UNIVERSITY OF OXFORD

Multiparameter Persistent Homology of Data

Oliver Vipond

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2021

Contents

To my parents, family and friends for all their love and support.

Abstract

We explore two distinct topics in the field of topological data analysis: invariants and metrics for multiparameter persistence modules, and the homology of random geometric simplicial complexes.

We define a computable, stable invariant for multiparameter persistence modules, the *multiparameter persistence landscape*, and exemplify this invariant to be sensitive to the topology and geometry of multifiltered data sets. We prove a local bi-Lipschitz equivalence between two well-studied metrics for multiparameter persistence modules: *the interleaving distance* and *the matching distance*. A consequence of this equivalence result is that the *multiparameter persistence landscape* is a locally complete invariant for finitely presented multiparameter persistence modules.

Finally we explore the asymptotic properties of Čech complexes built on compact Riemannian manifolds with non-empty boundary. We attain homological connectivity thresholds with identical leading terms. An upper threshold above which the Čech complex has homology isomorphic to the homology of the underlying manifold with high probability, and a lower threshold beneath which with high probability it does not.

1
Overview

Let us first give an overview of the work contained in this thesis. We shall give a self-contained high level description of the contents of each chapter, the author's contribution to the work, and the significance of this work to the author's field of research.

Contents

This thesis is structured as follows: Chapter 2, Chapter 3 and Chapter 4, each contain a paper written during the author's PhD studies. The paper associated to Chapter 2 has been published in the Journal of Machine Learning Research, and the papers associated to Chapter 3 and Chapter 4 are both under review at the first journal to which they were submitted. The author has also worked on a pair of computational papers[1] during his PhD which in the interest of length have not been included in this thesis.

[1] *A topological selection of folding pathways from native states of knotted proteins* (1), *Multiparameter persistent homology landscapes identify spatial patterns of immune cells in tumors* (2).

Research Context

The author's research lies at the intersection of Topology, Geometry and Data Science in a field called Topological Data Analysis (TDA). Topological Data Analysis encompasses an emerging set of analytic tools inspired by concepts from algebraic topology which leverage the underlying shape data form to provide new insights.

Persistent homology is a commonly used tool in the TDA toolkit. For those less familiar with persistent homology the author has written interactive slides[2] which give an intuitive introduction to 1-parameter and multiparameter persistent homology, and a demonstration of the code written for the paper associated to Chapter 2.

[2] These slides are best viewed on a PC rather than a small touch screen device.

Persistent homology studies the homology groups of a family of topological spaces built upon data. Homology is an invariant from algebraic topology which detects different dimensional holes in a

topological space. Degree zero homology detects connected components, degree one homology detects loops, degree two homology detects voids and higher degree homology detects higher dimensional cavities.

Persistent homology tracks how these homological feature persist throughout the family of topological spaces. The topological summary for a 1-parameter family is a barcode which marks the parameter values for births and deaths of homological features. The barcode can be thought of as a multiset of intervals in $\mathbb{R}$ and completely characterises the homology of the 1-parameter family.

Qualitatively one considers long-lived homological features detected as inherent to the data set and the short-lived features as noise. In applications it has been noted that sometimes short-lived features provide important discriminating information in classification. Hence an important aspect of the field of TDA is to build a statistical framework in which one can detect statistically significant topological features of large, medium and small persistence, and in particular can discriminate genuine short-lived features from noise.

There are natural situations where one may wish to build a richer structure of topological spaces on a data set and track the changes to the homology whilst varying multiple parameters. For example, when studying the topology of chemical compounds one may want to track both the chemical properties and spatial arrangement of atoms. The persistent homology theory becomes wildly more complicated when the family of topological spaces on our data set is indexed by multiple parameters. The associated family of homology groups is known as a multiparameter persistence module. Unlike the single parameter case where one may associate a barcode to a module, there is not an analogous complete discrete invariant in the multiparameter setting.

Chapter 2 Multiparameter Persistence Landscapes

In this work the author tries to resolve the issue that it is not possible to define a barcode for multiparameter persistence modules. The author introduces a family of new stable invariants for multiparameter persistence modules, *multiparameter persistence landscapes*, naturally extending a similar construction for 1-parameter persistence modules to multiparameter persistence modules.

The landscape invariant encodes a multiparameter module $M : \mathbf{R}^n \to \mathbf{Vect}$ ($\mathbf{Vect}$-valued functor from the poset category $\mathbf{R}^n$) as a sequence of 1-Lipschitz landscape functions $\lambda : \mathbb{R}^n \to \mathbb{R}$.

Living in Lesbegue p-space, the multiparameter landscapes are naturally endowed with a distance function and are well suited to

statistical analysis. There is a uniquely defined mean associated to a collection of landscapes, and considered as Banach space valued random variables, multiparameter landscapes satisfy the Strong Law of Large Numbers and a Central Limit Theorem. Furthermore the natural inner-product structure on the landscape functions gives rise to a positive-definite kernel, which can be leveraged by machine learning algorithms.

The multiparameter landscape functions are sensitive to homological features of large, medium and small persistence. The landscapes also have the advantage of being interpretable since they are closely related to another well-known invariant: the rank invariant. Moreover one can derive stable $\mathbb{R}$-valued numeric invariants from the landscape functions using the linear functionals in the dual space. We can produce confidence intervals and perform hypothesis tests on these numeric invariants which are viewed as $\mathbb{R}$-valued random variables.

In the 2-parameter case one can visualise multiparameter landscapes as a surface $\lambda : \mathbb{R}^2 \to \mathbb{R}$. In this work the author presents computational examples for 2-parameter persistence landscapes. These examples demonstrate a range of potential applications, showing that the landscapes are sensitive to both the topology and geometry of data.

The major, but inevitable, drawback of this work is that the multiparameter persistence landscapes are an incomplete invariant in the sense that distinct multiparameter modules may give rise to the same landscape function. The extent of this incompleteness is explored in *Chapter* 3.

The code used for computations in this work is available online.

Author's Contribution

This work was initiated and carried out by the author. He is grateful to the support from his supervisors Prof Ulrike Tillmann and Prof Vidit Nanda in the development of the write up of this work and navigating the publication process.

Chapter 3 Local Equivalence of Metrics for Multiparameter Persistence Modules

The central focus of the work in this section is to compare two metrics on multiparameter persistence modules (**Vect**-valued functors from the poset category $\mathbf{R}^n$). The author seeks to compare the interleaving distance, d_I, a complete but computationally infeasible metric, with the matching distance, d_0, an efficiently computable yet

incomplete metric. For single parameter persistence modules these two metrics coincide. The disparity is only apparent for multiparameter persistence modules.

The theoretical properties of the interleaving distance are well-behaved. The interleaving distance is universal on the space of multiparameter modules i.e. the most discriminative stable metric. However, using the interleaving distance in practice is not feasible computationally, since the interleaving distance is NP-hard to compute and approximate for multiparameter modules. Hence for practical applications one must compromise between discriminating power and computability of multiparameter modules invariants and metrics.

The matching distance has less favourable theoretical properties. It is incomplete in the sense that it is a pseudo-metric; distinct modules which are arbitrarily distant in the interleaving distance may have zero matching distance between them. However this sacrifice in discriminating power has the upshot that the matching distance is efficiently computable.

The main technical result of the work in this section shows that the matching distance d_0 distinguishes a finitely presented ($\mathbb{R}^n$-indexed) multiparameter persistence module M from all modules N in a d_I-neighbourhood of M. The author shows there is a d_I-open ball, B_M, centred at M in the space of finitely presented modules such that for all $N \in B_M$:

$$\frac{1}{34} d_I(M,N) \leq d_0(M,N) \leq d_I(M,N) \tag{1.1}$$

The radius of the open ball for which (1.1) is valid is dependent on M.

The upper bound $d_0 \leq d_I$ is not new and is a constructed property of the matching distance d_0. The local lower bound $\frac{1}{34} d_I \leq d_0$ is proven in this work and gives rise to a global equivalence of intrinsic metrics.

This result for multiparameter modules attains as a simple corollary equivalences of metrics on constructible spaces which can be embedded into the space of finitely presented modules. In particular, interlevel set persistence modules give rise to finitely presented 2-parameter persistence modules and inherit the local equivalence result.

This work provides a theoretical justification for using the incomplete invariant proposed in Chapter 2. The matching distance has the same discriminating power as the multiparameter persistence landscapes, hence the result (1.1) implies that there is a local d_I-neighbourhood about each multiparameter persistence module M within which the multiparameter persistence landscapes of M distinguishes M from all other modules in this neighbourhood.

The author gave a talk about this work at an online seminar.

Author's Contribution

This work was initiated and carried out by the author. A major inspiration for this work was the article of Carriere and Oudot: *Local Equivalence and Intrinsic Metrics Between Reeb Graphs*. In this work Carriere and Oudot conjecture that similar results to their local equivalence of metrics on Reeb graphs apply to "more general classes of metric spaces than Reeb graphs". The author is grateful to the support from his supervisors Prof Ulrike Tillmann and Prof Vidit Nanda, and useful technical conversations with Jacob Leygonie.

Chapter 4 Random Čech Complexes on Riemannian Manifolds with Boundary

The work in this section is independent of the previous sections and studies the asymptotic properties of Čech complexes built on Riemannian manifolds. In particular this work considers the random simplicial complex realised by the Čech complex associated to a Poisson point process on a Riemannian manifold and asks: when does the topology of the simplicial complex approximate that of the manifold?

The question of recovering the topology of a space from a finite sample has been studied previously and in differing contexts. Some authors consider submanifolds in Euclidean space and use the metric of the ambient Euclidean space when building the Čech complex. They provide explicit conditions to recover the homology of the manifold with high confidence. Naturally these explicit conditions are dependent on the curvature and nearness to self-intersection of the embedded manifold. In contrast this work bases the construction of the Čech complex on the intrinsic metric of the manifold independent of any embedding.

The asymptotic properties and the phase transition for which one can recover the homology with high probability when increasing the size of point sample and decreasing the radius over which the associated complex is constructed. A principal advantage of studying asymptotics is that results rely on fewer assumptions on the underlying manifold from which the point process is sampled.

