ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

# Persistence of generalized density functions

Master Thesis

Maxim Mikhaylov

January 19, 2038

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 $\mathbb{R}^n$ to infer the underlying topological structure of a dataset $P \subseteq \mathbb{R}^n$. The persistent homology of the sublevelset or superlevelset filtrations induced by these functions captures multi-scale topological features. A key desirable property is *stability*: small perturbations in the dataset result in similarly small changes in the persistence diagrams. A classic example is the nearest-neighbor distance function $f(P, x) := \min_{p \in P} d(x, p)$, for which the bottleneck distance between persistence diagrams $\mathrm{Dgm}_{f(P)}$ and $\mathrm{Dgm}_{f(Q)}$ is bounded by the Hausdorff distance $d_H(P, Q)$ between datasets [3]:

$$d_b(\mathrm{Dgm}_{f(P)}, \mathrm{Dgm}_{f(Q)}) \leq d_H(P, Q).$$

This thesis investigates *generalized density functions*, which are functions $f : \mathcal{P}(X) \times X \to \mathbb{R}$, and the conditions under which they satisfy a stability property of the form

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

for some finite constant $c$.

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

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

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

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

# Contents

## Chapter 1

# Introduction

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

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

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

where $d$ is a metric on $X$. The homology groups of the sublevel set filtration of this function are precisely the homology groups of the Čech filtration. This function satisfies a *stability* property:

$$d_b(\mathrm{Dgm}_{f(P)}, \mathrm{Dgm}_{f(Q)}) \leq d_H(P, Q), \tag{1.2}$$

where $d_b$ is the bottleneck distance between persistence diagrams and $d_H$ is the Hausdorff distance between point clouds. Stability ensures that small perturbations in the data lead to proportionally small changes in the persistence diagrams, guaranteeing that the topological features extracted from the data are robust to noise.

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

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

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

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

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

In this thesis, we investigate which functions are stable. We consider the following classes of functions:

1. Functions of the form $f(P, x) = \min_{p \in P} h(d(p, x))$, where $h$ is a monotone Lipschitz continuous function. These generalize the classical nearest-neighbor function while preserving stability.
2. Further generalization to $f(P, x) = \min_{p \in P} h(p, x)$, where $h$ is a Lipschitz function with respect to $p$.
3. Functions $f(P, x)$ that are Lipschitz with respect to point clouds *(PC-Lipschitz)*, where $\mathcal{P}(X)$ is equipped with the Hausdorff distance, and the space of real-valued functions $f_P : X \to \mathbb{R}$ is equipped with the $L_\infty$-norm.
4. Finally, we consider Morse functions whose restriction to the critical points is Lipschitz continuous.

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

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

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

## 1.1 Overview

The remainder of this thesis is organized as follows:

- Chapter 2 provides the necessary background on persistent homology, stability, as well as known stability results.
- Chapter 3 presents our main contributions: stability theorems for various classes of generalized density functions.
- Chapter 4 studies stability under common operations.
- Chapter 5 explores the properties of the space of $c$-stable functions, including its topological and algebraic structure.
- Chapter 6 concludes with directions for future work and open questions.

Chapter 2

# Background

This chapter reviews the mathematical foundations for stability of generalized density functions. We briefly review the core concepts of how real-valued functions induce filtrations for persistent homology and then focus on established stability theorems that serve as a foundation and motivation for this work. We assume the reader has a working knowledge of topological data analysis, mainly persistent homology.

