ABSTRACT

The stock market, with its inherent complexity and susceptibility to various economic, political, and social factors, presents a significant challenge in predicting major downturns or crashes. Traditional predictive methods often fall short in capturing the intricate dynamics of the market. This study explores the application of Topological Data Analysis (TDA) as an innovative approach to understanding and predicting stock market crashes. TDA utilizes concepts from algebraic topology to analyze the shape and structure of data in high-dimensional spaces, offering a unique perspective that complements conventional analysis techniques. In this research, we apply TDA to the S&P 500 index, covering data from 1980 to 2024, to identify significant market downturns. By constructing simplicial complexes and employing persistent homology, we capture the topological features of the market data and analyze their persistence across different scales. The persistence diagrams and barcode plots generated from this analysis provide insights into the market's topological structure and its evolution over time.

The findings demonstrate that TDA can effectively identify patterns and anomalies in market data that may precede significant downturns. The persistent topological features observed in the analysis are indicative of underlying market trends and potential instability. By leveraging these insights, TDA can serve as a powerful tool for enhancing market prediction models and improving risk management strategies. This study contributes to the growing body of literature on the application of topological methods in financial analysis. The integration of TDA into market analysis frameworks offers a novel approach to understanding the complex behavior of financial markets and provides a complementary tool for traditional economic and statistical models. The results of this research underscore the potential of TDA to enhance our ability to detect and respond to market crashes, ultimately contributing to more robust financial systems.

Keywords: TDA; Analysis

Acknowledgements

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Contents

Contents iii

List of Figures v

1 Introduction 1

2 Preliminaries 3
2.1 Simplicial Homology . . . 3
2.2 Singular Homology . . . 4
2.3 Homotopy Invariance . . . 4
2.4 Required Algebra [1] . . . 6
2.5 Computing Simplicial Homology over a PID . . . 6
2.6 Vietoris-Rips Complexes . . . 7
2.6.1 Filtration of Vietoris-Rips Complexes . . . 7
2.7 Barcodes in Persistent Homology . . . 7
2.7.1 Definition of Barcodes . . . 8
2.7.2 Construction of Barcodes . . . 8
2.7.3 Interpreting Barcodes [2] . . . 8

3 Geometric Reconstruction of Point Cloud Data 10
3.1 Complexes from Point Cloud . . . 10
3.2 Reconstructing the Manifold [3] . . . 11

4 Advanced Techniques in Topological Data Analysis 12
4.1 Takens' Embedding . . . 12
4.2 Persistence Diagrams . . . 13
4.3 Homological Derivatives . . . 14

5 Vectorized Representations of Persistent Homology 15
5.1 Real-valued Summaries . . . 15
5.2 Persistence Image . . . 15
5.3 Betti Curves . . . 16
5.4 Entropy Summary Function . . . 16
5.5 Persistence Landscapes . . . 16

6 Methodology 18
6.1 Data Source and Preparation . . . 18
6.1.1 Data Collection . . . 18
6.2 Data Preprocessing . . . 18
6.3 Takens' Embedding and Point Clouds . . . 18
6.3.1 Takens' Embedding Theorem . . . 18
6.3.2 Complex Time Series Embedding . . . 19
6.3.3 Choosing the Embedding Dimension and Time Delay . . . 19
6.3.4 Practical Considerations . . . 19
6.4 Baseline Model: First Derivative . . . 20
6.5 Topological Data Analysis (TDA) . . . 20
6.5.1 Computation of Persistence Diagrams . . . 20
6.6 Homological Derivatives . . . 20
6.6.1 Steps for Calculating Homological Derivatives . . . 21

7 Analysis Results 22
7.1 Close Price Over Time . . . 22
7.2 First Derivative . . . 22
7.3 Crash Probability Based on First Derivative . . . 23
7.4 Landscape Distances . . . 23
7.5 Norm of L^p Difference . . . 23
7.6 Landscape and Betti Curve Distances . . . 24

7.7 Crash Probability Based on Landscape Distance 25
7.8 Comparison of Methods 25

8 Conclusion **27**

9 Summary **29**

References **30**

List of Figures

1 Vietoris-Rips complex $VR_1(X)$ for a set of points X. 7
2 Filtration of Vietoris-Rips complexes for a set of points X at different scales $\epsilon_1, \epsilon_2, \epsilon_3$. . . 8
3 Vietoris-Rips complexes at different scales ϵ. 8
4 Example of a barcode with H0 (connected components), H1 (loops), and H2 (voids) features. 9
5 Close Price of the GSPC data from 1980 to 2024 22
6 First Derivative of the GSPC data 22
7 Crash Probability Based on First Derivative 23
8 Landscape Distances for the GSPC data 23
9 Norm of L^p Difference for the GSPC data 24
10 Landscape and Betti Curve Distances for the GSPC data 24
11 Crash Probability Based on Landscape Distance (GSPC) 25
12 Crash Probability Based on Landscape Distance (GPBUSD) 25
13 Crash Probability Based on Landscape Distance (NASDAQ) 26
14 Comparison of Baseline and Topological Detectors (GSPC) 26
15 Comparison of Baseline and Topological Detectors (NASDAQ) 27
16 Comparison of Baseline and Topological Detectors (GPBUSD) 27

Symbols

- H: Homology group. These groups capture information about the number of connected components, loops, and higher-dimensional holes in a dataset.
- C: Chain complex. This is a sequence of abelian groups connected by boundary operators, used in algebraic topology to define homology.
- ∂: Boundary operator. This maps an i-simplex to its boundary, which is an $(i-1)$-simplex.
- PID: Principal Ideal Domain. A type of ring in which every ideal is principal, meaning it can be generated by a single element. PIDs are useful for simplifying calculations in algebraic topology.
- β: Betti number. These numbers represent the rank of the homology groups and indicate the number of k-dimensional holes in a topological space at each dimension k.
- Δ: Simplicial complex. A combinatorial structure used to represent the topology of data, formed by combining simplices (points, line segments, triangles, etc.).
- ϵ: Epsilon. A small positive quantity used in various mathematical contexts, often as a distance threshold in topological data analysis.
- σ: n-simplex. The n-simplex spanned by a geometrically independent set of points $\{a_0, a_1, \ldots, a_n\}$ from R^N.
- $[v_0, v_1, \ldots v_n]$: Oriented n-simplex. An oriented n-simplex spanned by the ordered set of vertices $v_0, v_1, \ldots, v_n$.
- ∇d_M: Gradient of the distance function. A map $\nabla d_M : R^d \to R^d$.
- d_H: Hausdorff Distance. Given two subspaces K, K' of a metric space (X, d), the Hausdorff distance between them.
- d_{GH}: Gromov-Hausdorff Distance. The infimum over all $r \geq 0$ such that there exists a metric space (X, d) and compact subsets $C_1, C_2 \subseteq X$ which are isometric to K_1 and K_2 respectively such that $d_H(C_1, C_2) \leq r$.
- d_M: Distance function. A map $d_M : R^d \to R^+$ that gives the distance of each point in R^d from the subspace M.
- Φ: Time-delay embedding map. Defined by $\Phi(x) = (h(x), h(f(x)), h(f^2(x)), \ldots, h(f^{d-1}(x)))$.
- $\lambda_k(t)$: Persistence landscape. A piecewise linear function that records the persistence of features at different scales.
- $\beta_k(t)$: Betti curve. Represents the number of k-dimensional homological features present in the data at a given scale t.
- $\mathrm{VR}_\epsilon(X)$: Vietoris-Rips Complex. An abstract simplicial complex where a subset $\sigma \subseteq X$ is a simplex if and only if the pairwise distance between any two points in σ is at most ϵ.
- $C_\epsilon(X)$: Čech Complex. An abstract simplicial complex whose k-simplices are given by all $k+1$ points $\{x_{i_1}, \ldots, x_{i_{k+1}}\}$ such that $\bigcap_{j=0}^{k} B_\epsilon(x_{i_j}) \neq \emptyset$.
- $R_\epsilon(X)$: Rips Complex. An abstract simplicial complex whose k-simplices are determined by $k+1$ points $\{x_{i_1}, \ldots, x_{i_{k+1}}\}$ that are pairwise less than ϵ apart.
- $f_\#$: Chain map. A map between the singular chain complexes of X and Y.
- Distance(D_i, D_{i+1}): Distance between persistence diagrams. Calculated as $\sum_j \|(b_j, d_j)_i - (b_j, d_j)_{i+1}\|_p$.
- Total Persistence: Sum of the lifetimes of all features in the persistence diagram. Calculated as $\sum_{i=1}^{n}(d_i - b_i)$.
- Entropy: Entropy summary function. Calculated as $E = -\sum_{i=1}^{n} p_i \log p_i$.
- p_i: Normalized persistence of each feature. Calculated as $p_i = \frac{d_i - b_i}{\sum_{j=1}^{n}(d_j - b_j)}$.
- $\bigoplus$: Direct sum. This symbol is often used to represent the direct sum of groups or modules.

1 Introduction

The stock market represents a complex and dynamic system, characterized by its sensitivity to a multitude of economic, political, and social factors [4] [5]. Over the years, the behavior of stock markets has intrigued researchers and practitioners alike, primarily due to the challenges associated with predicting significant downturns or crashes. These market crashes, defined by rapid and often severe declines in stock prices, can lead to substantial economic repercussions, impacting investors, businesses, and the broader economy.Traditional methods of predicting market crashes have relied heavily on financial indicators and economic models. Indicators such as price-to-earnings ratios, interest rates, and economic growth rates have been used to forecast potential downturns. While these methods have provided valuable insights, they often fall short in capturing the intricate and multifaceted nature of market behavior. The underlying dynamics of the stock market are influenced by a myriad of interconnected factors, making it challenging to develop models that can accurately predict crashes.

In recent years, advancements in data analysis techniques have opened new avenues for exploring and understanding market behavior. One such innovative approach is Topological Data Analysis (TDA) [6] [7]. TDA is a field of study that applies concepts from algebraic topology to analyze the shape and structure of data in high-dimensional spaces. By examining the topological features of data, TDA provides a unique perspective that complements traditional statistical and machine learning methods.

Topological Data Analysis leverages the concept of shape to extract meaningful information from complex datasets [8]. Unlike traditional methods that focus on specific numerical values or trends, TDA examines the geometric and topological properties of data. This approach allows for the identification of patterns and structures that may not be apparent through conventional analysis techniques. By capturing the underlying shape of data, TDA can uncover hidden relationships and provide insights into the behavior of complex systems such as the stock market. At the heart of TDA is the concept of homology, a mathematical framework used to study the topological features of spaces. Homology groups capture information about the number of connected components, loops, and higher-dimensional holes in a dataset. These features, known as topological invariants, remain unchanged under continuous deformations, making them robust tools for analyzing the qualitative aspects of data. By constructing homology groups for different scales, TDA can track the evolution of topological features across various levels of resolution.

A key component of TDA is the construction of simplicial complexes, which are combinatorial structures used to represent the topology of data [9]. A simplicial complex is formed by combining simplices, which are generalizations of points, line segments, and triangles. By analyzing the simplicial complexes constructed from data, TDA can capture the topological features and study their behavior. This approach provides a powerful framework for understanding the shape of data and its underlying structure. The process of Topological Data Analysis involves several steps, starting with the construction of a filtration, which is a sequence of nested simplicial complexes. By varying a parameter, typically a distance threshold, TDA constructs a filtration that captures the evolution of topological features at different scales. This process allows for the identification of persistent features [10], which are those that remain present across multiple scales. These persistent features are often indicative of significant structures in the data and can provide valuable insights into the behavior of the system under study.

The output of TDA is often represented using persistence diagrams which visualize the birth and death of topological features across different scales. Persistence diagrams plot the birth and death times of features, providing a compact representation of the topological structure of the data. Persistence plots, on the other hand, represent features as horizontal bars, with the length of the bar indicating the persistence of the feature. These visualizations offer a clear and intuitive way to interpret the results of TDA and identify significant patterns in the data.

In the context of stock market analysis, TDA offers a novel approach to understanding market behavior and identifying potential crashes [11]. By examining the topological features of market data, TDA can uncover patterns and structures that may not be apparent through traditional methods. For instance, the persistence of certain topological features across different time scales may indicate underlying market trends or anomalies that could signal an impending crash. By leveraging the insights provided by TDA, researchers and practitioners can develop more robust and comprehensive models for predicting market behavior.