As a consequence the main result of this work has dependence only on the dimension of the underlying manifold and the homological dimension one wishes to recover. This work uses some similar techniques already present in the literature but also develops completely new arguments to take into account the effect of the bound-

ary.

The main result is stated with respect to the term $\Lambda := n\omega_d r^d$, the expected number of points of a uniform Poisson process of intensity n lying in a d-dimensional radius r ball. Depending on the asymptotic behaviour of Λ, the associated Čech complex at scale r built on the point process exhibits different behaviours.

There are three distinct regimes of behaviour for Λ as $n \to \infty, r \to 0$. In the subcritical regime ($\Lambda \to 0$) the connectivity of the Čech complex is very sparse and mostly disconnected, with the number of connected components growing at the same rate as the number of points. In the critical regime ($\Lambda \to \lambda \in (0, \infty)$) the Čech complex is sufficiently connected to exhibit non-trivial homology. However the number of connected components still grows linearly with the number of points. In the supercritical regime ($\Lambda \to \infty$) for sufficiently large Λ the Čech complex is connected, and for even larger Λ the point cloud covers the underlying manifold with high probability.

Analysis in the supercritical regime yields a sequence of increasing thresholds, (*homological connectivity thresholds*), such that if Λ is greater than the k^{th} threshold the Čech complex recovers the k^{th} homology of the underlying closed manifold with high probability. The intermediate homological connectivity thresholds interpolate between the thresholds for more commonly studied properties, from the 0^{th} homology which detects connectivity up to the d^{th} homology which detects coverage. We produce homological connectivity thresholds in the supercritical regime for which the Čech complex recovers the homology of a smooth compact manifold with non-trivial boundary.

The non-trivial boundary has a significant impact on the homological connectivity thresholds. As far as we know, our study of manifolds with boundary has uncovered a previously unobserved phenomenon occuring close to the boundary. Our analysis shows that close to the boundary a large number of spurious k-cycles appear which are not homological cycles inherent to the k^{th} homology of the underlying manifold. This phenomenon determines that the homological connectivity thresholds for manifolds with boundary are almost twice a big as those for a closed manifold.

Author's Contribution

This work was a collaboration between Prof Ulrike Tillmann, Dr Henry-Louis de Kergolay and the author. UT and the author have expertise in the topological aspect of this work and HdK has expertise in stochastic processes.

2
Multiparameter Persistence Landscapes

Contents

Abstract

An important problem in the field of Topological Data Analysis is defining topological summaries which can be combined with traditional data analytic tools. In recent work Bubenik introduced the persistence landscape, a stable representation of persistence diagrams amenable to statistical analysis and machine learning tools. In this paper we generalise the persistence landscape to multiparameter persistence modules providing a stable representation of the rank invariant. We show that multiparameter landscapes are stable with respect to the interleaving distance and persistence weighted Wasserstein distance, and that the collection of multiparameter landscapes faithfully represents the rank invariant. Finally we provide example calculations and statistical tests to demonstrate a range of potential applications and how one can interpret the landscapes associated to a multiparameter module.

2.1 *Introduction*

Topological and Geometric Data Analysis (TGDA) describes an emerging set of analytic tools which leverage the underlying shape data to produce topological summaries. These techniques have been particularly successful at providing new insight for high dimensional data sets, topological data structures and biological data sets (3; 4; 5). An ideal topological summary should discriminate well between different spaces, be stable to perturbations of the initial data, and be amenable to statistical analysis.

Persistent homology (PH) has become a ubiquitous tool in the TGDA arsenal. PH studies the homology groups of a family of topological spaces built upon data. Homology is an invariant from algebraic topology which detects different dimensional holes in a topological space. Degree zero homology detects connected components, degree one homology detects loops, degree two homology detects voids and higher degree homology detects higher dimensional cavities.

Persistent homology tracks how these homological feature persist throughout the family of topological spaces. The associated topological summary for a 1-parameter family is given in terms of a persistence diagram marking the parameter values for births and deaths of homological features. The persistence diagram may equivalently be thought of as a multiset of points in $\mathbb{R}^2$. For an introduction to persistent homology see (6) and (7).

Qualitatively one considers long-lived homological features detected as inherent to the data set and the short-lived features as noise.

Nevertheless in various applications it has been revealed that the short-lived features provide important discriminating information in classification (8). For example in analysing the topology of brain arteries it was found the 28^{th} longest persisting feature provided the most useful discriminating information (9). We would like therefore to have a statistical framework and topological summary in which one can detect statistically significant topological features of large, medium and small persistence, and in particular does not discard short-lived features as noise.

One can equip the collection of persistence diagrams with a natural pseudo-metric known as the bottleneck distance. The resulting metric space of persistence diagrams does not enjoy desired properties for traditional statistical analysis. For example, a collection of persistence diagrams may not have a well defined mean. As a result there have been various attempts to vectorize the persistence diagram, in order that the summary is more amenable to statistical analysis and machine learning techniques.

The article (10) introduces a stable vectorization of the persistence diagram, the persistence landscape, a function in Lebesgue p-space. The article (11) explores some further properties of the persistence landscape. The persistence landscape naturally enjoys unique means and one can perform traditional hypothesis tests upon the landscape function and numerical statistics derived from this function. Another popular stable vectorization of the persistence diagram is the persistence image (12) which has been shown to produce favourable classification results when combined with machine learning techniques.

There are several natural situations where we may wish to build a richer structure of topological spaces on a data set and track the changes to the homology whilst varying multiple parameters. For example in (13) the topology of chemical compounds is studied using 2-parameter filtrations. The PH theory becomes wildly more complicated when the family of topological spaces on our data set is indexed by multiple parameters. The associated family of homology groups is known as a multiparameter persistence module. The theory of multiparameter persistence modules is presented in (14) and unlike the single parameter case where we may associate a persistence diagram to a module, there is not an analogous complete discrete invariant in the multiparameter setting.

There exist various approaches to define invariants for multiparameter persistence modules in the literature. The rank invariant of a module has been studied in the context of H_0-modules and shape matching and has been shown to be stable when endowed with a matching distance (15; 16). An alternative approach uses algebraic

geometry to construct numeric functions on multiparameter persistence modules (17), generalizing the ring of algebraic functions on barcode space for the single parameter case (18). However, as noted by (19), the disadvantage of both of these approaches is that equipped with their vector space norms, these invariants are unstable with respect to the natural distance to consider on multiparameter modules, the interleaving distance. Another approach is to study algebraic invariants associated to the multigraded algebra structure of multiparameter persistence modules (20; 21).

In this article we introduce a family of new stable invariants for multiparameter persistence modules, naturally extending the results of (10) from the setting of single parameter persistence modules to multiparameter persistence modules. We shall consider persistence modules indexed by continuous real parameters. Our incomplete invariants, the *multiparameter persistence landscapes*, are derived from the rank invariant associated to a multiparameter persistence module, and they are continuous functions in Lesbegue p-space. The multiparameter persistence landscape reduces to a family of single parameter persistence landscapes: the multiparameter landscape of a persistence module M indexed over $\mathbb{R}^n$ is characterized by the fact that it restricts to the single parameter landscape function of the restriction of M to every line parallel to the main diagonal vector $\mathbf{1} = (1, 1, ..., 1)$.

Thus, whilst the single parameter persistence landscape faithfully represents the persistence diagram of a single parameter module (22), the multiparameter persistence landscape encodes the persistence diagrams of the family of single parameter persistence submodules which lie on the lines of slope $\mathbf{1}$ through the parameter space.

Living in Lesbegue p-space, the multiparameter landscapes are naturally endowed with a distance function and there is a uniquely defined mean associated to multiple landscapes. Considered as Banach space valued random variables, multiparameter landscapes satisfy the Strong Law of Large Numbers and the Central Limit Theorem and are thus are well suited to statistical analysis. The natural inner-product structure on the landscape functions gives rise to a positive-definite kernel, which can be leveraged by machine learning algorithms.

The multiparameter landscape functions are sensitive to homological features of large, medium and small persistence. The landscapes also have the advantage of being interpretable since they are closely related to the rank invariant. Moreover one can derive stable $\mathbb{R}$-valued numeric invariants from the landscape functions using the linear functionals in the dual space. We can produce confidence intervals and perform hypothesis tests on these numeric invariants which are viewed as $\mathbb{R}$-valued random variables.

In the 2-parameter case we visualise a multiparameter landscapes as a surface $\lambda : \mathbb{R}^2 \to \mathbb{R}$. We shall present computational examples in the 2-parameter case using the RIVET software presented in (23). These examples will serve a range of potential applications, demonstrating that the landscapes are sensitive to both the topology and geometry of data.

2.2 *Multiparameter Persistence*

The central objects of interest in multiparameter persistence theory are multiparameter persistence modules. These are algebraic objects which can be used to encode rich topological and geometric properties of a data set. The following example gives a construction of a multiparameter persistence module.

Example 2.2.1 (Sublevel-set Multiparameter Persistence Module)**.** *Let X be a topological space and $f : X \to \mathbb{R}^n$, called a filtering function. We can associate a family of topological subspaces indexed by vectors $\mathbf{a} = (a_1, ..., a_n) \in \mathbb{R}^n$ induced by f:*

$$X_{\mathbf{a}} = \{x \in X : f(x)_i < a_i \; \forall \; i = 1, ..., n\}$$

this is known as the sublevel-set filtration. For any $\mathbf{b} \in \mathbb{R}^n$ such that $a_i \leq b_i$ for all $i = 1, .., n$, we have an inclusion map $X_{\mathbf{a}} \hookrightarrow X_{\mathbf{b}}$. If we let H denote a singular homology functor with coefficients in a field then applying this functor to the collection $\{X_{\mathbf{a}}\}_{\mathbf{a} \in \mathbb{R}^n}$ and the appropriate inclusion maps gives rise to a family of vector spaces $\{H(X_{\mathbf{a}})\}_{\mathbf{a} \in \mathbb{R}^n}$ and linear maps $\{H(X_{\mathbf{a}}) \to H(X_{\mathbf{b}})\}_{\mathbf{a} \leq \mathbf{b}}$ known as a Sublevel-set Multiparameter Persistence Module.