## 2.1 Generalized density functions and filtrations

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

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

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

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

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

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

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

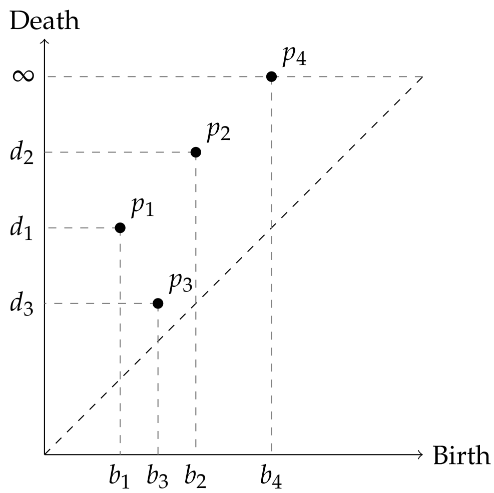


**Figure 2.1:** An example of a persistent diagram. A point $p_i$ is born at $b_i$ and dies at $d_i$. The point $p_4$ does not die, so its death time is $\infty$. The dashed line indicates the diagonal $\Delta$, where $b_i = d_i$.

Applying a homology functor $H_k$ to the filtration $\{f_P^{-1}(-\infty, a]\}_{a\in\mathbb{R}}$ yields a persistence module, which tracks the evolution of topological features as the parameter $a$ varies. The *persistence diagram*, denoted $\mathrm{Dgm}_k(f_P)$, is a multiset of points $(b_i, d_i)$ in the extended plane $\overline{\mathbb{R}}^2$, where $b_i$ represents the birth time and $d_i$ the death time of a $k$-dimensional feature. An example of a persistence diagram is shown in Figure 2.1.

## 2.2 Quantifying stability

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

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

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

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

As the Hausdorff distance requires $X$ to be a metric space, we will assume this throughout the thesis.

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

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

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

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

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

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

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

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

## 2.3 Known stability results

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

### 2.3.1 Čech filtration

As mentioned in Chapter 1, the Čech filtration can be represented as the sublevel set filtration of the nearest-neighbor distance function $f_{\mathrm{dist}}(P, x) := \min_{p \in P} d(x, p)$. The sublevel sets $f^{-1}_{\mathrm{dist},P}(-\infty, a]$ 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 [2, 5], if $P$ is finite, the Čech complex is homotopy equivalent to the union of balls.

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

**Theorem 2.6 (Stability of the Čech filtration, [3])** *Let $f_{\mathrm{dist},P}(x) := \min_{p \in P} d(x, p)$. Then for all finite $P, Q \subseteq \mathcal{P}(X)$ and any dimension $k \geq 0$, we have*

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

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

### 2.3.2 Weighted Čech filtration

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

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

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

*For $q = \infty$, we also define*

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

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

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

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

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

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

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

where similarly to the $L_\infty$-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 stable:

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

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

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

### 2.3.3 Stability of $L_\infty$-norm-bounded functions

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

**Theorem 2.9 (Stability of persistence diagrams, [4])** *Let $X$ be a triangulable space, and $g, h : X \to \mathbb{R}$ be two continuous tame functions. Then for all $k \in \mathbb{N}$, the persistence diagrams of $g$ and $h$ satisfy*

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

This theorem can be immediately adapted to the case of GDFs:

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

This theorem directly implies that if a GDF $f$ satisfies $\|f_P - f_Q\|_\infty \leq c \cdot d_H(P, Q)$ for some constant $c$, then $f$ is $c$-stable. This provides a powerful tool for proving stability of GDFs.

Chapter 3

# What functions are stable?

Chapter 4

# Operations on stable functions

Chapter 5

# The space of stable functions

Chapter 6

# Conclusion

Appendix A

# Weighted Čech complex can be represented as a GDF

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

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

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

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

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

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

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

The proposed GDF takes value

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

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

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

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

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

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

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

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

The proposed GDF takes value

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

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

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

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

$$(d(x, p)^q + g(p)^q)^{1/q} \leq a \tag{A.10}$$
$$d(x, p)^q + g(p)^q \leq a^q \tag{A.11}$$
$$d(x, p)^q \leq a^q - g(p)^q \tag{A.12}$$
$$d(x, p) \leq (a^q - g(p)^q)^{1/q} \tag{A.13}$$

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

# Bibliography

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

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

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

[4] 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.

[5] 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.

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.*