\ &\leq \frac{1}{\pi} \int_0^\infty t^{-1/2} \|\Sigma(p, P) - \Sigma(q, P)\|_{\text{op}} \cdot (\varepsilon + t)^{-1} \cdot (\varepsilon + t)^{-1} dt \end{aligned} \quad (\text{B.11})$$

$$= \frac{1}{\pi} \|\Sigma(p, P) - \Sigma(q, P)\|_{\text{op}} \int_0^\infty t^{-1/2} (\varepsilon + t)^{-2} dt \quad (\text{B.12})$$

$$= \|\Sigma(p, P) - \Sigma(q, P)\|_{\text{op}} \frac{1}{2} \varepsilon^{-3/2}. \quad (\text{B.13})$$

Now we bound $\|\Sigma(p, P) - \Sigma(q, P)\|_{\text{op}}$. We can split the term $k(\Sigma(p, P) - \Sigma(q, P))$ into three sums:

$$\begin{aligned} & k(\Sigma(p, P) - \Sigma(q, P)) \\ &= \sum_{x \in N_k(p, P) \cap N_k(q, P)} [(x - p)(x - p)^\top - (x - q)(x - q)^\top] + \\ & \quad \sum_{x \in N_k(p, P) \setminus N_k(q, P)} (x - p)(x - p)^\top + \sum_{x \in N_k(q, P) \setminus N_k(p, P)} (x - q)(x - q)^\top. \end{aligned} \quad (\text{B.14})$$

The first term is bounded by

$$\sum_{x \in N_k(p, P) \cap N_k(q, P)} 2\|x - p\|_2 \|p - q\|_2 + \|p - q\|_2^2 \leq k(2R_p \cdot d_H + d_H^2), \quad (\text{B.15})$$

where R_p is the radius of the k -neighbourhood of p and $d_H = d_H(P, Q)$.

The other two terms are bounded by kR_p^2 and kR_q^2 , respectively.

Combining these bounds, we have:

$$\|\Sigma(p, P) - \Sigma(q, P)\|_{\text{op}} \leq 2R_p \cdot d_H + d_H^2 + R_p^2 + R_q^2, \quad (\text{B.16})$$

and combining this with the bound of $\|A(p, P) - A(q, P)\|_{\text{op}}$, we get:

$$\|A(p, P) - A(q, P)\|_{\text{op}} \leq \frac{1}{2} \varepsilon^{-3/2} (2R_p \cdot d_H + d_H^2 + R_p^2 + R_q^2). \quad (\text{B.17})$$

Finally, plugging this into Equation (B.7), we have:

$$|h(x, p, P) - h(x, q, P)| \leq \varepsilon^{-1/2} d_H + \frac{1}{2} \varepsilon^{-3/2} (2R_p \cdot d_H + d_H^2 + R_p^2 + R_q^2) \cdot R_q. \quad (\text{B.18})$$

Via similar reasoning, we can bound $|h(x, p, P) - h(x, p, Q)|$:

$$|h(x, p, P) - h(x, p, Q)| = ||| \|A(p, P)(x - p)\|_2 - \|A(p, Q)(x - p)\|_2 ||| \quad (\text{B.19})$$

$$\leq \|A(p, P)(x - p) - A(p, Q)(x - p)\|_2 \quad (\text{B.20})$$

$$= \|(A(p, P) - A(p, Q))(x - p)\|_2 \quad (\text{B.21})$$

$$\leq \|A(p, P) - A(p, Q)\|_{\text{op}} \cdot \|x - p\|_2 \quad (\text{B.22})$$

$$\leq \frac{1}{2} \varepsilon^{-3/2} d_H \cdot (R_p^2 + R_q^2) \cdot R_p. \quad (\text{B.23})$$

By choosing a sufficiently large ε , we achieve the bounds needed by Theorem 3.13. \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] 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] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Extending persistence using Poincaré and Lefschetz duality. *Foundations of Computational Mathematics*, 9(1):79–103, 2009.
- [12] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. Applied Mathematics. American Mathematical Society, 2010.
- [13] Marc G Genton. Classes of kernels for machine learning: A statistics perspective. *Journal of machine learning research*, 2(Dec):299–312, 2001.
- [14] Loukas Grafakos et al. *Classical fourier analysis*, volume 2. Springer, 2008.
- [15] 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.
- [16] Allen Hatcher. *Algebraic topology*. Cambridge: Cambridge University Press, 2002.
- [17] Pepijn Roos Hoefgeest and Lucas Slot. The christoffel-darboux kernel for topological data analysis, 2022.
- [18] John M Lee. *Smooth manifolds*. Springer, 2003.
- [19] 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.
- [20] 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.
- [21] Yukio Matsumoto. *An introduction to Morse theory*, volume 208. American Mathematical Soc., 2002.
- [22] Goeffrey J McLachlan. Mahalanobis distance. *Resonance*, 4(6):20–26, 1999.
- [23] John Willard Milnor. *Morse theory*. Number 51. Princeton university press, 1963.
- [24] Edwin E Moise. *Geometric topology in dimensions 2 and 3*, volume 47. Springer Science & Business Media, 2013.
- [25] James R Munkres. *Topology*. Prentice Hall, 2nd edition, 2000.

- [26] William Ott and James Yorke. Prevalence. *Bulletin of the American Mathematical Society*, 42(3):263–290, 2005.
- [27] Jeff M. Phillips, Bei Wang, and Yan Zheng. Geometric inference on kernel density estimates, 2015.
- [28] Maximilian Schmah. *Topics in Persistent Homology: From Morse Theory for Minimal Surfaces to Efficient Computation of Image Persistence*. PhD thesis, Ruprecht Karl University of Heidelberg, 2023.
- [29] Patrick Schnider and Simon Weber. Introduction to topological data analysis lecture notes FS 2024. <https://ti.inf.ethz.ch/ew/courses/TDA24/Script.pdf>.
- [30] 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.