This study aims to explore the application of Topological Data Analysis for detecting stock market crashes. By analyzing the GSPC, GBPUSD and NASDAQ from 1980 to 2024, we demonstrate the effectiveness of TDA in identifying significant market downturns. The datasets, being one of the most widely followed equity indices, serves as an ideal dataset for this analysis. The index encompasses a

diverse range of companies and sectors, providing a comprehensive view of market behavior over time. In this study, we will apply TDA to the three stock symbols to identify topological features and analyze their persistence across different time scales. By constructing simplicial complexes from the market data, we will capture the topological structure and study its evolution. The resulting persistence diagrams will provide insights into the behavior of the market and help identify potential crashes. Through this analysis, we aim to demonstrate the utility of TDA as a complementary tool for market analysis and prediction.

The remainder of this report is structured as follows: In Chapter 2, we provide the necessary mathematical background on simplicial homology, singular homology, and homotopy invariance. Chapter 3 discusses the geometric reconstruction of point cloud data and its relevance to TDA. Chapter 4 delves into persistence homology and its applications to data analysis. Chapter 5 explores vectorized representations of persistent homology, including real-valued summaries and persistence images. Finally, Chapter 6 presents a topological pipeline for predicting future financial market crashes, with a detailed analysis of the datasets.

Through this study, we aim to highlight the potential of Topological Data Analysis as a powerful tool for understanding complex systems and predicting market behavior. By leveraging the unique insights provided by TDA, we can enhance our ability to detect market crashes and develop more effective strategies for managing financial risks.

2 Preliminaries

This chapter seeks to establish some definitions and results required for the thesis. Sections 2.1 and 2.2 discuss simplicial and singular homology theory [12]. The contents of this chapter also include the homotopy invariance [13] property of homology and the computation of simplicial homology modules using linear algebra.

2.1 Simplicial Homology

A set of points $\{a_0, a_1, \ldots, a_n\}$ in R^N is said to be geometrically independent iff the vectors $a_1 - a_0, \ldots, a_n - a_0$ are linearly independent.

Definition 2.1 (n-simplex) *The n-simplex σ spanned by a geometrically independent set of points $\{a_0, a_1, \ldots, a_n\}$ from R^N, is the set of all points*

$$\sigma = \left\{ \sum_{i=0}^{n} t_i a_i \mid \sum_{i=0}^{n} t_i = 1 \text{ and } t_i \geq 0 \text{ for all } i \in \{0, \ldots, n\} \right\}.$$

The set of points $\{a_0, a_1, \ldots, a_n\}$ are called the vertices of σ and n is the dimension of the n-simplex. The simplex spanned by a subset of the vertices is called a face of σ and the union of all faces of σ is called the boundary of σ. The standard n-simplex Δ^n is the simplex spanned by the standard basis vectors in $\{e_i\}_{i=1}^{n+1}$ in R^{n+1}.

Definition 2.2 (Orientation) *An orientation of an n-simplex is the equivalence class of an ordering of the vertex set under the equivalence relation that identifies orderings that differ by an even permutation.*

An oriented n-simplex spanned by the ordered set of vertices $v_0, v_1, \ldots, v_n$ is denoted by $[v_0, v_1, \ldots v_n]$.

Definition 2.3 (Geometric Simplicial Complex) *A geometric simplicial complex K in R^N is a collection of simplices in R^N such that:*

1. *Every face of a simplex in K also belongs to K, and*
2. *The intersection of two distinct simplices in K is a face of both of them.*

Definition 2.4 (Abstract Simplicial Complex) *A collection S of finite non-empty sets is said to be an abstract simplicial complex if every non-empty subset of an element of S also belongs to S. In this case, every element A of S is called a simplex of dimension $|A| - 1$. The vertex set V of an abstract simplicial complex S is the union of all singletons in S.*

Given a simplicial complex K, the collection of vertices of all the constituent simplices forms an abstract simplicial complex called the vertex scheme of K.

A bijective map between the vertex sets of two abstract simplicial complexes is an isomorphism if it maps every simplex in one complex to a simplex in the other. Every abstract simplicial complex can be shown to be isomorphic to the vertex scheme of some simplicial complex which is called its geometric realization. As a result, every abstract simplicial complex can be associated with a topological space determined by this geometric realization.

A related concept is that of a triangulation. A geometric simplicial complex K is said to be a triangulation of a topological space X, if there exists a homeomorphism $\gamma : K \to X$. A space that accepts a triangulation is said to be triangulable.

Example 2.1 *The geometric realization K of the abstract simplicial complex with the vertex set*

$$\{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}\}$$

whose simplices are given by

$$\{a, b, c\}, \{c, a, d\}, \{c, e, d\}, \{e, f, d\}, \{a, e, f\}, \{a, b, f\}$$

and their non-empty subsets is depicted in Figure 1.2. K is homeomorphic to a Möbius strip and thus defines a triangulation for this space.

Definition 2.5 (p-chains) *Given a simplicial complex K, the free abelian group generated by all the oriented p-simplices in K is called the group of p-chains in K and is denoted by $C_p(K)$.*

Definition 2.6 (Boundary Map) *The p-th boundary map is the linear homomorphism $\partial_p : C_p(K) \to C_{p-1}(K)$ that maps each oriented p-simplex $[v_0, v_1, \ldots, v_p]$ as follows:*

$$\partial_p[v_0, v_1, \ldots, v_p] = \sum_{i=0}^{n} (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_p].$$

It can easily be shown that $\partial_p \circ \partial_{p+1} = 0$. As a result, one can associate the following chain complex with every simplicial complex K:

$$\cdots \to C_{p+1}(K) \xrightarrow{\partial_{p+1}} C_p(K) \xrightarrow{\partial_p} C_{p-1}(K) \to \cdots \to C_0 \to 0.$$

The kernel of the map ∂_p is called the group of p-cycles and is denoted by $Z_p(K)$ while the image of ∂_{p+1} is called the group of p-boundaries and is denoted by $B_p(K)$. As $\partial_p \circ \partial_{p+1} = 0$, $Im(\partial_{p+1}) = B_p(K) \subseteq \ker(\partial_p) = Z_p(K) \subseteq C_p(K)$.

Definition 2.7 (Homology) *The p-th homology group of a simplicial complex K, denoted by $H_p(K)$, is given by:*

$$H_p(K) = \frac{Z_p(K)}{B_p(K)}.$$

The rank of $H_p(K)$ is called the p-th Betti number of K and is denoted by $\beta_p(K)$.

2.2 Singular Homology

Definition 2.8 (n-singular simplex) *Given a space X, a singular n-simplex of X is a continuous map $\sigma : \Delta^n \to X$.*

One can then define singular homology groups of X by constructing chain groups and boundary maps in a similar manner as outlined previously using singular n-simplices.

Remark 2.1 *Similar to how simplicial homology was built by considering simplices as building blocks, a cubical homology theory can also be developed by using n-cubes given by:*

$$I^n = [0, 1] \times \cdots \times [0, 1] \text{ (n-times).}$$

Remark 2.2 *While chain groups for simplicial and singular homology groups were defined using Z-linear combinations, the same can be done with any base ring R. Using this, one can define homology modules corresponding to any choice of a ground ring.*

2.3 Homotopy Invariance

Let X and Y be topological spaces and $f : X \to Y$ be a continuous map. For a given singular n-simplex σ of X, $f \circ \sigma$ is a singular n-simplex of Y. Thus, for each n we can define a map $f_\# : C_n(X) \to C_n(Y)$ by linearly extending the map that assigns $\sigma \mapsto f \circ \sigma$.

Proposition 2.1 *The maps $f_\# : C_n(X) \to C_n(Y)$, define a chain map $f_\#$ between the singular chain complexes of X and Y:*

$$\begin{array}{ccccccccc} \cdots \to C_{n+1}(X) & \xrightarrow{\partial_{n+1}} & C_n(X) & \xrightarrow{\partial_n} & C_{n-1}(X) & \to \cdots \to & C_0(X) & \to 0 \\ & & \downarrow f_\# & & \downarrow f_\# & & \downarrow f_\# & & \\ \cdots \to C_{n+1}(Y) & \xrightarrow{\partial_{n+1}} & C_n(Y) & \xrightarrow{\partial_n} & C_{n-1}(Y) & \to \cdots \to & C_0(Y) & \to 0 \end{array}$$

As the maps $f_\# : C_n(X) \to C_n(Y)$ are linear homomorphisms by construction, all that is left to be shown is that each square in the diagram commutes. Given an n-simplex σ in X,

$$f_\# \circ \partial_n(\sigma) = f_\# \left(\sum_{i=0}^{n} (-1)^i \sigma[v_0, \ldots, \hat{v}_i, \ldots, v_n] \right) = \sum_{i=0}^{n} (-1)^i f(\sigma[v_0, \ldots, \hat{v}_i, \ldots, v_n])$$

$$= \sum_{i=0}^{n} (-1)^i (f\sigma)[v_0, \ldots, \hat{v}_i, \ldots, v_n] = \partial_n \circ f_\#(\sigma).$$

The proposition thus holds as a result of the linearity of ∂_n and $f_\#$.

Proposition 2.2 *Given topological spaces X, Y and Z and maps f, g such that $X \xrightarrow{f} Y \xrightarrow{g} Z$, the following hold:*

$$(g \circ f)_{\#} = g_{\#} \circ f_{\#},$$
$$(g \circ f)_{*} = g_{*} \circ f_{*}.$$

Theorem 2.1 *Homotopic maps $f, g : X \to Y$ induce the same homomorphism at the homology level, $f_* = g_* : H_n(X) \to H_n(Y)$.*

Given the space $\Delta^n \times I$, let $\Delta^n \times \{0\} = [e_0, e_1, \ldots, e_n]$ and $\Delta^n \times \{1\} = [v_0, v_1, \ldots, v_n]$ such that e_i maps to v_i under the projection $\Delta^n \times I \to \Delta^n$. The space $\Delta^n \times I$ can then be expressed as a union of $(n+1)$-simplices $\{[e_0, e_1, \ldots, e_i, v_i, \ldots, v_n]\}_{i=0}^{n}$.

Let $F : X \times I \to Y$ represent the homotopy from f to g. Given any simplex σ in $C_n(X)$, we have the map

$$F \circ (\sigma \times 1) : \Delta^n \times I \to Y.$$

We can thus define a family of maps $P : C_n(X) \to C_{n+1}(Y)$ such that for any singular n-simplex σ in X,

$$P(\sigma) = \sum_{i=0}^{n} (-1)^i (F(\sigma \times 1))[e_0, e_1, \ldots, e_i, v_i, \ldots, v_n].$$

Consider the maps ∂P and $P\partial : C_n(X) \to C_n(Y)$,

$$P\partial(\sigma) = \sum_{j<i} (-1)^{i+j} (F(\sigma \times 1))[e_0, \ldots, e_j, v_j, \ldots, \hat{v}_i, \ldots, v_n]$$

$$+ \sum_{j>i} (-1)^{i+j+1} (F(\sigma \times 1))[e_0, \ldots, \hat{e}_i, \ldots, e_j, v_j, \ldots, v_n],$$

$$\partial P(\sigma) = \sum_{j\leq i} (-1)^{i+j} (F(\sigma \times 1))[e_0, \ldots, \hat{e}_j, \ldots, e_i, v_i, \ldots, v_n]$$

$$+ \sum_{j\geq i} (-1)^{i+j+1} (F(\sigma \times 1))[e_0, \ldots, e_i, v_i, \ldots, \hat{v}_j, \ldots, v_n].$$

$$P\partial(\sigma) + \partial P(\sigma) = F(\sigma \times 1)[v_0, v_1, \ldots, v_n] - F(\sigma \times 1)[e_0, e_1, \ldots, e_n] = g_{\#}(\sigma) - f_{\#}(\sigma).$$

Given $\sigma \in Z_n(X)$,

$$g_{\#}(\sigma) - f_{\#}(\sigma) = \partial P(\sigma) + P\partial(\sigma) = \partial P(\sigma) \in B_n(Y).$$

From the previous statement, one can see that $f_{\#}$ and $g_{\#}$ map a cycle in X to homologous cycles in Y. Thus, the map induced at the homology level by f and g are the same.

Proposition 2.3 *Homotopically equivalent spaces X and Y have isomorphic homology groups $H_n(X)$ and $H_n(Y)$ for all $n \in N \cup \{0\}$.*

Let $f : X \to Y$ be a homotopy equivalence and $g : Y \to X$ be its homotopy inverse. It then follows from Proposition 1.3.2 and Theorem 1.3.1 that $g_* f_* = 1_{H_n(X)}$ and $f_* g_* = 1_{H_n(Y)}$. Thus, $H_n(X) \cong H_n(Y)$.