In the case of point cloud data in Euclidean space we can capture the spatial distribution of the points with our filtering function. Let $P = \{p_1, ..., p_n\} \subset \mathbb{R}^N = X$ denote a point cloud in Euclidean space. We may define our filtering function from this point cloud $f_P : \mathbb{R}^N \to \mathbb{R}^n$ to have the first parameter be the distance function from the point cloud $f_P(\mathbf{x})_1 = \min_{p \in P} \|x - p\|$, (where $\| \cdot \|$ denotes the Euclidean norm). The remaining coordinates can be used to filter our space by other parameters of interest. For example, one could choose the second parameter to be a density function in order to introduce robustness to outliers of the point cloud. Alternatively a second parameter could be chosen to track a numeric property associated to the points of the point cloud such as charge or mass in the case that the points in the point cloud represent atoms (13).

The resulting sublevel-set multiparameter persistence module will then encode the topology of the spatial distribution of the point

cloud and its interdependence with other chosen filtration parameters. We provide such constructions and example computations in Section 2.6.

We can give a compact mathematical description of multiparameter persistence modules as objects of functor categories following (24) or as graded modules over polynomial rings following (25; 23).

Let P_n denote the monoid ring of the monoid $([0,\infty)^n, +)$ over a field $\mathbb{F}$. Equivalently one may think of P_n as a pseudo-polynomial ring $\mathbb{F}[x_1, ..., x_n]$ in which exponents are only required to be non-negative and can be non-integral. Let A_n denote the polynomial ring $\mathbb{F}[x_1, ..., x_n]$ or analogously the monoid ring of $(\mathbb{N}^n, +)$ over $\mathbb{F}$.

Let $\mathbf{P}$ denote the category associated to the poset P, so that $\mathbf{R}^n$ and $\mathbf{Z}^n$ denote the categories associated to the posets $(\mathbb{R}^n, \leq)$ and $(\mathbb{Z}^n, \leq)$ under the standard coordinate-wise partial orders. Let **Vect** denote the category of vector spaces and linear maps over $\mathbb{F}$, and **vect** denote the subcategory of finite dimensional vector spaces. Moreover, for a category $\mathbf{C}$, let us denote the $\mathbf{C}$-valued functors on $\mathbf{P}$ by $\mathbf{C}^{\mathbf{P}}$.

Definition 2.2.2 (Multiparameter Persistence Module)**.** *Let M be a module over the ring P_n. We say M is a persistence module if M is an $\mathbb{R}^n$-graded P_n-module. That is to say M has a decomposition as a $\mathbb{F}$-vector space $M = \bigoplus_{\mathbf{a}\in\mathbb{R}^n} M_{\mathbf{a}}$ compatible with the action of P_n:*

$$m \in M_{\mathbf{a}} \implies \mathbf{x}^{\mathbf{b}} \cdot m \in M_{\mathbf{a}+\mathbf{b}}$$

We require a morphism of $\mathbb{R}^n$-graded modules $f : M \to N$ to be compatible with the module structure $f(\mathbf{x}^{\mathbf{b}} \cdot m) = \mathbf{x}^{\mathbf{b}} \cdot f(m)$ and respect the grading of the modules so that $m \in M_{\mathbf{a}}$ implies $f(m) \in N_{\mathbf{a}}$.

In the setting of a sublevel-set persistence module the vector space at each grade is $H(X_{\mathbf{a}})$, and the action of $\mathbf{x}^{\mathbf{b}}$ on $H(X_{\mathbf{a}})$ is given by the linear map of homology groups induced by the inclusion $X_{\mathbf{a}} \hookrightarrow X_{\mathbf{a}+\mathbf{b}}$.

Definition 2.2.3 (Multiparameter Persistence Module)**.** *Let M be an element of the functor category $\mathbf{Vect}^{\mathbf{R}^n}$. We say M is a multiparameter persistence module. A morphism of persistence modules is a natural transformation $M \Rightarrow N$.*

Definition 2.2.2 and Definition 2.2.3 are equivalent as realised by an equivalence of categories between $\mathbb{R}^n$-graded P_n-**Mod** and $\mathbf{Vect}^{\mathbf{R}^n}$ (26).

The theory of persistent homology is well developed for topological spaces filtered by a single parameter. Under appropriate finiteness conditions, one can record the homological features of the filtered

topological space captured by the single parameter persistence module in a multiset of intervals known as the barcode. Moreover the Krull-Schmidt Decomposition Theorem establishes the barcode as a complete invariant (27). No such discrete complete invariant exists for multiparameter modules. However, associated to a pointwise finite dimensional multiparameter module is a family of single parameter modules whose collection of barcodes is known as the fibered barcode (23). Since we are working with persistence modules induced by real world finite data, the pointwise finite dimensionality assumption we require for the fibered barcode to exist is not a serious restriction. We shall see later that we will be able to reduce the computation of our multiparameter module invariant, the multiparameter persistence landscapes, to queries of the fibered barcode.

Definition 2.2.4 (Fibered Barcode (23))**.** *Let* $\mathbf{L}$ *denote the subposet of* $\mathbf{R}^n$ *corresponding to a positively sloped line* $L \subset \mathbb{R}^n$*. Let* $\iota_L : (\mathbf{R}, \|\cdot\|_\infty) \to (\mathbf{R}^n, \|\cdot\|_\infty)$ *denote the isometric embedding with* $\iota_L(\mathbf{R}) = \mathbf{L}$ *and* $\iota_L(0) \in \{x_n = 0\}$*. For* $M \in \mathbf{vect}^{\mathbf{R}^n}$ *the composite* $M^L = M \circ \iota_L$ *is a pointwise finite-dimensional single parameter persistence module, and thus has an associated barcode* $\mathcal{B}(M^L)$*. Let* $\mathcal{L}$ *denote the set of positively sloped lines. The collection* $\{\mathcal{B}(M^L) : L \in \mathcal{L}\}$ *is known as the fibered barcode of* M*.*

In view of applications of multiparameter persistence to data analysis we would like a distance function to compare multiparameter persistence modules. The interleaving distance, d_I, is a pseudo-metric on multiparameter persistence modules and provides a notion of algebraic similarity. Details of the interleaving distance can be found in the Appendix 2.8.2. This pseudo-metric has been shown to be the most discriminating, stable metric on multiparameter persistence modules (26). A comprehensive mathematical formulation of metrics for generalized persistence modules is presented in (28).

Despite the strong theoretical properties which make the interleaving distance a good candidate to compare multiparameter persistence modules, it has a couple of undesirable properties for data analysis. The interleaving distance has been shown to be NP-hard to compute (29; 30) and so is not computable in applications. Moreover the interleaving distance has the characteristics of an L^∞-style distance, in that it measures the worst place two modules match up over the parameter space. In contrast, an L^p-style distance, for $p \in [1, \infty)$, is sensitive to the difference of two modules over the whole parameter space.

2.2.1 *Interval Decomposable Modules*

Given the complicated nature of unconstrained multiparameter persistence modules, it is common to consider subclasses of multiparam-

eter persistence modules. These modules admit a discrete complete description in analogy with single parameter modules. We can use the decomposition of this subclass of modules into simple summands to define matching distances between theses modules. Recall that we denote the category of finite dimensional vector spaces as **vect**.

Definition 2.2.5 (Interval Decomposable Modules)**.** *We define a subposet $I \leq \mathbb{R}^n$ to be an interval if $s,t \in I, s \leq r \leq t$ implies $r \in I$ and for any $s,t \in I \; \exists \; r_i \in I$ connecting s and t, $s = r_0 \leq r_1 \geq r_2 \leq r_3 \geq ... \leq r_n = t$. The interval module $\mathbb{1}_{\mathbf{I}} \in \mathbf{vect}^{\mathbf{R}^n}$ associated to an interval I has a one dimensional vector space at each $a \in I$ and internal isomorphisms given by the identity wherever possible. We say a module $M \in \mathbf{vect}^{\mathbf{R}^n}$ is interval decomposable if $M \cong \bigoplus_{j \in \mathcal{J}} \mathbb{1}_{\mathbf{I}_j}$, for some indexed set of intervals $\{I_j : j \in \mathcal{J}\}$.*

The Krull-Schmidt-Remak-Azumaya Theorem guarantees that the decomposition of an interval decomposable module is unique up to reordering. We can thus assign the indexed set of intervals in the decomposition of a module M to be the barcode $\mathcal{B}(M) = \{I_j : j \in \mathcal{J}\}$.

Definition 2.2.6 (ε-Matching)**.** *Let $\{I_j : j \in \mathcal{J}\}$ and $\{J_k : k \in \mathcal{K}\}$ be indexed sets of intervals. We say a partial bijection $\sigma : \mathcal{J} \nrightarrow \mathcal{K}$ is an ε-matching if $d_I(\mathbb{1}_{I_j}, \mathbb{1}_{J_{\sigma(j)}}) \leq \varepsilon$ for matched intervals and $d_I(\mathbb{1}_{I_j}, 0), d_I(0, \mathbb{1}_{J_k}) \leq \varepsilon$ for unmatched intervals.*

Definition 2.2.7 (Bottleneck Distance)**.** *Let M and N be interval decomposable modules. The bottleneck distance between the modules is given by*

$$d_B(M,N) = \inf\{\varepsilon \geq 0 : \mathcal{B}(M), \mathcal{B}(N) \text{ admit an } \varepsilon\text{-matching}\}.$$

One would hope to attain a result analogous to the isometry theorem for ordinary one dimensional persistent homology relating the bottleneck distance and the interleaving distance. In the single parameter case an ε-interleaving induces an ε-matching between summands (31). In contrast, the interleaving distance and bottleneck distance do not coincide for multiparameter interval decomposable modules. Certainly the bottleneck distance provides an upper bound on the interleaving distance. However general interleavings of interval decomposable multiparameter modules do not necessarily induce a matching of interval summands. This is best illustrated by an example provided in (32) for which the optimal matching between 1-interleaved modules is a 3-matching.

We can further define a Wasserstein distance for interval decomposable modules.

Definition 2.2.8 (p-Wasserstein Distance)**.** *Let M, N be interval decomposable persistence modules with barcodes $\{I_j : j \in \mathcal{J}\}$ and $\{J_\kappa : \kappa \in \mathcal{K}\}$ respectively. Append a collection of empty intervals to each barcode of cardinality the size of the other barcode. For a matching $\sigma : \mathcal{J} \to \mathcal{K}$, let $\varepsilon_j = d_I(\mathbb{1}_{I_j}, \mathbb{1}_{J_{\sigma(j)}})$. The p-Wasserstein distance is given by*

$$d_{W_p}(M, N) = \inf_{\sigma:\mathcal{J}\to\mathcal{K}} \left[\sum_{\mathcal{J}} \varepsilon_j^p\right]^{\frac{1}{p}}.$$

The bottleneck distance is simply the ∞-Wasserstein distance. If we wish to place extra emphasis on intervals with large persistence we may use the persistence weighted p-Wasserstein distance (Definition 2.2.10). In order to ensure the persistence weighted p-Wasserstein distance is well defined we need to check that intervals are Lebesgue measurable.

Proposition 2.2.9. *If $I \subset \mathbb{R}^n$ is an interval of the partially ordered set $\mathbb{R}^n$, then I is Lebesgue measurable.*

Proof. Recall the following characterisation of Lebesgue measurable sets. A set $I \subset \mathbb{R}^n$ is Lebesgue measurable if and only if for all $\varepsilon > 0$ there exists an open set U and a closed set C such that $U \subseteq I \subseteq C$, and the Lebesgue measure $\mu(C \setminus U) < \varepsilon$.

To establish intervals are measurable it suffices to show that for any interval $I \subset \mathbb{R}^n$ there is some open $U \subset \mathbb{R}^n$ and some closed set $C \subset \mathbb{R}^n$ (open and closed under the standard Euclidean topology) such that $U \subseteq I \subseteq C$ and $C \setminus U$ has zero measure.

Let $\overline{I}$ and I° denote closure and interior of the subset $I \subset \mathbb{R}^n$ in the standard Euclidean topology respectively. We shall first show that the interior of the closure of an interval $(\overline{I})^\circ$ coincides with the interior of that interval $(\overline{I})^\circ = I^\circ$. If $\mathbf{x} \in (\overline{I})^\circ$, then there is some ∞-norm ball contained in $\overline{I}$ containing $\mathbf{x}$ in its interior. Thus there exist $\mathbf{a}, \mathbf{b} \in \overline{I}$ with $a_i < x_i < b_i$ for all $i = 1, ..., n$. The interval I is dense in $\overline{I}$ so we can perturb $\mathbf{a}$ and $\mathbf{b}$ to $\mathbf{a}', \mathbf{b}' \in I$ with $a_i' < x_i < b_i'$. Since I is an interval containing $\mathbf{a}', \mathbf{b}'$ we have that the open set $\{\mathbf{y} \in \mathbb{R}^n : a_i' < y_i < b_i'\}$ is contained in I, and so $\mathbf{x} \in I^\circ$.

The set $\overline{I}$ is closed and so it is Lebesgue measurable and in particular $\overline{I} \setminus (\overline{I})^\circ$ has zero measure. Thus $\overline{I} \setminus I^\circ = \overline{I} \setminus (\overline{I})^\circ$ also has zero measure and we have found the requisite $U = I^\circ$ and $C = \overline{I}$ witnessing that I is measurable. □

Definition 2.2.10 (Persistence Weighted p-Wasserstein Distance)**.** *Let M, N be interval decomposable persistence modules with barcodes $\{I_j : j \in \mathcal{J}\}$ and $\{J_\kappa : \kappa \in \mathcal{K}\}$. For a union of a pair of intervals $I \cup J \subset \mathbb{R}^n$ let $|I \cup J|$ denote the Lebesgue measure. The persistence weighted p-Wasserstein distance is given by*

$$d_{\overline{W}_p}(M,N) = \inf_{\sigma:\mathcal{J}\to\mathcal{K}} \left[\sum_{\mathcal{J}} |I_j \cup J_{\sigma(j)}|\varepsilon_j^p\right]^{\frac{1}{p}}.$$

The p-landscape distance we introduce in the following section is similar to the persistence weighted p-Wasserstein Distance and can be defined for persistence modules which do not admit an interval decomposition.

In Section 2.4 we will show that our invariant is stable with respect to interleaving distance and the persistence weighted Wasserstein distance. In particular the L^p vector space norms on multiparameter landscapes provide stable, computable distance functions for multiparameter persistence modules.

2.3 *Persistence Landscapes*

In this section we shall recall the definition of the single parameter persistence landscape and its properties. We shall generalize the definition to multiparameter persistence modules and show which properties of the single parameter persistence landscape are preserved.

From this point onward all single parameter persistence modules we consider shall be pointwise finite dimensional in order that they admit an interval decomposition (27). The multiparameter persistence modules we consider will be pointwise finite dimensional, but will not necessarily admit an interval decomposition.

2.3.1 *Single Parameter Persistence Landscapes*

The *persistence landscape* associated to a single parameter persistence module is defined in (10). The persistence landscape is derived from the rank invariant of a module.

Definition 2.3.1 (Rank Invariant)**.** *Let $M \in \boldsymbol{vect}^{\boldsymbol{R}}$ be a persistence module then for $a \leq b$ the function $\beta^{\cdot,\cdot}$ giving the corresponding Betti number is the rank invariant of M:*

$$\beta^{a,b} = \dim(\mathrm{Im}(M_a \to M_b)).$$

Definition 2.3.2 (Rank Function)**.** *The rank function* $\mathrm{rk} : \mathbb{R}^2 \to \mathbb{R}$ *is given by*

$$\mathrm{rk}(b,d) = \begin{cases} \beta^{b,d} & \text{if } b \leq d \\ 0 & \text{otherwise.} \end{cases}$$

Definition 2.3.3 (Rescaled Rank Function)**.** *The rescaled rank function* $r : \mathbb{R}^2 \to \mathbb{R}$ *is supported on the upper half plane:*

$$r(m,h) = \begin{cases} \beta^{m-h,m+h} & \text{if } h \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Observe that the rank function has support contained in the upper triangular half of the plane with the coordinates corresponding to "births" and "deaths", whilst the rescaled rank function has support contained in the upper half plane with coordinates corresponding to "midpoints" and "half-lifes".

Definition 2.3.4 (Persistence Landscape (10))**.** *The persistence landscape is a function* $\lambda : \mathbb{N} \times \mathbb{R} \to \overline{\mathbb{R}}$*, where* $\overline{\mathbb{R}}$ *denotes the extended real numbers,* $[-\infty, \infty]$*. The function* $\lambda(k,t) : \mathbb{R} \to \overline{\mathbb{R}}$ *is defined by*

$$\lambda(k,t) = \sup\{h \geq 0 : \beta^{t-h,t+h} \geq k\}.$$

The value $\lambda(k,t)$ gives the maximal radius of an interval centred at t that is contained in at least k intervals of the barcode. The persistence landscape and persistence diagram of a suitably well-behaved single parameter module carry the same information (22).

Alternatively the persistence landscape of a single parameter module can be derived from the landscape functions of the modules' interval summands; see Figure 2.1.

Definition 2.3.5 (Persistence Landscape (10))**.** *Let* M *be a single parameter persistence module with associated persistence diagram given by the indexed set* $\{(b_j, d_j) : j \in \mathcal{J}\}$*. The persistence landscape of* M *may be equivalently defined as:*

$$\lambda(k,t) = \text{kmax}_{j \in \mathcal{J}}\{\lambda((b_j,d_j))(1,t)\}$$

where kmax *denotes the* k^{th} *largest value of the indexed set and* $\lambda((b_j,d_j))$ *is the landscape associated to the interval module* $\mathbb{1}_{(b_j,d_j)}$*.*

Lemma 2.3.6 ((10))**.** *The persistence landscape has the following properties:*

1. $\lambda(k,t) \geq 0$.
2. $\lambda(k,t) \geq \lambda(k+1,t)$.
3. $\lambda(k,t)$ *is 1-Lipschitz.*

The first two properties are immediate from the definition and the third property is proved by (10).

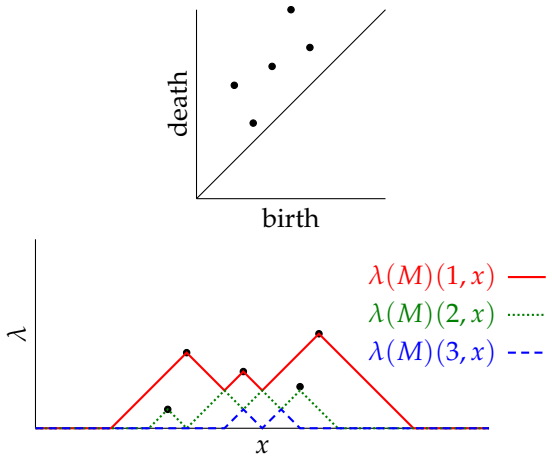


Figure 2.1: We show the persistence diagram of a single parameter module M on the left and the associated persistence landscapes on the right. One can see that $\lambda(M)(k,t)$ is the kmax of the landscape functions $\lambda((b_j,d_j))(1,t)$ of the interval summands of M.

2.3.2 *Multiparameter Persistence Landscapes*

Let us define the multiparameter persistence landscape in analogy with the single parameter case. The rank invariant, rank function and rescaled rank function defined above generalise naturally to multiparameter persistence modules:

Definition 2.3.7 (Rank Invariant). *Let $M \in \mathbf{vect}^{\mathbf{R}^n}$ be a multiparameter persistence module, then for $\mathbf{a} \leq \mathbf{b}$ the function $\beta^{\cdot,\cdot}$ giving the corresponding Betti number is the rank invariant of M:*

$$\beta^{\mathbf{a},\mathbf{b}} = \dim(\mathrm{Im}(M_{\mathbf{a}} \to M_{\mathbf{b}})).$$

Definition 2.3.8 (Multiparameter Rank Function). *The rank function $\mathrm{rk} : \mathbb{R}^{2n} \to \mathbb{R}$ is given by*

$$\mathrm{rk}(\mathbf{b},\mathbf{d}) = \begin{cases} \beta^{\mathbf{b},\mathbf{d}} & \text{if } \mathbf{b} \leq \mathbf{d} \\ 0 & \text{otherwise.} \end{cases}$$

Definition 2.3.9 (Rescaled Multiparameter Rank Function). *The rescaled rank function $r : \mathbb{R}^{2n} \to \mathbb{R}$*

$$r(\mathbf{m},\mathbf{h}) = \begin{cases} \beta^{\mathbf{m}-\mathbf{h},\mathbf{m}+\mathbf{h}} & \text{if } \mathbf{h} \geq \mathbf{0} \\ 0 & \text{otherwise.} \end{cases}$$

One could perform statistical analysis directly to the rank function and rescaled rank function. Endowed with the L^p normed vector space structure these functions are not stable with respect to the interleaving distance for all $p \in [1,\infty]$.