Remark 2.3 *The singular homology groups of a simplicial complex are the same as its simplicial homology groups. Using the homotopy invariance property of singular homology, we can also conclude that the singular homology groups of a triangulable space coincide with the simplicial homology of its triangulation.*

2.4 Required Algebra [1]

Definition 2.9 (Graded Ring [14]) *A ring R is called graded if there exists a family of subgroups $\{R_i\}_{i\in Z}$ of R such that:*

1. *$R = \bigoplus_{i\in Z} R_i$ as abelian groups, and*
2. *$R_n R_m \subseteq R_{n+m}$ for all $n, m \in Z$.*

A graded ring R is said to be non-negatively graded if $R_n = 0$ for all $n < 0$. For a given $i \in Z$, any element in R_i is called a homogeneous element of degree i.

Definition 2.10 (Graded Modules [14]) *A module M over a graded ring R is said to be a graded R-module if there exists a family of subgroups $\{M_i\}_{i\in Z}$ of M such that:*

1. *$M = \bigoplus_{i\in Z} M_i$ as abelian groups, and*
2. *$R_n M_m \subseteq M_{n+m}$ for all $n, m \in Z$.*

A graded R-module is said to be non-negatively graded if $M_n = 0$ for all $n < 0$.

Definition 2.11 (-shift) *Given a graded ring R and $\alpha \in Z$, one can define a new ring denoted by $P_\alpha R$ as $\bigoplus_{i\in Z} R_{\alpha+i}$ by shifting the gradation on R by α.*

Theorem 2.2 *Given a finitely generated module M over a PID D, there exist unique non-zero elements $d_1, \ldots, d_m \in D$, where $d_1 \mid d_2 \mid \ldots \mid d_m$ and $\beta \in N \cup \{0\}$ such that M is isomorphic to a direct sum of cyclic D-modules as follows:*

$$M \cong D^\beta \oplus \left(\bigoplus_{i=1}^{m} D/d_i D \right).$$

A structure theorem for graded modules over a PID can be defined in a similar fashion as described in Theorem 1.4.1.

Theorem 2.3 *Given a graded module M over a graded PID D, there exist unique homogeneous elements $d_1, \ldots, d_m \in D$ such that $d_1 \mid d_2 \mid \ldots \mid d_m$ and $\alpha_1, \ldots, \alpha_n \;\; \gamma_1, \ldots, \gamma_m \in Z$ such that:*

$$M \cong \left(\bigoplus_{i=1}^{n} X_{\alpha_i} D \right) \oplus \left(\bigoplus_{j=1}^{m} X_{\gamma_j} D/d_j D \right).$$

2.5 Computing Simplicial Homology over a PID

Simplicial homology [15] over a PID D can be easily computed using results from linear algebra. For most practical applications, the ground ring $Z/2Z$ is preferred.

A matrix M_n associated with boundary map $\partial_n : C_n(K) \to C_{n-1}(K)$ with the cycles as basis is constructed. This matrix can be reduced to the normal form where $d_1 \mid d_2 \mid \ldots \mid d_{k_n}$. Using the structure theorem, we can make the following observations:

1. The elements greater than 1 in the set $\{d_1, \ldots, d_{k_n}\}$ correspond to the torsion coefficients in the decomposition of $H_{n-1}(X)$.
2. The number of zero columns in M_n, denoted by α_n, represents the rank of $Z_n(X)$.
3. The $(n-1)$-th Betti number of K, is given by $\beta_{n-1} = \alpha_{n-1} - k_n$.

Thus, we can determine the homology modules of K over D for all dimensions. Similarly, the structure theorem over graded PIDs can be used to determine the graded homology modules.

2.6 Vietoris-Rips Complexes

Vietoris-Rips complexes [16] are a type of simplicial complex used to approximate the shape of a set of points in a metric space. They are particularly useful in topological data analysis for studying the topological features of point cloud data.

Definition 2.12 (Vietoris-Rips Complex) *Given a set of points X in a metric space and a distance parameter ϵ, the Vietoris-Rips complex $VR_\epsilon(X)$ is defined as the abstract simplicial complex where a subset $\sigma \subseteq X$ is a simplex if and only if the pairwise distance between any two points in σ is at most ϵ.*

Example 2.2 *Consider a set of points $X = \{x_1, x_2, x_3, x_4\}$ in a metric space, and let the distance parameter $\epsilon = 1$. If the pairwise distances between the points are given by the following matrix:*

$$\begin{pmatrix} 0 & 0.5 & 1.2 & 0.8 \\ 0.5 & 0 & 0.6 & 1.1 \\ 1.2 & 0.6 & 0 & 0.7 \\ 0.8 & 1.1 & 0.7 & 0 \end{pmatrix}$$

Then the Vietoris-Rips complex $VR_1(X)$ will include the simplices formed by points whose pairwise distances are less than or equal to 1. Specifically, the 1-simplices (edges) are:

$$\{\{x_1, x_2\}, \{x_1, x_4\}, \{x_2, x_3\}, \{x_3, x_4\}\}$$

The 2-simplices (triangles) are:

$$\{\{x_1, x_2, x_4\}, \{x_2, x_3, x_4\}\}$$

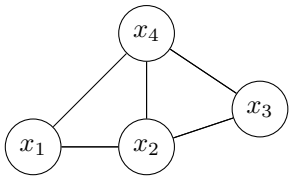

Figure 1: Vietoris-Rips complex $VR_1(X)$ for a set of points X.

2.6.1 Filtration of Vietoris-Rips Complexes

A filtration of Vietoris-Rips complexes [17] is a sequence of nested complexes that represent the data at different scales. As the distance parameter ϵ increases, more simplices are added, capturing the topological features of the data at various levels of granularity.

Definition 2.13 (Filtration [18]) *Given a set of points X and an increasing sequence of distance parameters $\epsilon_1 < \epsilon_2 < \cdots < \epsilon_n$, the corresponding filtration is the sequence of Vietoris-Rips complexes:*

$$VR_{\epsilon_1}(X) \subseteq VR_{\epsilon_2}(X) \subseteq \cdots \subseteq VR_{\epsilon_n}(X)$$

Filtrations are used in persistent homology to study how topological features evolve over different scales, providing insights into the structure and shape of the data [19].

Example 2.3 *Consider the set of points $X = \{x_1, x_2, x_3, x_4\}$ and distance parameters $\epsilon_1 = 0.5$, $\epsilon_2 = 1.0$, and $\epsilon_3 = 1.5$. The filtration is:*

$$VR_{0.5}(X) \subseteq VR_{1.0}(X) \subseteq VR_{1.5}(X)$$

This section on Vietoris-Rips complexes provides the necessary background to understand their application in TDA for analyzing the topological features of data.

2.7 Barcodes in Persistent Homology

Barcodes [20] are a pivotal tool in persistent homology, a method within TDA that tracks the birth and death of topological features across various scales. The concept of barcodes stems from the need to analyze data beyond conventional statistical measures, focusing instead on the shape and connectivity of data points.

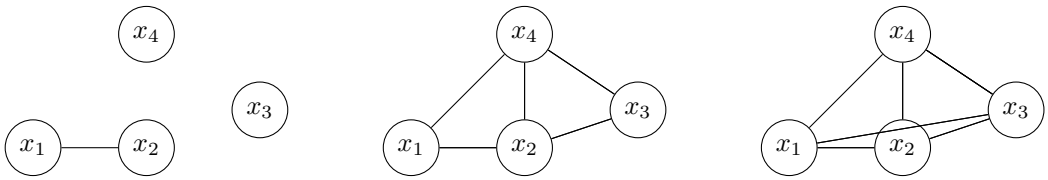

Figure 2: Filtration of Vietoris-Rips complexes for a set of points X at different scales $\epsilon_1, \epsilon_2, \epsilon_3$.

2.7.1 Definition of Barcodes

Definition 2.14 (Barcode) *A barcode is a graphical representation of the persistence of topological features in a filtered simplicial complex. It consists of a collection of intervals, where each interval corresponds to a feature that appears (birth) and disappears (death) as the filtration parameter changes.*

Each interval in a barcode represents the lifespan of a homological feature. Longer intervals typically indicate more significant features, while shorter intervals may correspond to noise.

2.7.2 Construction of Barcodes

To construct a barcode, we follow these steps:
1. Create a filtration of the simplicial complex by varying the scale parameter. 2. Compute the homology at each scale. 3. Record the birth and death of each homological feature.

Example 2.4 *Consider the set of points $X = \{x_1, x_2, x_3, x_4\}$ with the following pairwise distances:*

$$\begin{pmatrix} 0 & 0.5 & 1.2 & 0.8 \\ 0.5 & 0 & 0.6 & 1.1 \\ 1.2 & 0.6 & 0 & 0.7 \\ 0.8 & 1.1 & 0.7 & 0 \end{pmatrix}$$

As we increase the scale parameter ϵ, we obtain the following Vietoris-Rips complexes:

- *At $\epsilon = 0.5$: Only vertices and edges $\{x_1, x_2\}$ and $\{x_2, x_3\}$ are present.*
- *At $\epsilon = 1.0$: Additional edges $\{x_1, x_4\}$, $\{x_3, x_4\}$, and triangles $\{x_1, x_2, x_4\}$ and $\{x_2, x_3, x_4\}$ appear.*
- *At $\epsilon = 1.5$: The complex becomes fully connected, and more triangles $\{x_1, x_3, x_4\}$ appear.*

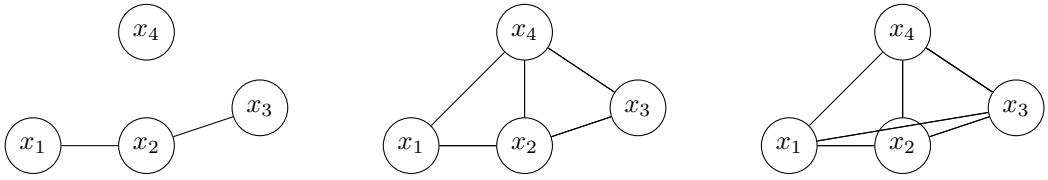

Figure 3: Vietoris-Rips complexes at different scales ϵ.

2.7.3 Interpreting Barcodes [2]

Each bar in a barcode represents a topological feature:

- **H0 features** (connected components): Bars represent connected components in the data. A long bar indicates a persistent component.
- **H1 features** (loops): Bars represent loops or holes. A long bar indicates a significant loop.
- **H2 features** (voids): Bars represent higher-dimensional voids.

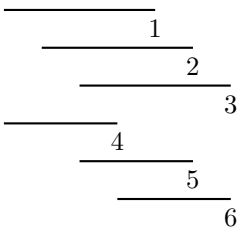

Figure 4: Example of a barcode with H0 (connected components), H1 (loops), and H2 (voids) features.

In this example, the barcode shows three H0 features, two H1 features, and one H2 feature. The length of each bar represents the persistence of the feature. Barcodes provide a visual summary of the topological features of the data, offering insights into its structure that are robust to noise and scale variations. They are widely used in various fields such as biology, neuroscience, and sensor networks to reveal hidden structures and patterns in complex datasets. This section on barcodes in persistent homology provides the necessary background to understand their application in TDA for analyzing the topological features of data.

3 Geometric Reconstruction of Point Cloud Data

Before delving into topics in TDA, it is necessary to gain an understanding of the topological information of interest present in the data and the steps involved in estimating the same. This chapter explores these ideas by considering the specific example of point cloud data [21]. Here, a point cloud X is assumed to be a finite collection of points $\{x_1, \ldots, x_n\}$ in R^d sampled i.i.d using a probability measure μ with a compact support M. It is the topology of this underlying space M that one wishes to capture by using TDA. The first step towards estimating the topological features of M using X is constructing structures on the point cloud which capture the required topological information.

3.1 Complexes from Point Cloud

A space X_ϵ can be naturally constructed from a given point cloud $X = \{x_1, \ldots, x_n\}$ by considering the union of closed balls of radius ϵ centered at $\{x_1, \ldots, x_n\}$.