Example 2.3.10 (Rank Function Instability). *For $\varepsilon > 0$ and $N \in \mathbb{N}$ consider the multiparameter persistence module M with presentation $M =$*

$\langle \{(a_i, \mathbf{0})\}_{i=1,\dots,N} \mid \{\mathbf{x}^\varepsilon \cdot a_i\}_{i=1,\dots,N} \rangle$. *The persistence module M is such that* $\| \operatorname{rk}_M \|_\infty = N, \| \operatorname{rk}_M \|_p = \infty$ *for all* $p \in [1, \infty)$, *and* $d_I(M, 0) = \varepsilon$. *Details of presentations and the interleaving distance d_I for multiparameter persistence modules can be found in Appendix 2.8.*

We wish to define a stable invariant and we derive a landscape function from the rank invariant.

Definition 2.3.11 (Multiparameter Persistence Landscape)**.** *The multiparameter persistence landscape considers the maximal radius over which k features persist in every (positive) direction through $\mathbf{x}$ in the parameter space* $\lambda : \mathbb{N} \times \mathbb{R}^n \to \overline{\mathbb{R}}$.

$$\lambda(k, \mathbf{x}) = \sup\{\varepsilon \geq 0 : \beta^{\mathbf{x}-\mathbf{h}, \mathbf{x}+\mathbf{h}} \geq k \text{ for all } \mathbf{h} \geq \mathbf{0} \text{ with } \|\mathbf{h}\|_\infty \leq \varepsilon\}.$$

It is worth noting that when restricted to a single parameter persistence module this definition coincides with the single parameter persistence landscape (10). If there are multiple persistence modules under consideration we shall denote the landscape associated to module M as $\lambda(M)(k, \mathbf{x})$.

Lemma 2.3.12. *The multiparameter persistence landscape has the following properties:*

1. $\lambda(k, \mathbf{x}) \geq 0$.
2. $\lambda(k, \mathbf{x}) \geq \lambda(k+1, \mathbf{x})$.
3. $\lambda(k, \mathbf{x})$ *is 1-Lipschitz.*

Proof. The first two properties follow immediately from the definition. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and let $k \in \mathbb{N}$. Without loss of generality assume that $\lambda(k, \mathbf{x}) \geq \lambda(k, \mathbf{y})$ and that also $r = \lambda(k, \mathbf{x}) \geq \|\mathbf{x} - \mathbf{y}\|_\infty = \delta$. We seek to show that $\lambda(k, \mathbf{y}) \geq \lambda(k, \mathbf{x}) - \|\mathbf{x} - \mathbf{y}\|_\infty$.

For any $\varepsilon \geq \mathbf{0}$ such that $\|\varepsilon\|_\infty \leq r - \delta$ let us define $\mathbf{h} = (|x_i - y_i| + \varepsilon_i)_i$. We observe that $\|\mathbf{h}\|_\infty \leq \|\mathbf{x} - \mathbf{y}\|_\infty + \|\varepsilon\|_\infty \leq r$ and in particular that:

$$\mathbf{x} - \mathbf{h} \leq \mathbf{y} - \varepsilon \leq \mathbf{y} + \varepsilon \leq \mathbf{x} + \mathbf{h}.$$

Thus $\|\mathbf{h}\|_\infty \leq r$ means the map $M(\mathbf{x} - \mathbf{h} \leq \mathbf{x} + \mathbf{h})$ which factors through $M(\mathbf{y} - \varepsilon \leq \mathbf{y} + \varepsilon)$ has rank at least k. Since ε was arbitrary we see that $\lambda(k, \mathbf{y}) \geq r - \delta$. □

In extending the persistence landscape definition to multiparameter persistence modules we encountered a choice of p-norm for the ball in $\mathbb{R}^n$ over which we ask features persist. Whilst Lemma 2.3.12 holds for all choices of p-norm it transpires that the most natural choice is the ∞-norm. This choice considerably simplifies computation of the multiparameter persistence landscape and gives rise to a number of further properties which we explore in this section.

Lemma 2.3.13. *Let M be a multiparameter persistence module with rank invariant $\beta^{\cdot,\cdot}_M$. For all $\mathbf{h} \geq \mathbf{0}$ with $\|\mathbf{h}\|_\infty = h$ we have that $\beta_M^{\mathbf{x}-h\mathbf{1},\mathbf{x}+h\mathbf{1}} \leq \beta_M^{\mathbf{x}-\mathbf{h},\mathbf{x}+\mathbf{h}}$.*

Proof. For all $\mathbf{h} \geq \mathbf{0}$ with $\|\mathbf{h}\|_\infty = h$ we have that $\mathbf{x} - h\mathbf{1} \leq \mathbf{x} - \mathbf{h} \leq \mathbf{x} + \mathbf{h} \leq \mathbf{x} + h\mathbf{1}$. Hence the linear map $M(\mathbf{x} - h\mathbf{1} \leq \mathbf{x} + h\mathbf{1})$ factors through the map $M(\mathbf{x} - \mathbf{h} \leq \mathbf{x} + \mathbf{h})$ and thus the result follows. □

In view of Lemma 2.3.13, we see that in order to compute the value of a multiparameter persistence landscape $\lambda(k, \mathbf{x})$ at the point $\mathbf{x} = \mathbf{x}_0$, we only need to compute $\sup\{\varepsilon \geq 0 : \beta^{\mathbf{x}_0-\varepsilon\mathbf{1},\mathbf{x}_0+\varepsilon\mathbf{1}} \geq k\}$. Thus we only need to compute a single barcode in the fibered barcode to compute the landscape value at a point.

Proposition 2.3.14. *Let M be a persistence module and let $\mathbf{L}$ be a line of slope $\mathbf{1}$ through the parameter space $\mathbf{R}^n$. Let $\iota_L : (\mathbf{R}, \|\cdot\|_\infty) \to (\mathbf{R}^n, \|\cdot\|_\infty)$ denote the isometric embedding with $\iota_L(\mathbf{R}) = \mathbf{L}$ and $\iota_L(0) \in \{x_n = 0\}$. The restriction of the multiparameter landscape of M to L and the single parameter persistence landscape of the persistence module M^L coincide, $\lambda(M^L)(k,t) = \lambda(M)(k, \iota_L(t))$.*

Proof. Following the landscape definitions we see that if $\lambda(M)(k, \iota_L(t)) > h$ then we have that $\beta_M^{\iota_L(t)-h\mathbf{1},\iota_L(t)+h\mathbf{1}} = \beta_{M^L}^{t-h,t+h} \geq k$ and so $\lambda(M^L)(k,t) > h$ thus $\lambda(M^L)(k,t) \geq \lambda(M)(k, \iota_L(t))$. Conversely, if $\lambda(M^L)(k,t) > h$ then $\beta_M^{\iota_L(t)-h\mathbf{1},\iota_L(t)+h\mathbf{1}} \geq k$, and so by Lemma 2.3.13 we have that $\beta_M^{\iota_L(t)-\mathbf{h},\iota_L(t)+\mathbf{h}} \geq \beta_M^{\iota_L(t)-h\mathbf{1},\iota_L(t)+h\mathbf{1}} \geq k$ for all $\mathbf{h} \geq \mathbf{0}$ with $\|\mathbf{h}\|_\infty = h$ and thus $\lambda(M^L)(k,t) \leq \lambda(M)(k, \iota_L(t))$. □

Let kmax denote the function on finite multisets mapping a finite multiset S to the k^{th} largest value of the multiset if $|S| \geq k$ and 0 otherwise. An immediate consequence of Proposition 2.3.14 is that like the single parameter persistence landscape, the multiparameter persistence landscape also admits a decomposition as the kmax of a series of simple landscape functions when our module is interval decomposable.

Proposition 2.3.15. *The multiparameter persistence landscape of an interval decomposable module $M \cong \bigoplus_j \mathbb{1}_{I_j}$ can be equivalently defined as:*

$$\lambda(M)(k, \mathbf{x}) = \text{kmax}_j\{\lambda(\mathbb{1}_{I_j})(1, \mathbf{x})\}.$$

Proposition 2.3.16 follows from Proposition 2.3.14 and the fact that the map from persistence diagram to persistence landscape is invertible for finitely presented single parameter persistence modules (22). For completeness, we provide a proof of Propostion 2.3.16. See Appendix 2.8.1 and Definition 2.8.4 for multiparameter persistence module presentations.

Proposition 2.3.16. *Let $M \in \mathbf{vect}_{fin}^{\mathbb{R}^n}$ be a finitely presented multiparameter persistence module and let $\lambda : \mathbb{N} \times \mathbb{R}^n \to \mathbb{R}$ be the associated multiparameter persistence landscape. Using λ we can recover the rank invariant of M on the set of pairs of parameter values which lie on lines of slope $\mathbf{1}$, $\{(\mathbf{a}, \mathbf{a} + r\mathbf{1}) : \mathbf{a} \in \mathbb{R}^n,\ r \geq 0\}$.*

Proof. The rank invariant on the set $\{(\mathbf{a}, \mathbf{a} + r\mathbf{1}) : \mathbf{a} \in \mathbb{R}^n,\ r \geq 0\}$ is derived from the landscape as follows:

$$\beta^{\mathbf{a},\mathbf{a}+r\mathbf{1}} = \limsup_{\delta \to 0^+} \max \left\{ k \in \mathbb{N} : \lambda(k, \mathbf{a} + \frac{(r+\delta)}{2}\mathbf{1}) \geq \frac{r+\delta}{2} \right\}.$$

Consider a pair of parameter values $(\mathbf{a}, \mathbf{a} + r\mathbf{1})$. If $\beta^{\mathbf{a},\mathbf{a}+r\mathbf{1}} \geq k$, since M is finitely presented, $\beta^{\mathbf{a},\mathbf{a}+(r+\delta)\mathbf{1}} \geq k$ for all sufficiently small δ. Let $\mathbf{c}_\delta = \mathbf{a} + \frac{(r+\delta)}{2}\mathbf{1}$. By Lemma 2.3.13 $\beta_M^{\mathbf{c}_\delta - \mathbf{h}, \mathbf{c}_\delta + \mathbf{h}} \geq k$ for all $\mathbf{h} \geq \mathbf{0}$ with $\|\mathbf{h}\|_\infty \leq \frac{(r+\delta)}{2}$, and thus we have that $\lambda(k, \mathbf{c}_\delta) \geq \frac{(r+\delta)}{2}$ for sufficiently small δ.