Definition 3.1 (Nerve of a cover) *Given a cover $U = \{U_\alpha\}_{\alpha \in A}$ of a space Y, the nerve of U is the abstract simplicial complex $N(U)$ whose k-simplices are determined by $k+1$ elements of U that have a non-empty intersection. That is,*

$$[U_{i_0}, \ldots, U_{i_k}] \in N(U) \iff \bigcap_{n=0}^{k} U_{i_n} \neq \emptyset.$$

Theorem 3.1 (Nerve Theorem [22]) *Let $U = \{U_\alpha\}_{\alpha \in A}$ be a cover of the space Y such that for any $A_0 \subseteq A$, the intersection $\bigcap_{\alpha \in A_0} U_\alpha$ is either contractible or empty. Then, the space Y is homotopically equivalent to $N(U)$.*

The Čech complex of the point cloud $\hat{X}$ for a given $\epsilon > 0$, denoted by $C_\epsilon(X)$, is the nerve of the covering $\{B_\epsilon(x_i)\}_{i=1}^n$ of X_ϵ. As the intersection of closed balls in R^d is convex and hence contractible, it follows from the Nerve Theorem that the Čech complex is homotopically equivalent to the ϵ-thickening of X.

Definition 3.2 (Čech Complex [23]) *For a given point cloud $X = \{x_1, \ldots, x_n\}$ and $\epsilon > 0$, the Čech complex $C_\epsilon(X)$ is the abstract simplicial complex whose k-simplices are given by all $k+1$ points $\{x_{i_1}, \ldots, x_{i_{k+1}}\}$ such that*

$$\bigcap_{j=0}^{k} B_\epsilon(x_{i_j}) \neq \emptyset.$$

Definition 3.3 (Rips Complex) *For a given point cloud $X = \{x_1, \ldots, x_n\}$ and $\epsilon > 0$, the Rips complex $R_\epsilon(X)$ is the abstract simplicial complex whose k-simplices are determined by $k+1$ points $\{x_{i_1}, \ldots, x_{i_{k+1}}\}$ that are pairwise less than ϵ apart.*

The Rips complex is an instance of a flag complex and is completely determined by the ϵ-connectivity graph given by its 1-simplices. It also should be noted that the Rips complex may lie in a Euclidean space of dimension greater than that in which the point cloud is embedded.

Proposition 3.1 *For any given point-cloud X and $\epsilon > 0$,*

$$C_\epsilon(X) \subseteq R_{2\epsilon}(X) \subseteq C_{2\epsilon}(X).$$

Both these inclusions follow from the definitions of the Čech Rips complex and the triangle inequality property of the Euclidean metric.

Remark 3.1 *The number of intersections that need to be computed to determine the Čech complex of a point cloud combined with the size of the resultant complex often poses practical difficulties. On the other hand, the ϵ-connectivity graph is more tractable to compute and store, making the Rips complex a more computationally viable option.*

In cases where the point cloud is very large, even the construction of the Rips complex might prove difficult. In such scenarios, other complexes like the weak and strong witness complexes are preferred. These complexes are constructed on a smaller subset of the point cloud called the set of landmark points.

3.2 Reconstructing the Manifold [3]

Definition 3.4 (Hausdorff Distance) *Given two subspaces K, K' of a metric space (X, d), the Hausdorff distance d_H between them is given by*

$$d_H(K, K') = \sup_{w \in X} \left| \inf_{x \in K} d(w, x) - \inf_{y \in K'} d(w, y) \right|.$$

The Hausdorff distance d_H defines a metric on the set of compact subspaces of a metric space (X, d). This can be used to define a distance function between two compact metric spaces as follows. This distance function can be viewed as a measure of how close the compact spaces are to being isometric.

Definition 3.5 (Gromov-Hausdorff Distance) *Let (K_1, d_1) and (K_2, d_2) be two compact metric spaces. The Gromov-Hausdorff distance $d_{GH}(K_1, K_2)$ is the infimum over all $r \geq 0$ such that there exists a metric space (X, d) and compact subsets $C_1, C_2 \subseteq X$ which are isometric to K_1 and K_2 respectively such that $d_H(C_1, C_2) \leq r$.*

Consider the point cloud $X \subseteq R^d$ sampled using the probability measure μ with compact support M. As both X and M are compact subsets of R^d, we can compute the Hausdorff distance between them. Additionally, we can define a map $d_M : R^d \to R^+$ that gives the distance of each point in R^d from the subspace M,

$$d_M(y) = \inf_{m \in M} \|m - y\|_2.$$

Under this setting, the ϵ-offsets of M correspond to the sublevel sets $d_M^{-1}([0, \epsilon])$ of the distance function d_M.

Proposition 3.2 *The distance function d_M as defined, satisfies the following properties:*

1. *d_M is 1-Lipschitz, i.e. $|d_M(y) - d_M(z)| \leq \|y - z\|$*
2. *d_M^2 is semiconcave, i.e. the map $y \mapsto \|y\|^2 - d_M^2(y)$, is convex.*

As a result of these properties, one can define the gradient of the distance function, $\nabla d_M : R^d \to R^d$. In this case, a point $y \in R^d$ is said to be α-critical if $\|\nabla d_M\| \leq \alpha$. For an $\alpha \in (0, 1)$, the α-reach of d_M is the maximum value of R such that $d_M^{-1}((0, R])$ does not contain an α-critical point.

Theorem 3.2 (Reconstruction Theorem [24]) *For X and M in R^d as defined before, let $d_H(X, M) < \epsilon$ and $reach_\alpha(d_M) \leq R$ for some $\alpha > 0$. Then, for any $\beta \in [4\epsilon/\alpha^2, R - 3\epsilon]$ and $\gamma \in (0, R)$, X_β the β-offset of X is homotopically equivalent to M_γ, the γ-offset of M, when $\epsilon \leq \frac{R}{5+4/\alpha^2}$.*

The Reconstruction Theorem establishes that under certain regularity conditions on the manifold, small offsets of the underlying manifold are homotopically equivalent to the β-offset of X for suitably chosen values of β. We also know from using the Nerve Theorem, that this β-offset of X is homotopically equivalent to the Čech complex $C_\beta(X)$ constructed on the point cloud. When the underlying manifold is a compact Riemannian manifold, we also have the following result.

Theorem 3.3 *For a compact Riemannian manifold M, there exists $\eta > 0$ such that for all $\epsilon < \eta$, the ϵ-thickening of the manifold M_ϵ is homotopically equivalent to M.*

These results provide a justification for using Čech complexes built on the point cloud to infer the topology of the underlying manifold.

From a practical perspective, there are a few challenges that need to be addressed before we can utilize these results to analyze data. The validity of the Reconstruction Theorem rests on certain assumptions on the nature of the underlying manifold. Even if we were to work under the assumption that these are satisfied for our dataset, we still need a procedure to determine a suitable value of ϵ for building the Čech complex $C_\epsilon(X)$. Following this, we also need appropriate homotopically invariant objects or descriptors to quantify the topology of this complex.

The Betti numbers [25] associated with the simplicial homology groups of the complexes are a good solution to the second issue due to their relative ease of computation. The other problems, while being more tricky to overcome, can be conveniently circumvented by taking into consideration all possible values of ϵ instead of choosing just one. These ideas form the basis for persistent homology, a central technique in TDA, that is discussed in the next chapter.

4 Advanced Techniques in Topological Data Analysis

In this chapter, we explore advanced techniques in topological data analysis (TDA) [26] [7] [6] that enhance the ability to analyze complex datasets and uncover significant patterns. We begin with Takens' embedding theorem, which transforms univariate time series into higher-dimensional spaces, allowing for a richer representation of the data's dynamics. This is followed by a discussion on persistence diagrams, which capture the birth and death of topological features across different scales, providing a comprehensive summary of the data's shape. We also introduce homological derivatives, a method for measuring changes in topological features over time by computing distances between successive persistence diagrams. Additionally, we examine Betti curves, which track the number of homological features at various dimensions as a function of scale, offering insights into the data's topological complexity. These advanced techniques provide powerful tools for identifying and analyzing significant topological changes in datasets, enabling more effective detection of critical events such as market crashes. This chapter highlights the versatility and efficacy of TDA in various applications, particularly in the realm of financial market analysis.

4.1 Takens' Embedding

Takens' embedding theorem [27] [28] is a fundamental result in dynamical systems theory, which provides a method for reconstructing the state space of a dynamical system from a sequence of observations. This theorem is particularly useful for transforming a univariate time series into a higher-dimensional space, thereby allowing for a richer representation of the data's underlying dynamics.

Definition 4.1 (Takens' Embedding Theorem) *Given a smooth dynamical system with an attractor $\mathcal{A}$ in a m-dimensional manifold, and a smooth observation function $h : \mathcal{A} \to R$, the time-delay embedding map $\Phi : \mathcal{A} \to R^d$ is defined as:*

$$\Phi(x) = (h(x), h(f(x)), h(f^2(x)), \ldots, h(f^{d-1}(x)))$$

where f is the evolution function of the dynamical system, and d is the embedding dimension. According to Takens' theorem, if $d > 2m$, then Φ is an embedding, meaning that Φ is a diffeomorphism on its image, thus preserving the topological properties of the attractor.

Explanation: The theorem essentially states that the geometry and topology of the original state space can be captured by looking at delayed coordinates of a single observable. This allows us to reconstruct the state space of the system from a time series of scalar observations. The choice of time delay τ and embedding dimension d are crucial for the reconstruction to be accurate.

Mathematical Justification: The embedding theorem relies on the Whitney embedding theorem from differential topology, which ensures that a smooth m-dimensional manifold can be embedded in R^{2m}. Takens extended this result to time-delay embeddings of dynamical systems.

Example 4.1 *Given a time series* $[2900, 2910, 2905, 2890, 2880, 2875, 2895]$*, with* $d = 3$ *and* $\tau = 1$*:*

$$\begin{aligned} Y_{t_1} &= [2900, 2910, 2905] \\ Y_{t_2} &= [2910, 2905, 2890] \\ Y_{t_3} &= [2905, 2890, 2880] \\ Y_{t_4} &= [2890, 2880, 2875] \end{aligned}$$

This results in a set of 3-dimensional vectors.

Example 4.2 *Consider a more complex time series that represents daily stock prices:*

$$[100, 102, 101, 105, 107, 106, 110, 111, 115, 117, 116]$$

with $d = 4$ and $\tau = 1$:

$$Y_{t_1} = [100, 102, 101, 105]$$
$$Y_{t_2} = [102, 101, 105, 107]$$
$$Y_{t_3} = [101, 105, 107, 106]$$
$$Y_{t_4} = [105, 107, 106, 110]$$
$$Y_{t_5} = [107, 106, 110, 111]$$
$$Y_{t_6} = [106, 110, 111, 115]$$
$$Y_{t_7} = [110, 111, 115, 117]$$
$$Y_{t_8} = [111, 115, 117, 116]$$

This results in a set of 4-dimensional vectors, providing a more detailed view of the stock price dynamics.

Choosing the Embedding Dimension and Time Delay:

- **Embedding Dimension (d):** The embedding dimension should be chosen such that it captures the complexity of the system. According to Takens' theorem, d should be greater than twice the dimension of the attractor.
- **Time Delay ():** The time delay τ should be chosen to ensure that the coordinates are sufficiently separated in time to provide new information, but not so far apart that they become uncorrelated.

Practical Considerations:

- **False Nearest Neighbors (FNN):** A method for determining the appropriate embedding dimension by examining the proportion of points that have false nearest neighbors in phase space as a function of d.
- **Mutual Information:** A method for selecting the time delay by identifying the time lag that maximizes the mutual information between $y(t)$ and $y(t + \tau)$.

Takens' embedding theorem is a powerful tool for reconstructing the dynamics of a system from time series data. By transforming the data into a higher-dimensional space, it allows for a more comprehensive analysis of the system's behavior and provides the foundation for applying topological data analysis techniques to time series data.

4.2 Persistence Diagrams

Persistence diagrams [29] capture the birth and death of topological features in the data. For each time window, we calculate the persistence diagram using the Vietoris-Rips complex.

Definition 4.2 (Persistence Diagram) *The birth and death of a feature (e.g., a loop) are represented as points (b_i, d_i) in the diagram. The lifespan (persistence) of the feature is $d_i - b_i$.*

Example 4.3 *Consider a persistence diagram with the following points representing features:*

$$\{(1, 4), (2, 5), (3, 6)\}$$

These points indicate the birth and death times of features in the data.

To visualize the persistence diagram, imagine a 2D plot where each point (b_i, d_i) represents the birth and death times of a topological feature. The diagonal line $y = x$ represents features that have zero lifespan. Features further away from the diagonal have longer persistence and are often more significant.

Example 4.4 *Suppose we have a simplicial complex constructed [30] from the following 2D points:* $(0, 0)$, $(1, 0)$, $(0, 1)$, $(1, 1)$. *The Vietoris-Rips complex at different scales can be represented as adjacency matrices.*

At a small scale (e.g., $\epsilon = 0.5$), only the vertices are considered:

$$\begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

At a larger scale (e.g., $\epsilon = 1.5$), edges form between the points:

$$\begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{matrix}$$

At an even larger scale (e.g., $\epsilon = 2.5$), the entire structure becomes a filled square:

$$\begin{matrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{matrix}$$

From these adjacency matrices, we construct the Vietoris-Rips complexes and compute the persistence diagrams.

4.3 Homological Derivatives

We compute the distances between successive persistence diagrams to measure changes in the topological features over time.

Definition 4.3 (Homological Derivative [31]) *The distance between two persistence diagrams D_i and D_{i+1} is given by:*

$$\textit{Distance}(D_i, D_{i+1}) = \sum_j \|(b_j, d_j)_i - (b_j, d_j)_{i+1}\|_p$$

where $\|\cdot\|_p$ denotes the L_p norm.

Example 4.5 *Given two successive persistence diagrams:*

$$D_1 = \{(1, 4), (2, 5)\}$$

$$D_2 = \{(1, 5), (2, 6)\}$$

The distance is calculated as:

$$\begin{aligned} \textit{Distance}(D_1, D_2) &= \sqrt{(1-1)^2 + (4-5)^2} + \sqrt{(2-2)^2 + (5-6)^2} \\ &= \sqrt{0+1} + \sqrt{0+1} \\ &= \sqrt{2} \end{aligned}$$

Example 4.6 *For more complex data, we can represent persistence diagrams as matrices where rows correspond to different features, and columns correspond to their birth and death times. For example:*

$$D_1 = \begin{pmatrix} 1 & 4 \\ 2 & 5 \end{pmatrix}$$

$$D_2 = \begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix}$$

We calculate the distance between these matrices using an appropriate norm.

These advanced techniques in TDA provide powerful tools for identifying and analyzing significant topological changes in datasets, enabling more effective detection of critical events such as market crashes. This chapter highlights the versatility and efficacy of TDA in various applications, particularly in the realm of financial market analysis.

5 Vectorized Representations of Persistent Homology

Persistent homology is a powerful tool in topological data analysis that provides insights into the shape and structure of data. To make persistent homology more applicable in machine learning and statistical analysis, vectorized representations [32] are often used. This chapter explores various vectorized representations of persistent homology, including real-valued summaries, persistence images, Betti curves, entropy summary functions, and persistence landscapes.

5.1 Real-valued Summaries

Real-valued summaries of persistent homology are essential for simplifying the interpretation of topological features. These summaries convert the detailed information in persistence diagrams into numerical values that can be easily analyzed and compared.

Definition 5.1 (Birth and Death Statistics) *One common approach is to calculate basic statistics on the birth and death [33] times of features. For example, the mean and variance of birth and death times can provide insights into the average lifespan and variability of topological features.*

Definition 5.2 (Total Persistence) *Another important summary is total persistence [33], which is the sum of the lifetimes of all features in the persistence diagram. This measure captures the overall topological complexity of the data.*

$$\textit{Total Persistence} = \sum_{i=1}^{n} (d_i - b_i)$$

where b_i and d_i are the birth and death times of the i-th feature, respectively.

Example 5.1 (Applications of Real-valued Summaries) *Real-valued summaries are widely used in various applications, such as shape analysis in computer vision, where they help distinguish between different shapes based on their topological features. In bioinformatics, these summaries can be used to compare the structure of biological molecules.*

5.2 Persistence Image

The persistence image [34] is a two-dimensional representation of persistence diagrams. This approach transforms the persistence diagram into a grid where each cell's value is determined by the persistence of features within that cell. This transformation allows for the application of machine learning techniques on the resulting images.

Definition 5.3 (Construction of Persistence Images) *To construct a persistence image, the persistence diagram is first divided into a grid. Each point in the diagram is mapped to a corresponding cell in the grid, and the value of each cell is calculated based on the persistence of the points it contains. Typically, Gaussian functions are used to smooth the values across the grid.*

$$I(x, y) = \sum_{i=1}^{n} \exp\left(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma^2} \right)$$

where (x_i, y_i) are the coordinates of points in the persistence diagram and σ is a smoothing parameter.

Example 5.2 (Applications in Machine Learning) *Persistence images are particularly useful in machine learning because they enable the use of convolutional neural networks (CNNs) [35] and other image-based techniques. This approach has been applied to various tasks, including shape classification and time series analysis.*

Example 5.3 (Visualization of Persistence Images) *Examples of persistence images can be found in numerous studies. Visualization of these images helps in understanding the distribution and importance of topological features within the data.*

5.3 Betti Curves

Betti curves [36], denoted as $\beta_k(t)$, represent the number of k-dimensional homological features (e.g., connected components, loops) present in the data at a given scale t. These curves provide insights into the topological complexity of the data across different scales.

Definition 5.4 (Betti Curves) *The Betti curve for dimension k is a function that maps each scale t to the number of k-dimensional features present at that scale. These curves are derived from the persistence diagram by counting the number of features that are born and persist at each scale.*

$$\beta_k(t) = \sum_{i=1}^{n} \mathbf{1}_{[b_i, d_i)}(t)$$

where $\mathbf{1}_{[b_i,d_i)}$ is the indicator function that is 1 if t is within the interval $[b_i, d_i)$ and 0 otherwise.

Example 5.4 (Interpretation of Betti Curves) *Betti curves are useful for understanding the evolution of topological features over different scales. Peaks in the Betti curve indicate scales where a large number of features are present, which may correspond to significant structures in the data.*

Example 5.5 (Applications of Betti Curves) *Betti curves have applications in various fields, including materials science, where they help analyze the porosity and connectivity of materials, and in neuroscience, where they are used to study the connectivity of brain networks.*

5.4 Entropy Summary Function

The entropy summary function [37] measures the uncertainty or randomness in the persistence diagrams. It quantifies the spread of feature lifetimes, providing a single numerical value that captures the complexity of the topological features in the data.

Definition 5.5 (Entropy Summary Function) *The entropy of a persistence diagram is calculated by considering the lifetimes of the features. The entropy summary function E is defined as:*

$$E = -\sum_{i=1}^{n} p_i \log p_i$$

where p_i represents the normalized persistence of each feature, calculated as:

$$p_i = \frac{d_i - b_i}{\sum_{j=1}^{n}(d_j - b_j)}$$

Example 5.6 (Interpretation of Entropy) *Entropy captures the diversity of topological features in the data. Higher entropy values indicate a wider range of feature lifetimes, suggesting more complex structures. Lower entropy values suggest a more uniform distribution of feature lifetimes.*

Example 5.7 (Applications of Entropy Summary Functions) *Entropy summary functions are used in various fields, such as signal processing and texture analysis. In these applications, entropy helps quantify the complexity and heterogeneity of the data.*

5.5 Persistence Landscapes

Persistence landscapes [38] are vectorized representations of persistence diagrams that enable statistical analysis and comparison. A persistence landscape is a collection of piecewise linear functions that summarize the topological features' persistence in the data.

Definition 5.6 (Persistence Landscapes) *A persistence landscape λ is defined for each feature in the persistence diagram. It is a piecewise linear function that records the persistence of features at different scales. The landscape function for a feature born at b and dying at d is defined as:*

$$\lambda_k(t) = \max(0, \min(t - b, d - t))$$

for each dimension k. The overall persistence landscape is the collection of these functions for all features in the persistence diagram.

Example 5.8 (Statistical Analysis with Persistence Landscapes) *Persistence landscapes allow for the application of traditional statistical methods to topological data. For example, the mean and variance of persistence landscapes can be calculated, enabling comparisons between different data sets.*

Example 5.9 (Applications of Persistence Landscapes) *Persistence landscapes are used in various applications, including time series analysis and classification tasks. They provide a robust and interpretable way to analyze and compare the topological features of different data sets.*

6 Methodology

This chapter outlines the methodology employed to analyze and detect significant patterns in financial market data using Topological Data Analysis (TDA). The process begins with data collection and preprocessing, followed by the application of TDA techniques to uncover the topological features of the data. We also describe the methods used for visualization and interpretation of the results, providing a comprehensive approach to understanding market behavior and potential crash indicators.

6.1 Data Source and Preparation

6.1.1 Data Collection

The primary data source for this study is Yahoo Finance, accessed using the `yfinance` library in Python. We focused on historical data for major financial indices such as the (GSPC), GBP/USD, and NASDAQ, covering the period from their inception to May 2024. The 'Close' price was chosen for analysis as it reflects the final trading sentiment of the market each day.

```
import yfinance as yf
import pandas as pd
import matplotlib.pyplot as plt

# Load S&P 500 historical data
SP500 = yf.Ticker("^GSPC")
sp500_df = SP500.history(period="max")
```

6.2 Data Preprocessing

To ensure the data is suitable for time series analysis, it was necessary to address gaps caused by weekends and holidays. We resampled the data to obtain evenly spaced daily values, using forward fill to handle missing entries. This preprocessing step is crucial for the accurate application of time series techniques.

```
# Resample the data to ensure evenly spaced daily values
start_year = '1980'
price_resampled_df = sp500_df['Close'].resample('D').ffill()
price_resampled_df = price_resampled_df[start_year:]

# Plot the resampled data
plt.plot(price_resampled_df)
plt.title('Price of S&P 500 (Resampled)')
plt.xlabel('Date')
plt.ylabel('Close Price')
plt.show()
```

6.3 Takens' Embedding and Point Clouds

6.3.1 Takens' Embedding Theorem

Takens' embedding theorem is a fundamental result in the theory of dynamical systems. It states that the dynamics of a smooth, deterministic system can be reconstructed from a sequence of observations of a single variable. This is achieved by constructing a delay embedding, where the system's state space is reconstructed using time-delayed copies of the observed variable.

Theorem 6.1 (Takens' Embedding Theorem) *Let M be a compact manifold of dimension d. For pairs (ϕ, h) where $\phi : M \to M$ is a smooth map and $h : M \to R$ is a smooth function, it is a generic property that the map $F : M \to R^{2d+1}$ defined by $F(x) = (h(x), h(\phi(x)), \ldots, h(\phi^{2d}(x)))$ is an embedding.*

This theorem facilitates the transformation of a univariate time series [39] into a higher-dimensional space, allowing for the extraction of topological features.

```
from sklearn.preprocessing import StandardScaler
import numpy as np

# Standardize the time series data
scaler = StandardScaler()
scaled_prices = scaler.fit_transform(
    price_resampled_df.values.reshape(-1, 1)
)

# Apply Takens' embedding
embedding_dimension = 3
time_delay = 2
embedded_data = np.array([
    scaled_prices[i:i + embedding_dimension * time_delay:time_delay].flatten()
    for i in range(len(scaled_prices) - embedding_dimension * time_delay + 1)
])
```

6.3.2 Complex Time Series Embedding

Consider a more complex time series that represents daily stock prices:

$$[100, 102, 101, 105, 107, 106, 110, 111, 115, 117, 116]$$

with $d = 4$ and $\tau = 1$:

$$\begin{aligned}
Y_{t_1} &= [100, 102, 101, 105] \\
Y_{t_2} &= [102, 101, 105, 107] \\
Y_{t_3} &= [101, 105, 107, 106] \\
Y_{t_4} &= [105, 107, 106, 110] \\
Y_{t_5} &= [107, 106, 110, 111] \\
Y_{t_6} &= [106, 110, 111, 115] \\
Y_{t_7} &= [110, 111, 115, 117] \\
Y_{t_8} &= [111, 115, 117, 116]
\end{aligned}$$

This results in a set of 4-dimensional vectors, providing a more detailed view of the stock price dynamics.

6.3.3 Choosing the Embedding Dimension and Time Delay

- **Embedding Dimension (d):** The embedding dimension should be chosen such that it captures the complexity of the system. According to Takens' theorem, d should be greater than twice the dimension of the attractor.
- **Time Delay ():** The time delay τ should be chosen to ensure that the coordinates are sufficiently separated in time to provide new information, but not so far apart that they become uncorrelated.

6.3.4 Practical Considerations

- **False Nearest Neighbors (FNN) [40]:** A method for determining the appropriate embedding dimension by examining the proportion of points that have false nearest neighbors in phase space as a function of d.
- **Mutual Information:** A method for selecting the time delay by identifying the time lag that maximizes the mutual information between $y(t)$ and $y(t + \tau)$.

Takens' embedding theorem is a powerful tool for reconstructing the dynamics of a system from time series data. By transforming the data into a higher-dimensional space, it allows for a more comprehensive analysis of the system's behavior and provides the foundation for applying topological data analysis techniques to time series data.