Conversely, suppose that $\lambda(k, \mathbf{c}_\delta) \geq \frac{(r+\delta)}{2}$ for some small postitive δ. This implies that $\beta^{\mathbf{a},\mathbf{a}+r\mathbf{1}} \geq \beta^{\mathbf{a},\mathbf{a}+(r+\frac{\delta}{2})\mathbf{1}} \geq k$. □

The following example illustrates two modules with distinct rank invariants which are not distinguished by their multiparameter persistence landscapes.

Example 2.3.17. *Denote by $[\mathbf{a}, \mathbf{b}]$ the rectangular interval module with opposite vertices $\mathbf{a}, \mathbf{b}$. Let $M = [(0,1),(10,2)] \oplus [(4,1),(6,2)]$ and $N = [(0,1),(6,2)] \oplus [(4,1),(10,2)]$ be interval decomposable 2-parameter persistence modules. These two modules have different rank invariant ($\beta_M^{(0,1),(10,2)} = 1 \neq 0 = \beta_N^{(0,1),(10,2)}$) but the same multiparameter persistence landscapes, see Figure 2.2.*

One would hope that a multiparameter module invariant could distinguish the modules in the previous example. The multiparameter persistence landscape fails to distinguish these modules since the rank invariants of these modules coincide on all pairs of parameter values lying on lines of slope $\mathbf{1}$. This in turn occurs since the overlap between the summands in the x_1-coordinate is greater than their persistence in the x_2-coordinate. Thus altering x_1, x_2 at the same rate we cannot detect the interaction between the summands with the multiparameter persistence landscape. A simple reparametrisation scaling parameters x_i appropriately would allow us to distinguish these modules and motivates the following definition.

Definition 2.3.18 ($\mathbf{w}$-Weighted Persistence Landscape)**.** *Let $\mathbf{w} \in \{\mathbf{u} \in \mathbb{R}^n : u_i > 0, \|\mathbf{u}\|_\infty = 1\}$ be a weighting vector corresponding to a rescaling of the parameter space $\mathbb{R}^n$. Define the $\mathbf{w}$-weighted infinity norm*

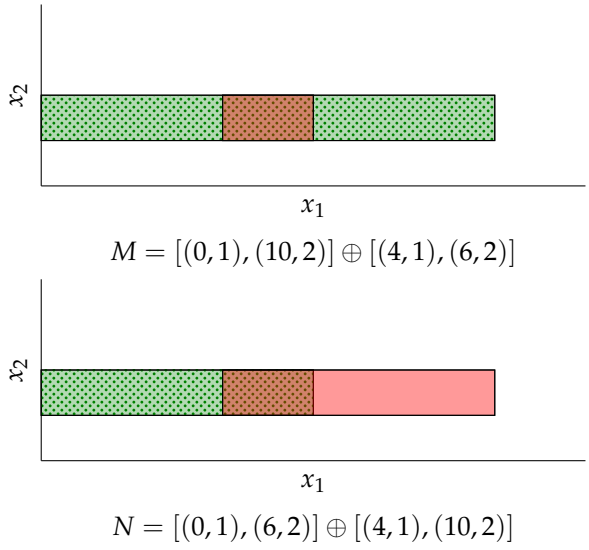


Figure 2.2: We illustrate a pair of interval decomposable multiparameter persistence modules, $M, N \in \mathbf{vect}^{\mathbf{R}^2}$, which have distinct rank invariants but the same multiparameter persistence landscapes. Each module is the direct sum of two rectangular intervals. The first summand of each module is shaded with green dots and the second summand is shaded in solid red.

to be $\|\mathbf{h}\|_\infty^{\mathbf{w}} = \|(w_i h_i)_i\|_\infty$. *The* $\mathbf{w}$*-Weighted Persistence Landscape is a function* $\lambda_{\mathbf{w}} : \mathbb{N} \times \mathbb{R}^n \to \overline{\mathbb{R}}$.

$$\lambda_{\mathbf{w}}(k, \mathbf{x}) = \sup\{\varepsilon \geq 0 : \beta^{\mathbf{x}-\mathbf{h}, \mathbf{x}+\mathbf{h}} \geq k \text{ for all } \mathbf{h} \geq \mathbf{0} \text{ with } \|\mathbf{h}\|_\infty^{\mathbf{w}} \leq \varepsilon\}.$$

Remark 2.3.19. The ordinary multiparameter persistence landscape is the $\mathbf{1}$-weighted persistence landscape.

The example illustrated in Figure 2.2 exhibits the dependence on the relative scaling of the significant parameter values in a multiparameter persistence module and highlights the importance of normalisation or choice of weighting vectors $\mathbf{w}$ in practical applications. Consider the weighted landscape with weighting $\mathbf{w} = (\frac{1}{10}, 1)$ for the examples in Figure 2.2. This weighted landscape distinguishes the two modules, indeed $\lambda_{\mathbf{w}}(M)(1,(5,1.5)) = 0.5$ and $= \lambda_{\mathbf{w}}(N)(1,(5,1.5)) = 0.1$.

The weighting hyperparameter for multiparameter persistence landscapes is analogous to the Poisson parameter used for the Poisson-weighted persistence landscape kernel introduced in (11), and the Weighting Function for persistence images introduced in (12). Determining an appropriate choice of these hyperparameters is dependent on a given applied setting.

In the case that your multiparameter persistence module is a sublevel set filtration of a simplicial complex K with filtration function $f = \prod_{i=1}^n f_i : K \to \mathbb{R}^n$, it may be desirable to choose a weighting such that the distributions of filtration values in each coordinate have similar variance. For example, in Section 2.6.2 we apply a log transform to the bandwidth parameter to ensure the distributions of filtration values in our bifiltration have similar variance in each coordinate. For

convenience of notation let us use shorthand notation for component wise multiplication $\mathbf{a} \odot \mathbf{b} = (a_i b_i)_i$.

Definition 2.3.20 (**w**-Rescaling). *Let $\varphi_{\mathbf{w}} \in Aut(\mathbb{R}^n)$ denote the invertible rescaling $\varphi_{\mathbf{w}}(\mathbf{x}) = \mathbf{w} \odot \mathbf{x_i}$ for $\mathbf{w} \in \{\mathbf{u} \in \mathbb{R}^n : u_i > 0\}$.*

The following proposition makes precise the relationship between the weighted landscape and rescaling the parameter space.

Proposition 2.3.21. *Let $\mathbf{w}$ be a rescaling vector and let $\lambda_{\mathbf{w}}$ denote the function taking a module to its $\mathbf{w}$-weighted persistence landscape. Let $(\varphi_{\mathbf{w}})^*$ denote the pull back of $\varphi_{\mathbf{w}}$. The $\mathbf{w}$-weighted persistence landscape is given by $\lambda_{\mathbf{w}} = (\mathrm{id} \times \varphi_{\mathbf{w}^{-1}})^* \circ \lambda_{\mathbf{1}} \circ (\varphi_{\mathbf{w}})^*$ and so the following diagram commutes:*

$$\begin{array}{ccc} \mathbf{vect}^{\mathbf{R}^n} & \xleftarrow{\quad (\varphi_{\mathbf{w}})^* \quad} & \mathbf{vect}^{\mathbf{R}^n} \\ \downarrow{\scriptstyle \lambda_{\mathbf{1}}} & & \downarrow{\scriptstyle \lambda_{\mathbf{w}}} \\ L^p(\mathbb{N} \times \mathbb{R}^n) & \xrightarrow{(\mathrm{id} \times \varphi_{\mathbf{w}^{-1}})^*} & L^p(\mathbb{N} \times \mathbb{R}^n). \end{array}$$

Proof. Let $M \in \mathbf{vect}^{\mathbf{R}^n}$. It is a straight forward definition chase to see the diagram commutes. Direct computation yields the result:

$$\begin{aligned} \lambda_{\mathbf{w}}(M)(k, \mathbf{x}) &= \sup\{\varepsilon \geq 0 : \beta_M^{\mathbf{x}-\mathbf{h}, \mathbf{x}+\mathbf{h}} \geq k, \text{ for all } \mathbf{h} \geq \mathbf{0} \text{ with } \|\mathbf{h}\|_\infty^{\mathbf{w}} \leq \varepsilon\} \\ &= \sup\{\varepsilon \geq 0 : \beta_{M \circ \varphi_{\mathbf{w}}}^{\mathbf{w}^{-1} \odot (\mathbf{x}-\mathbf{h}), \mathbf{w}^{-1} \odot (\mathbf{x}+\mathbf{h})} \geq k, \text{ for all } \mathbf{h} \geq \mathbf{0} \text{ with } \|\mathbf{w}^{-1} \odot \mathbf{h}\|_\infty^{\mathbf{w}} \leq \varepsilon\} \\ &= \sup\{\varepsilon \geq 0 : \beta_{M \circ \varphi_{\mathbf{w}}}^{\mathbf{w}^{-1} \odot \mathbf{x}-\mathbf{t}, \mathbf{w}^{-1} \odot \mathbf{x}+\mathbf{t}} \geq k, \text{ for all } \mathbf{t} \geq \mathbf{0} \text{ with } \|\mathbf{t}\|_\infty \leq \varepsilon\} \\ &= \lambda_{\mathbf{1}}(M \circ \varphi_{\mathbf{w}})(k, \mathbf{w}^{-1} \odot \mathbf{x}) = (\mathrm{id} \times \varphi_{\mathbf{w}^{-1}})^* \circ \lambda_{\mathbf{1}} \circ (\varphi_{\mathbf{w}})^*(M)(k, \mathbf{x}). \end{aligned}$$

□

2.4 *Stability and Injectivity*

In this section we shall show that the multiparameter landscapes are stable with respect to the interleaving distance and persistence-weighted Wasserstein distance. We will then provide an injectivity result that shows the collection of weighted persistence landscapes derived from a persistence module contains almost all the information in the rank invariant of that module.