6.4 Baseline Model: First Derivative

As a baseline for comparison, we computed the first derivative of the time series over a sliding window. The first derivative measures the rate of change in price, highlighting regions of significant changes and potential market volatility.

```
# Compute the first derivative of the time series
window_size = 31
sliding_window = np.array([
    scaled_prices[i:i + window_size]
    for i in range(len(scaled_prices) - window_size + 1)
])
first_derivative = np.diff(sliding_window, axis=1)

# Plot the first derivative
plt.plot(first_derivative)
plt.title('First Derivative of S&P 500 Close Prices')
plt.xlabel('Time')
plt.ylabel('First Derivative')
plt.show()
```

6.5 Topological Data Analysis (TDA)

6.5.1 Computation of Persistence Diagrams

Topological Data Analysis involves the calculation of persistence diagrams, which summarize the topological features of the point clouds derived from the embedded time series. These diagrams capture the birth and death of homological features as the filtration parameter varies.

Theorem 6.2 (Persistence of Homology) *Let K be a filtered simplicial complex, i.e., a sequence of subcomplexes $\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K$. The p-persistent k-th homology group of K at index i is defined as $H_k^{i,p}(K) = \mathrm{Im}(H_k(K_i) \to H_k(K_{i+p}))$, where H_k denotes the k-th homology group.*

This theorem underpins the computation of persistence diagrams, enabling the tracking of topological features across different scales.

```
import persim
from ripser import ripser

# Compute persistence diagrams
diagrams = ripser(embedded_data)['dgms']

# Visualize a persistence diagram
persim.plot_diagrams(diagrams, show=True)
```

6.6 Homological Derivatives

Homological derivatives provide a sophisticated method for analyzing the evolution of topological features in time series data. By calculating the distances between successive persistence diagrams, we obtain a detailed measure of how topological features change over time, helping to identify significant events such as market crashes.

Definition 6.1 (Wasserstein Distance) *The Wasserstein distance [41] between two persistence diagrams D_1 and D_2 is defined as the minimum cost of matching points in D_1 with points in D_2, where the cost is given by the p-th power of the Euclidean distance between matched points, summed over all matched points, and minimized over all possible matchings.*

6.6.1 Steps for Calculating Homological Derivatives

1. **Embedding Time Series**: Transform the univariate time series into higher-dimensional space using Takens' embedding.
2. **Sliding Window**: Apply a sliding window to the embedded data to generate overlapping point clouds, capturing local behaviors over time.
3. **Persistence Diagrams**: Compute persistence diagrams for each point cloud, representing the topological features.
4. **Distance Calculation**: Calculate the distance between successive persistence diagrams to quantify changes in topological features.

```
from persim import wasserstein

# Calculate distances between successive persistence diagrams
distances = [wasserstein(diagrams[i], diagrams[i + 1]) for i in range(len(diagrams) - 1)]

# Plot the distances (homological derivatives)
plt.plot(distances)
plt.title('Homological Derivatives of S&P 500 Data')
plt.xlabel('Time')
plt.ylabel('Wasserstein Distance')
plt.show()
```

By tracking homological derivatives, we can detect significant changes in the market's structure. Peaks in the derivative plot often correspond to periods of substantial topological change, which may indicate market crashes. This approach provides a nuanced and detailed analysis compared to traditional financial metrics, offering a powerful tool for early warning of market instability. In summary, this comprehensive methodology leverages Topological Data Analysis to capture and analyze the topological characteristics of financial market data. By combining TDA with traditional time series analysis techniques, we gain deeper insights into market behavior and potential crash indicators, enhancing our ability to predict and manage financial risks.

7 Analysis Results

7.1 Close Price Over Time

The close price of the GSPC data from 1980 to 2024 shows significant events like the dot-com bubble and the 2008 financial crisis. These events are marked by sharp declines in the index, which we aim to detect using TDA.

Figure 5: Close Price of the GSPC data from 1980 to 2024

In Figure 4, the y-axis represents the close price of the GSPC data, and the x-axis represents the date. The plot shows a general upward trend over the years, with noticeable dips during significant market events. The sharp declines around the year 2000 and 2008 correspond to the dot-com bubble and the financial crisis, respectively. These events are critical for our crash detection analysis.

7.2 First Derivative

The first derivative of the close price highlights regions with significant changes. Peaks in this graph correspond to periods of high volatility, which can indicate market crashes.

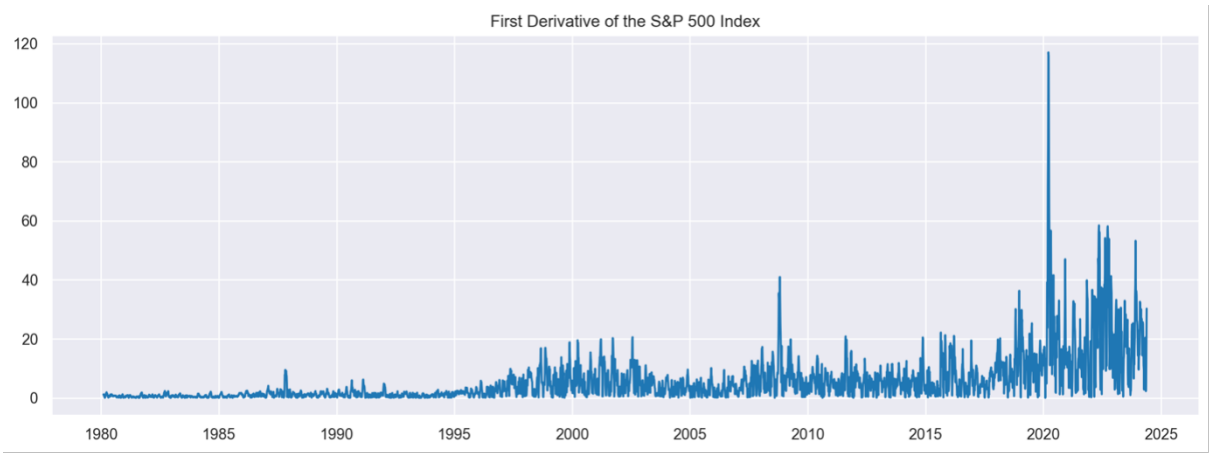

Figure 6: First Derivative of the GSPC data

In Figure 5, the y-axis represents the magnitude of the first derivative, and the x-axis represents the date. The peaks indicate periods of rapid changes in the close price, suggesting high volatility. Notable peaks around the years 2000 and 2008 align with the significant market downturns observed in Figure 1. The first derivative serves as a baseline measure for detecting market crashes.

7.3 Crash Probability Based on First Derivative

Figure 3 shows the crash probability based on the first derivative. The left graph displays the crash probability over time, with a threshold set at 0.3. The right graph highlights periods where the crash probability exceeded this threshold, indicating potential market crashes.

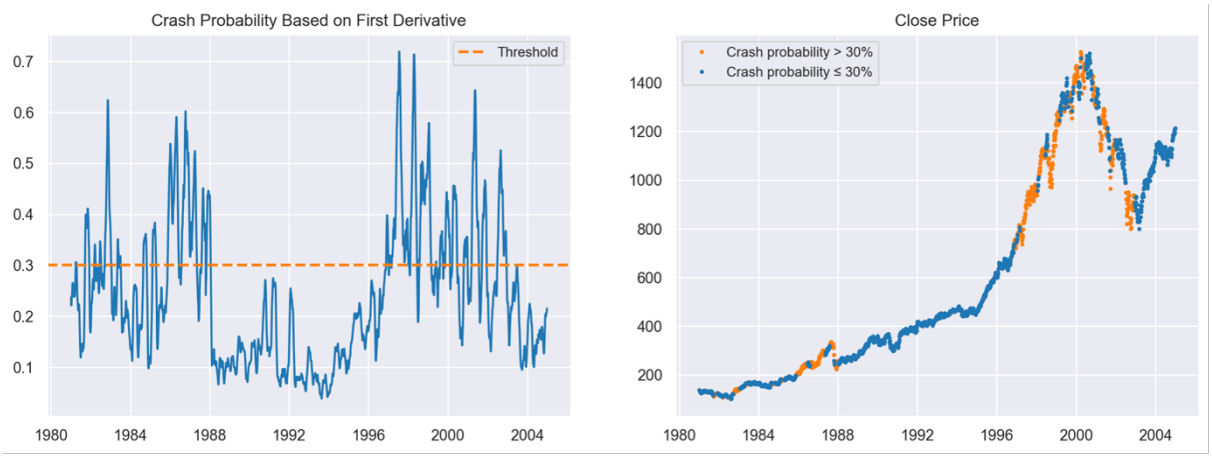

Figure 7: Crash Probability Based on First Derivative

In Figure 6, the left graph's y-axis represents the crash probability, and the x-axis represents the date. The threshold line at 0.3 separates periods of high crash probability. The right graph plots the close price with points marked in orange where the crash probability exceeds 0.3, highlighting potential crash periods. This visualization helps in understanding the effectiveness of the first derivative in predicting market crashes.

7.4 Landscape Distances

The landscape distances measure the change in topological features between successive windows. Peaks in this graph correspond to significant changes in the market, potentially indicating crashes.

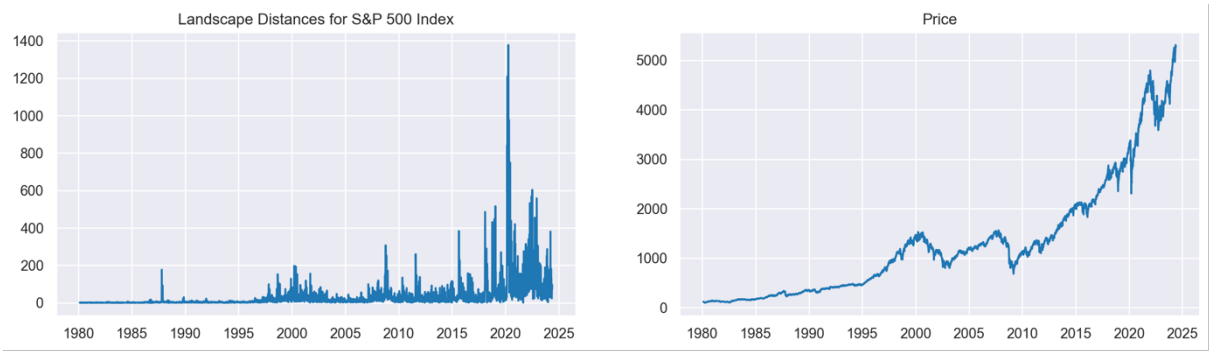

Figure 8: Landscape Distances for the GSPC data

In Figure 7, the y-axis represents the landscape distance, and the x-axis represents the date. The peaks indicate significant topological changes in the point clouds, suggesting potential market crashes. Notable peaks around 2000 and 2008 align with the market downturns observed in previous figures. Landscape distances provide a more nuanced measure of market changes compared to the first derivative.

7.5 Norm of L^p Difference

The Norm of L^p difference provides another measure of topological change. This method also identifies significant events in the market's history.

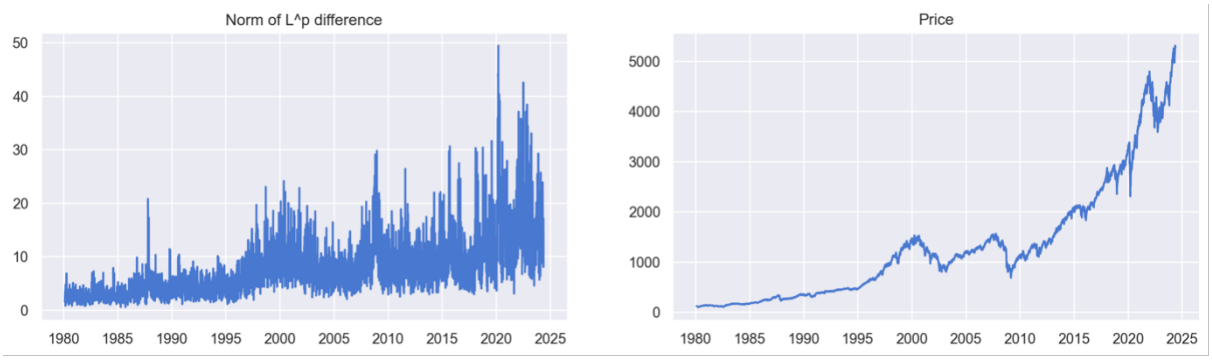

Figure 9: Norm of L^p Difference for the GSPC data

In Figure 8, the y-axis represents the norm of L^p differences, and the x-axis represents the date. The peaks indicate significant topological changes, similar to the landscape distances. The measure captures notable events such as the dot-com bubble and the financial crisis, indicating its effectiveness in detecting market crashes.

7.6 Landscape and Betti Curve Distances

Figure 6 compares the landscape and Betti curve distances [38]. Both graphs highlight significant market events, with peaks corresponding to periods of substantial topological change.

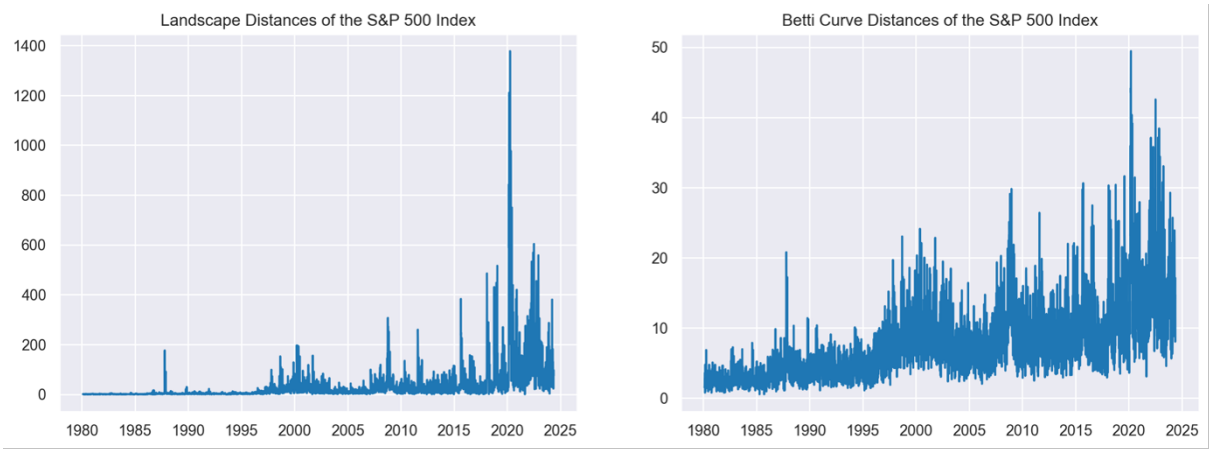

Figure 10: Landscape and Betti Curve Distances for the GSPC data

In Figure 9, the left graph shows the landscape distances, and the right graph shows the Betti curve distances. Both measures capture significant market events, such as the dot-com bubble and the financial crisis. The comparison highlights the consistency of TDA measures in detecting market crashes.

7.7 Crash Probability Based on Landscape Distance

The following figures present the crash probability based on landscape distance. The left graph shows the crash probability over time, with a threshold set at 0.3. The right graph highlights periods where the crash probability exceeded this threshold, indicating potential market crashes.

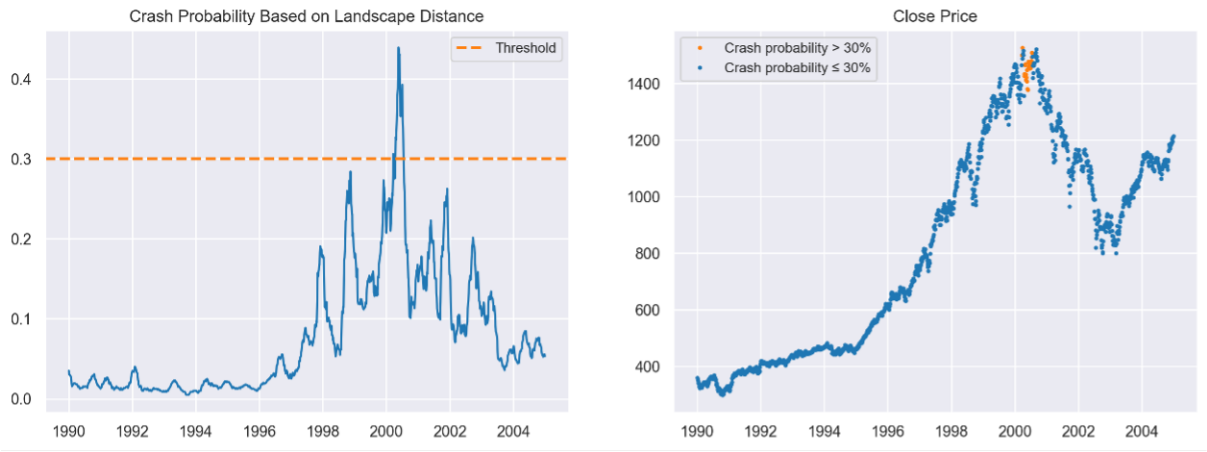

Figure 11: Crash Probability Based on Landscape Distance (GSPC)

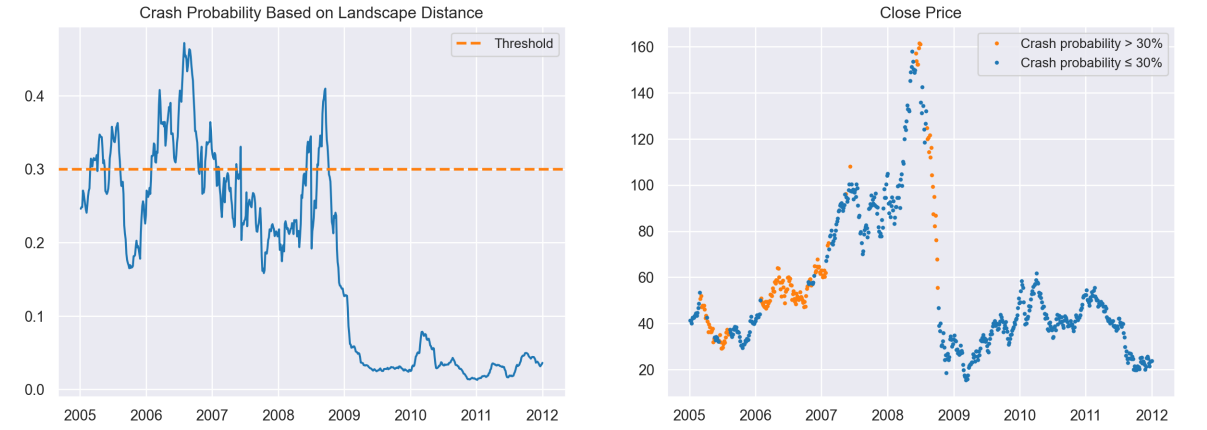

Figure 12: Crash Probability Based on Landscape Distance (GPBUSD)

7.8 Comparison of Methods

In this analysis, we compare the effectiveness of a baseline detector and a topological detector in identifying market crashes. The topological detector leverages Topological Data Analysis (TDA) techniques, which provide a unique perspective by examining the shape and structure of the data in high-dimensional spaces.

Baseline Detector

The baseline detector uses traditional methods to calculate crash probabilities. The approach involves measuring the first derivative of the market index to capture significant changes and volatility. The peaks in the first derivative graph correspond to periods of high volatility, which can indicate potential market crashes.

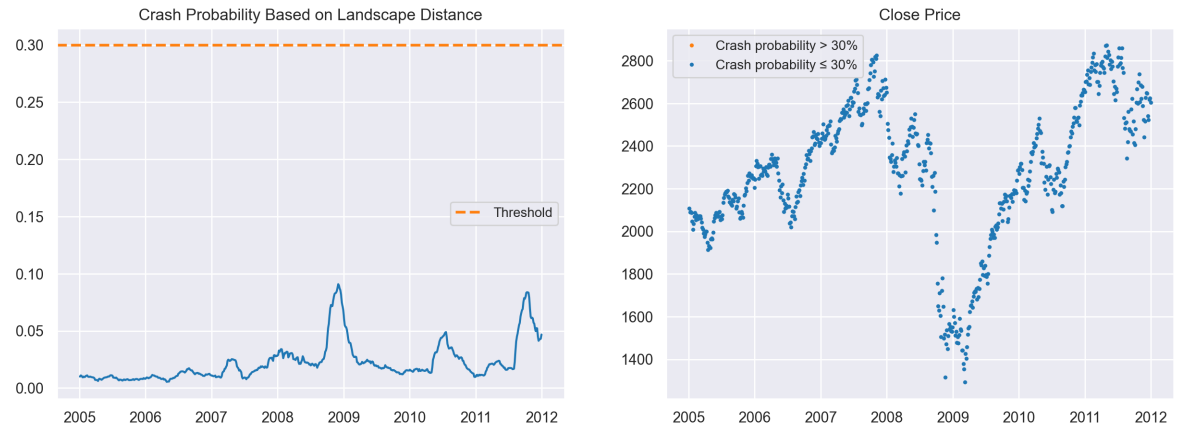

Figure 13: Crash Probability Based on Landscape Distance (NASDAQ)

Topological Detector

The topological detector employs TDA to analyze the market index. This approach involves constructing simplicial complexes and computing persistence diagrams to capture the topological features of the data. The persistence diagrams track the birth and death of topological features (e.g., connected components, loops) across different scales.

Detailed Analysis of Major Market Crashes

Dot-com Crash (March 11, 2000 – October 9, 2004) The topological detector identified significant changes in the market's topological structure during the dot-com crash. The baseline detector also captured this period, but the topological approach provided a more nuanced view.

Subprime Mortgage Crisis (December 2007 – June 2009) Both detectors identified the subprime mortgage crisis. The topological detector was more robust to noise and provided clearer indications of the market downturn.

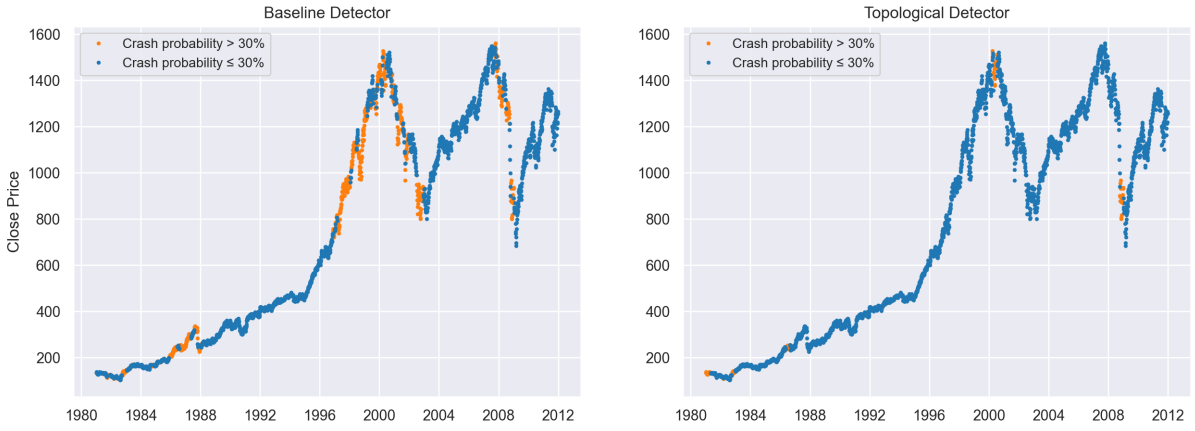

Figure 14: Comparison of Baseline and Topological Detectors (GSPC)

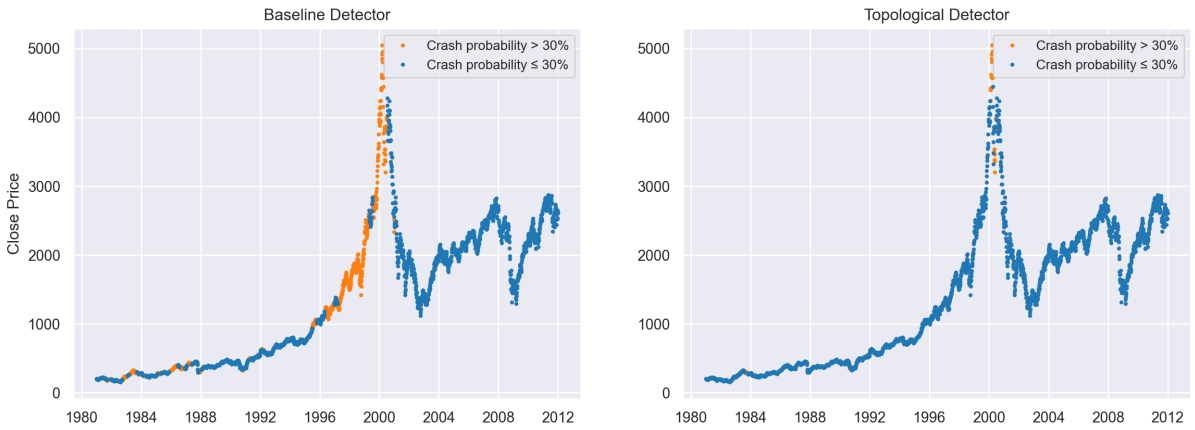

Figure 15: Comparison of Baseline and Topological Detectors (NASDAQ)