For each $p \in [1, \infty]$ let us define a p-distance on the space of multiparameter landscapes completely analogously with the definition as in (10) where we implicitly are viewing our landscapes as elements of Lebesgue space $L^p(\mathbb{N} \times \mathbb{R}^n)$. Our landscapes are all measurable since they are continuous; however they may be unbounded. We can either choose to permit infinite distances or alternatively truncate our landscapes to a bounded region if we wish to ensure that our distances are finite.

Definition 2.4.1 (p-Landscape Distance)**.** *Let M, N be multiparameter persistence modules. The p-landscape distance between M and N, $d_\lambda^{(p)}(M, N)$, is defined to be:*

$$d_\lambda^{(p)}(M, N) = \|\lambda(M) - \lambda(N)\|_p.$$

2.4.1 Stability

Unlike the infinity norm of rank function and rank invariant, the infinity norm of multiparameter persistence landscapes is stable with respect to the interleaving distance.

Theorem 2.4.2 (Multiparameter Persistence Landscape Stability)**.** *If $M, N \in \boldsymbol{vect}^{\mathbb{R}^n}$ are multiparameter persistence modules then the ∞-landscape distance of the multiparameter persistence landscapes is bounded by the interleaving distance d_I.*

$$d_\lambda^{(\infty)}(M, N) \leq d_I(M, N).$$

Proof. Suppose M, N are ε-interleaved. Let $\mathbf{x} \in \mathbb{R}^n$ and assume without loss of generality that $r = \lambda(M)(k, \mathbf{x}) \geq \lambda(N)(k, \mathbf{x})$ and that also $\lambda(M)(k, \mathbf{x}) \geq \varepsilon$.

For any $\mathbf{h} \geq \mathbf{0}$ with $\|\mathbf{h}\|_\infty < r - \varepsilon$ we have that $\|\mathbf{h} + \varepsilon\mathbf{1}\|_\infty < r$. Since $r = \lambda(M)(k, \mathbf{x})$ we know that the map $M(\mathbf{x} - (\mathbf{h} + \varepsilon\mathbf{1}) \leq \mathbf{x} + (\mathbf{h} + \varepsilon\mathbf{1}))$ has rank at least k.

An ε-interleaving gives rise to commutative diagram:

$$\begin{array}{ccc} M(\mathbf{x} - (\mathbf{h} + \varepsilon\mathbf{1})) & \longrightarrow & M(\mathbf{x} + (\mathbf{h} + \varepsilon\mathbf{1})) \\ & \searrow \quad \nearrow & \\ & N(\mathbf{x} - \mathbf{h}) \longrightarrow N(\mathbf{x} + \mathbf{h}) & \end{array}$$

Thus we see that the map $N(\mathbf{x} - \mathbf{h} \leq \mathbf{x} + \mathbf{h})$ has rank at least k. Hence $\lambda(N)(k, x) \geq r - \varepsilon$ and thus the infinity norm distance between $\lambda(M)$ and $\lambda(N)$ is at most ε. □

Corollary 2.4.3 (Multiparameter Sublevel-set Landscape Stability Theorem)**.** *Let $f, g : X \to \mathbb{R}^n$ be filtering functions and let $M(f), M(g)$ denote the induced sublevel-set multiparameter persistence modules. The sublevel-set persistence modules satisfy:*

$$d_\lambda^{(\infty)}(M(f), M(g)) \leq \|f - g\|_\infty.$$

Proof. $M(f), M(g)$ are $\|f-g\|_\infty$-interleaved. □

In practical applications we may wish to truncate our landscapes to a bounded region of the parameter space $R \subset \mathbb{R}^n$. Since $\|h\mathbb{1}_R\|_p \leq |R|\|h\mathbb{1}_R\|_\infty$, the ∞-landscape distance stability result yields a coarse bound for the p-landscape distance.

Corollary 2.4.4. *Let M, N be multiparameter persistence modules and $R \subset \mathbb{N} \times \mathbb{R}^n$ a Lebesgue measurable subset of the parameter space. The p-landscape distance restricted to the region R is stable with respect to the interleaving distance:*

$$\|(\lambda(M) - \lambda(N))\mathbb{1}_R\|_p \leq |R| d_I(M, N).$$

The weighted landscapes also satisfy stability with respect to the interleaving distance. This can be shown directly or using Proposition 2.3.21 and the stability result in the unweighted case.

Corollary 2.4.5 (Multiparameter $\mathbf{w}$-Weighted Landscape Stability)**.** *Let M, N be multiparameter persistence modules. For unit weightings $\mathbf{w} \in \{\mathbf{u} \in \mathbb{R}^n : u_i > 0, \|\mathbf{u}\|_\infty = 1\}$ the ∞-Landscape distance of the weighted Multiparameter Persistence Landscapes is bounded by the interleaving distance.*

$$\|\lambda_{\mathbf{w}}(M) - \lambda_{\mathbf{w}}(N)\|_\infty \leq d_I(M, N).$$

Proof. Suppose that M, N are ε-interleaved and that $\mathbf{w}$ is a unit weighting. Since $\mathbf{w}$ is a unit weighting $(\varphi_{\mathbf{w}})^*(M), (\varphi_{\mathbf{w}})^*(N)$ are also ε-interleaved. We attain that:

$$\begin{aligned}
\|\lambda_{\mathbf{w}}(M) - \lambda_{\mathbf{w}}(N)\|_\infty &= \|(\mathrm{id} \times \varphi_{\mathbf{w}^{-1}})^* \circ \lambda_{\mathbf{1}} \circ (\varphi_{\mathbf{w}})^*(M) - (\mathrm{id} \times \varphi_{\mathbf{w}^{-1}})^* \circ \lambda_{\mathbf{1}} \circ (\varphi_{\mathbf{w}})^*(N)\|_\infty \\
&= \|\lambda_{\mathbf{1}} \circ (\varphi_{\mathbf{w}})^*(M) - \lambda_{\mathbf{1}} \circ (\varphi_{\mathbf{w}})^*(N)\|_\infty \\
&\leq d_I((\varphi_{\mathbf{w}})^*(M), (\varphi_{\mathbf{w}})^*(N)) \leq d_I(M, N).
\end{aligned}$$

The first equality is a result of Proposition 2.3.21 and the penultimate inequality is a direct application of Theorem 2.4.2. □

The p-landscape distance restricted to interval decomposable modules is stable with respect to the persistence weighted p-Wasserstein distance.

Proposition 2.4.6 (p-Landscape Distance Stability of Interval Decomposable Modules)**.** *Let M, N be interval decomposable multiparameter modules with barcodes consisting of finitely many intervals. The p-landscape distance is stable with respect to the persistence weighted p-Wasserstein distance:*

$$d_\lambda^{(p)}(M, N) \leq d_{\overline{W}_p}(M, N).$$

Proof. Let us use the shorthand notation $\lambda_M = \lambda(M)$, and suppose M, N have barcodes $\{I_j : j \in \mathcal{J}\}$ and $\{J_\kappa : \kappa \in \mathcal{K}\}$ with equal cardinality by appending a set of empty intervals of the cardinality of the other set. Recall that the landscape for M can be expressed as a pointwise maximum, $\lambda_M(k, \mathbf{x}) = \operatorname{kmax}_{\mathcal{J}} \lambda_{\mathbb{1}_{I_j}}(1, \mathbf{x})$. Let $\sigma : \mathcal{J} \to \mathcal{K}$ be any bijection realising the persistence weighted p-Wasserstein distance, and let $\varepsilon_j = d_I(\mathbb{1}_{I_j}, \mathbb{1}_{J_{\sigma j}})$.

$$\begin{aligned}
d^{(p)}_{\lambda_\infty}(M,N)^p = \|\lambda_M - \lambda_N\|_p^p &= \sum_{k=1}^{\infty} \int_{\mathbb{R}^n} |\lambda_M(k,\mathbf{x}) - \lambda_N(k,\mathbf{x})|^p d\mu \\
&= \int_{\mathbb{R}^n} \sum_{k=1}^{\infty} |\operatorname{kmax}_{\mathcal{J}} \lambda_{\mathbb{1}_{I_j}}(1,\mathbf{x}) - \operatorname{kmax}_{\mathcal{K}} \lambda_{\mathbb{1}_{J_\kappa}}(1,\mathbf{x})|^p d\mu \\
&\leq \int_{\mathbb{R}^n} \sum_{j \in \mathcal{J}} |\lambda_{\mathbb{1}_{I_j}}(1,\mathbf{x}) - \lambda_{\mathbb{1}_{J_{\sigma(j)}}}(1,\mathbf{x})|^p d\mu \\
&\leq \int_{\mathbb{R}^n} \sum_{j \in \mathcal{J}} \varepsilon_j^p \mathbb{1}_{\{I_j \cup J_{\sigma(j)}\}} d\mu \\
&= \sum_{j \in \mathcal{J}} |I_j \cup J_{\sigma(j)}| \varepsilon_j^p = d_{\overline{W}_p}(M,N)^p.
\end{aligned}$$

The inequality between the second and third line follows from the general fact that for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ the sum $\sum |u_i - v_i|^q$ is minimized by ordering the components of each tuple. The fourth line bounds the third line by Theorem 2.4.2 applied to the matched interval summands. □

2.4.2 *Injectivity*

We now show that the collection of weighted landscapes associated to a module preserves almost all the information contained in the rank invariant of finitely presented persistence modules. See Appendix 2.8.1 and Definition 2.8.4 for multiparameter persistence module presentations.