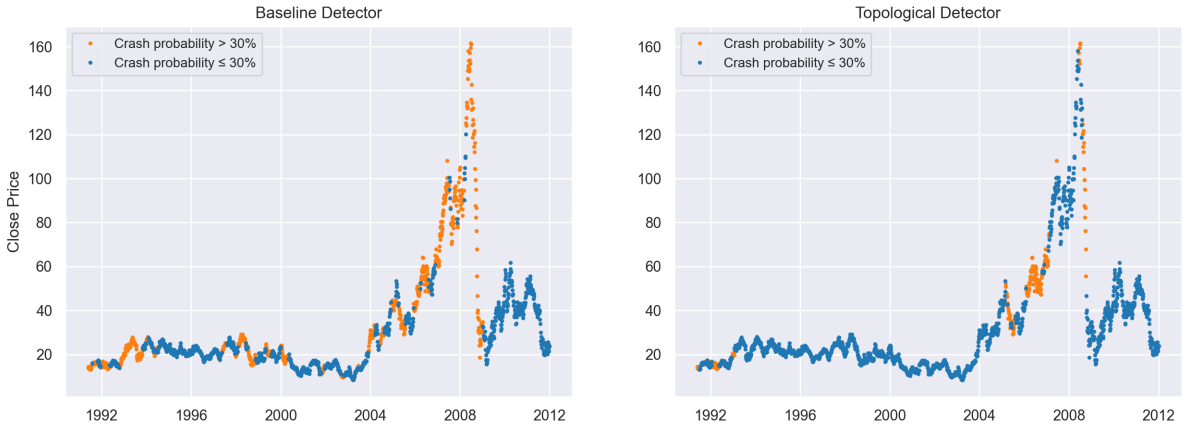

Figure 16: Comparison of Baseline and Topological Detectors (GPBUSD)

8 Conclusion

This study demonstrates the effectiveness of Topological Data Analysis (TDA) in detecting stock market crashes. TDA provides a unique approach by examining the shape and structure of data in high-dimensional spaces, revealing patterns that traditional financial metrics may miss. By applying TDA to the historical data of the GSPC from 1980 to 2024, we were able to identify significant market downturns, including the dot-com bubble and the 2008 financial crisis. The methodology involved transforming the time series data into point clouds using Takens' embedding, followed by the calculation of persistence diagrams to summarize the topological features of these point clouds. We then measured the distances between successive persistence diagrams to track the evolution of these features over time. This approach allowed us to detect significant changes in the market's structure, providing early warning signals of potential crashes.

The results of the analysis showed that TDA could effectively capture significant changes in the market, offering a more nuanced view compared to traditional methods. Specifically, the landscape and Betti curve distances proved to be reliable indicators of market volatility, often highlighting periods of instability before traditional financial metrics. When compared to the first derivative of the time series, which served as our baseline model, TDA offered a more comprehensive understanding of market dynamics. Moreover, this study revealed that TDA could complement traditional methods by providing additional layers of analysis. While the first derivative indicated periods of high volatility, it sometimes failed to capture subtle shifts in the market's structure that TDA could detect. This highlights the potential for integrating TDA with conventional crash detection models to enhance their accuracy and robustness.

However, it is important to acknowledge the limitations of the study. TDA is computationally in-

tensive and requires significant resources, which may limit its practical application in real-time trading environments. Additionally, the complexity of TDA might pose challenges for financial analysts unfamiliar with topological concepts. Future research should focus on improving the efficiency of TDA algorithms and exploring ways to make them more accessible to practitioners. Furthermore, applying TDA to other financial indices and markets could provide additional validation of its effectiveness in crash detection. Expanding the scope of this research to include different types of assets, such as commodities or cryptocurrencies, could offer valuable insights into the broader applicability of TDA in financial analysis.

In conclusion, this study underscores the potential of Topological Data Analysis as a powerful tool for detecting stock market crashes. By revealing hidden patterns and structures within financial data, TDA can provide early warning signals of market instability, complementing traditional methods and offering a more comprehensive understanding of market dynamics. As the financial landscape continues to evolve, integrating advanced analytical techniques like TDA could play a crucial role in enhancing market prediction models and mitigating the risks associated with market crashes.

9 Summary

Detecting stock market crashes is crucial for investors and policymakers. Traditional methods, while useful, may not fully capture the complexity of market behaviors. Topological Data Analysis (TDA) offers a promising alternative by examining the shape of data in higher-dimensional spaces. This study applied TDA to the GSPC, demonstrating its effectiveness in identifying significant market events. By comparing TDA with traditional methods, we highlighted its ability to capture subtle changes in the market's structure. The use of TDA in financial markets is a relatively new area of research. This study shows that TDA can provide valuable insights into market dynamics, complementing traditional methods. However, there are limitations to consider, such as the complexity of TDA and the computational resources required. Future research could focus on improving the efficiency of TDA algorithms and exploring their integration with other analytical techniques. Additionally, applying TDA to other financial indices and markets could provide further validation of its effectiveness in crash detection.

References

[1] R. GORDON, "Graded artin algebras," 11 1982.

[2] S. Basu, N. Karisani, and L. Parida, "Sequents, barcodes, and homology," 2022.

[3] S. Faigenbaum-Golovin and D. Levin, "Manifold reconstruction and denoising from scattered data in high dimension," *Journal of Computational and Applied Mathematics*, vol. 421, p. 114818, Mar. 2023.

[4] D. Shah, H. Isah, and F. Zulkernine, "Stock market analysis: A review and taxonomy of prediction techniques," *International Journal of Financial Studies*, vol. 7, p. 26, 05 2019.

[5] K. El-Wassal, "The development of stock markets: In search of a theory," *International Journal of Economics and Financial Issues*, vol. 3, pp. 607–624, 09 2013.

[6] P. Yen and S. A. Cheong, "Using topological data analysis (tda) and persistent homology to analyze the stock markets in singapore and taiwan," *Frontiers in Physics*, vol. 9, p. 572216, 03 2021.

[7] B. Pujiarto, "With topological data analysis, predicting stock market crashes," *IJIIS: International Journal of Informatics and Information Systems*, vol. 4, pp. 63–70, 03 2021.

[8] W. L. D. Oliveira, L. Tunstall, U. Lupo, and A. Medina-Mardones, "Detecting stock market crashes with topological data analysis," *Towards Data Science*, November 2019. 7 min read.

[9] W. Guo and A. G. Banerjee, "Identification of key features using topological data analysis for accurate prediction of manufacturing system outputs," *Journal of Manufacturing Systems*, vol. 43, pp. 225–234, 2017. High Performance Computing and Data Analytics for Cyber Manufacturing.

[10] P. T.-W. Yen and S. A. Cheong, "Using topological data analysis (tda) and persistent homology to analyze the stock markets in singapore and taiwan," *Frontiers in Physics*, vol. 9, 2021.

[11] C. Chang and H. Lin, "A topological based feature extraction method for the stock market," *Data Science in Finance and Economics*, vol. 3, no. 3, pp. 208–229, 2023.

[12] N. Guglielmi, A. Savostianov, and F. Tudisco, "Quantifying the structural stability of simplicial homology," *Journal of Scientific Computing*, vol. 97, Aug. 2023.

[13] C. Malin, "An elementary proof of the homotopy invariance of stabilized configuration spaces," 2022.

[14] M. Refai and R. Abu-Dawwas, "On properties of graded rings and graded modules," *Proyecciones (Antofagasta)*, vol. 41, 11 2022.

[15] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete and Computational Geometry*, vol. 33, pp. 249–274, 02 2005.

[16] N. Saleh, T. Titz Mite, and S. Witzel, "Vietoris–rips complexes of platonic solids," *Innovations in Incidence Geometry: Algebraic, Topological and Combinatorial*, vol. 21, p. 17–34, May 2024.

[17] M. Guerra, A. De Gregorio, U. Fugacci, G. Petri, and F. Vaccarino, "Homological scaffold via minimal homology bases," *Scientific Reports*, vol. 11, p. 5355, 03 2021.

[18] D. R. Sheehy, "Linear-size approximations to the vietoris-rips filtration," 2013.

[19] M. E. Aktas, E. Akbas, and A. E. Fatmaoui, "Persistence homology of networks: methods and applications," *Applied Network Science*, vol. 4, no. 1, p. 61, 2019.

[20] R. Ghrist, "Barcodes: The persistent topology of data," *BULLETIN (New Series) OF THE AMERICAN MATHEMATICAL SOCIETY*, vol. 45, 02 2008.

[21] M. Ritter, "Geometric reconstruction and visualization of point clouds by second order tensor neighborhoods," 02 2021.

[22] U. Bauer, M. Kerber, F. Roll, and A. Rolle, "A unified view on the functorial nerve theorem and its variations," *Expositiones Mathematicae*, vol. 41, p. 125503, Dec. 2023.

[23] S. Dantchev and I. Ivrissimtzis, "Efficient construction of the Čech complex," *Computers Graphics*, vol. 36, no. 6, pp. 708–713, 2012. 2011 Joint Symposium on Computational Aesthetics (CAe), Non-Photorealistic Animation and Rendering (NPAR), and Sketch-Based Interfaces and Modeling (SBIM).

[24] F. Chazal and B. Michel, "An introduction to topological data analysis: Fundamental and practical aspects for data scientists," *Frontiers in Artificial Intelligence*, vol. 4, 2021.

[25] M. Kahle and E. Meckes, "Limit theorems for betti numbers of random simplicial complexes," 2011.

[26] F. Chazal and B. Michel, "An introduction to topological data analysis: fundamental and practical aspects for data scientists," 2021.

[27] K. Kim and J.-H. Jung, "Exact multi-parameter persistent homology of time-series data: Fast and variable topological inferences," 2024.

[28] J. Robinson, "A topological delay embedding theorem for infinite-dimensional dynamical systems," *Nonlinearity*, vol. 18, p. 2135, 07 2005.

[29] A. Elchesen, I. Hartsock, J. A. Perea, and T. Rask, "Learning on persistence diagrams as radon measures," 2022.

[30] M. Hajij, G. Zamzmi, T. Papamarkou, V. Maroulas, and X. Cai, "Simplicial complex representation learning," 2022.

[31] V. Pérez and C. Miranda-Neto, "Homological aspects of derivation modules and critical case of the herzog–vasconcelos conjecture," *Collectanea Mathematica*, vol. 73, pp. 203–219, 04 2022.

[32] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, "Persistence images: A stable vector representation of persistent homology," *Journal of Machine Learning Research*, vol. 18, no. 8, pp. 1–35, 2017.

[33] G. Carlsson and A. Zomorodian, "The theory of multidimensional persistence," *Discrete and Computational Geometry*, vol. 42, pp. 71–93, 06 2007.

[34] A. Asaad, *Persistent Homology for Image Analysis*. PhD thesis, 01 2020.

[35] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015.

[36] Guzel and A. Kaygun, "Classification of stochastic processes with topological data analysis," 06 2022.

[37] N. Atienza, R. Gonzalez-Diaz, and M. Soriano-Trigueros, "A new entropy based summary function for topological data analysis," *Electronic Notes in Discrete Mathematics*, vol. 68, pp. 113–118, 07 2018.

[38] P. Bubenik and P. Dłotko, "A persistence landscapes toolbox for topological statistics," *Journal of Symbolic Computation*, vol. 78, 12 2014.

[39] G. Papacharalampous, H. Tyralis, and D. Koutsoyiannis, "Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: a multiple-case study from greece," *Water Resources Management*, pp. 1–33, 12 2018.

[40] A. Krakovská, K. Mezeiová, and H. Budáčová, "Use of false nearest neighbours for selecting variables and embedding parameters for state space reconstruction," *Journal of Complex Systems*, vol. 2015, p. 12, 03 2015.

[41] A. Stéphanovitch, U. Tanielian, B. Cadre, N. Klutchnikoff, and G. Biau, "Optimal 1-wasserstein distance for wgans," 2023.