Definition 2.4.7 (Weighted Landscape Space)**.** *Let* $\mathbf{W} := \{\mathbf{u} \in \mathbb{R}^n : u_i > 0, \|\mathbf{u}\|_\infty = 1\}$ *denote the set of unit weights. Each multiparameter module* M *gives rise to a function* $\lambda_{\mathbf{W}}(M) : \mathbf{W} \to L^\infty(\mathbb{N} \times \mathbb{R}^n)$ *mapping each unit weight to the associated weighted landscape of that module. We define weighted landscape space to be this function space equipped with the metric:*

$$d_{\mathbf{W}}(\lambda_{\mathbf{W}}(M), \lambda_{\mathbf{W}}(N)) = \|\lambda_{\mathbf{W}}(M) - \lambda_{\mathbf{W}}(N)\|_\infty = \sup_{\mathbf{w} \in \mathbf{W}} \{\|\lambda_{\mathbf{w}}(M) - \lambda_{\mathbf{w}}(N)\|_\infty\}.$$

Proposition 2.4.8. *Let* $\mathbf{vect}^{\mathbf{R}^n}_{fin}$ *denote the space of finitely presented multiparameter persistence modules. Let us define an equivalence relation on*

$\mathbf{vect}_{fin}^{\mathbf{R}^n}$ *identifying* $M \sim N$ *if the rank invariant of* M *and* N *coincide. Denote the quotient space of finitely presented multiparameter persistence modules under this equivalence relation by* $\mathbf{vect}_{fin}^{\mathbf{R}^n} / \sim$. *The map* $\lambda_{\mathbf{W}} : M \mapsto \{(\mathbf{w}, \lambda_{\mathbf{w}}(M)) : w_i > 0, \|\mathbf{w}\|_\infty = 1\}$ *is injective and* 1*-Lipschitz on* $\mathbf{vect}_{fin}^{\mathbf{R}^n} / \sim$, *where we equip the quotient space with the distance induced by the interleaving distance, and the weighted landscape space with the metric* $d_{\mathbf{W}}$.

Proof. We shall first show that the map λ is injective. Let $M \in \mathbf{vect}_{\text{fin}}^{\mathbf{R}^n}$ and recall Proposition 2.3.16. For all $\mathbf{a} < \mathbf{b}$ there is some unit rescaling vector $\mathbf{w}$ such that the pair of parameter values $(\varphi_{\mathbf{w}}(\mathbf{a}), \varphi_{\mathbf{w}}(\mathbf{b}))$ lie on a line of slope $\mathbf{1}$. Thus we can recover the rank invariant of M from the collection $\{(\mathbf{w}, \lambda_{\mathbf{w}}(M))\}$. The fact that the map $\lambda_{\mathbf{W}}$ is 1-Lipschitz is an immediate consequence of Corollary 2.4.5 Multiparameter $\mathbf{w}$-Weighted Landscape Stability. □

Since $\lambda_{\mathbf{W}}$ is 1-Lipschitz we can compute a lower bound on the interleaving distance between modules from the collection of weighted landscapes. We would be interested to investigate further the relationship between the landscape distance and the interleaving distance to understand when the landscape distance provides a good lower bound for the interleaving distance.

2.5 *Statistics on Multiparameter Landscapes*

A principal advantage of working with landscapes as a summary statistic for our data is that we are always able to take the pointwise mean of a collection of landscapes. Whilst the mean of a collection of landscapes is not necessarily the landscape of some module, one can still interpret the mean landscape. The location of local maxima of the mean landscape correspond to the parameter values above which significant features of the sample landscapes live, which in turn correspond to significant topological features of the data set from which the sample landscapes have been derived.

The space of persistence landscapes endowed with the p-landscape distance naturally spans a subspace of Lebesgue space, a Banach space. We would like to perform statistical analysis on a set landscapes produced from data sets to distinguish significant topological signals from sampling noise. In Appendix 2.9 we review relevant results from the theory of Banach Space valued random variables. In this section we apply these results to multiparameter persistence landscapes. We attain the same collection of results enjoyed by the single parameter persistence landscape established in (10).

2.5.1 *Convergence Results for Multiparameter Landscapes*

We shall take the same probabilistic approach as in (10) in viewing multiparameter landscapes derived from a data set as a Banach space valued random variable. The model for applying statistical analysis to persistence landscapes will trace the following general setup.

Suppose X is a Borel measurable random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ thought of as sampling data from some distribution. Further let $\Lambda = \Lambda(X)$ denote the multiparameter persistence landscape associated to some multifiltration of the data X, so that in summary $\Lambda : (\Omega, \mathcal{F}, \mathbb{P}) \to L^p(\mathbb{N} \times \mathbb{R}^n)$ for $1 \leq p < \infty$ is a random variable taking values in a real, separable Banach Space.

Let $\{X_i\}$ be i.i.d copies of X and $\{\Lambda_i\}$ their associated landscapes. Denoting the pointwise mean of the first n landscapes by $\overline{\Lambda}^n$ and applying the general theory of probability in Banach spaces presented in Appendix 2.9 we attain several results. Observe that in practice we may be required to truncate our multiparameter landscapes to a bounded region in order to satisfy the finiteness criteria in the convergence results.

Associated to a well-behaved Banach space valued random variable $\Lambda : (\Omega, \mathcal{F}, \mathbb{P}) \to L^p(\mathbb{N} \times \mathbb{R}^n)$ is a set function $I_\Lambda : \mathcal{F} \to \mathcal{B}$ called the Pettis Integral of Λ. This can be thought of as the expectation of a Banach space valued random variable. For more details see Appendix 2.9 and (33).

Theorem 2.5.1 (Strong Law of Large Numbers)**.** *With our notation as in the above discussion $\overline{\Lambda}^n \to I_\Lambda(\Omega)$ almost surely if and only if $\mathbb{E}[\|\Lambda\|] < \infty$.*

Theorem 2.5.2 (Central Limit Theorem)**.** *Let us consider the landscapes endowed with the p-landscape distance for $p \geq 2$. If $\mathbb{E}[\|\Lambda\|] < \infty$ and $\mathbb{E}[\|\Lambda^2\|] < \infty$, then $\sqrt{n}(\overline{\Lambda}^n - I_\Lambda(\Omega))$ converges weakly to a Gaussian random variable $G(\Lambda)$ with the same covariance structure as Λ.*

The central limit theorem for the landscapes induces a central limit theorem for associated real valued random variables and facilitates the computation of approximate confidence intervals.

Corollary 2.5.3. *Let us consider the landscapes endowed with the p-landscape distance for $p \geq 2$. Suppose $\mathbb{E}[\|\Lambda\|] < \infty$ and $\mathbb{E}[\|\Lambda^2\|] < \infty$. If $f \in L^p(\mathbb{N} \times \mathbb{R}^n)^*$, so that $Y = f(\Lambda)$ is a real valued random variable, then $\sqrt{n}(\overline{Y}^n - \mathbb{E}[Y]) \to \mathcal{N}(0, Var(Y))$ converges in distribution.*

Corollary 2.5.4 (Approximate Confidence Intervals)**.** *Suppose Y is a real-valued random variable attained from a functional applied to the multiparameter landscape Λ satisfying the conditions of Corollary 2.5.3. Let $\{Y_i\}_{i=1}^n$ be i.i.d. instances of this random variable and $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i -$*

$\overline{Y}_n)^2$ *the sample variance. An approximate* $(1-\alpha)$ *confidence interval for* $\mathbb{E}[Y]$ *is given by:* $[\overline{Y}_n - z_{\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}, \overline{Y}_n + z_{\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}]$*, where* $z_{\frac{\alpha}{2}}$ *is the* $\frac{\alpha}{2}$ *critical value for the normal distribution.*

In practice, a functional of choice could be given by integrating the landscapes over a subset R of the parameter domain $f_R(\Lambda) = \int_R \Lambda \, d\mu$. These functionals can be used to establish the significance of homological features in different regions of the parameter space.

We remark that recent work has attained confidence bands for single parameter persistence landscapes (34; 35). It is shown in (34) that the single parameter persistence landscapes satisfy a uniform central limit theorem, and moreover the rate of convergence is computed. It would be interesting to see similar analysis performed in the multiparameter setting.

2.6 *Example Computations and Machine Learning Applications*

In this section we shall present example computations of multiparameter persistence landscapes and demonstrate a simple application of machine learning to the persistence landscapes. We use the RIVET software (36) for computations of 2-parameter persistence modules presented in (23) together with (37) to compute multiparameter persistence landscapes from the 2-parameter persistence modules. RIVET supports the fast computation of multigraded Betti numbers (see Definition 2.8.5) and an interactive visualisation for 2-parameter persistence modules. The software computes a data structure associated to a module which facilitates real time queries of the fibered barcode. As far as we know, RIVET is the only publicly available TDA software package supporting multiparameter persistent homology calculations.

The software supports a range of input formats including: point cloud, metric space, algebraic chain complex, and explicit bifiltered complex. In particular we shall use the software to calculate and query the fibered barcode associated to a module along a selection of one dimensional slices of the parameter space.

In order to reduce computational cost, RIVET approximates multiparameter modules with a discretization. These approximations can be taken to arbitrary accuracy with respect to the interleaving distance, see Appendix 2.8.3. Discretization is not the only approach one can take in computations involving continuous multiparameter modules. In contrast (21) develops a primary decomposition of modules which facilitates a finite description of a wide class of persistence modules which would require infinitely many generators if discretized. Our only obstruction to using this approach rather

than discretization is the lack of available software to cope with these presentations.

Computation of the module with RIVET is the most computationally expensive procedure in our calculations. Details of the time and space complexity of the algorithm may be found in (23), loosely if m denotes the size of the filtered complex associated to the input data, in the worst case one requires time $O(m^5)$ and space $O(m^5)$ to compute the data structure which admits fast queries of the fibered barcode. Further details of the software and complexity may be found in (23).

In theory, since our landscape is derived solely from the rank invariant, we need not calculate the full module and fibered barcode. Recall that the value of the multiparameter persistence landscape at each point can be calculated using the single parameter persistence landscape associated to the line of slope $\mathbf{1}$ passing through that point. Thus we could reduce the computation of the multiparameter landscape in any dimension to repeated single parameter persistent homology calculations. This reduction would be highly parallelizable and likely to provide significant speedup.

Proposition 2.6.1. *Let $M \in \mathbf{vect}^{\mathbf{R}^2}$ be a multiparameter persistence module derived from a simplicial complex with m simplices. Let $\varepsilon > 0$ be some tolerance value and $[0, R] \times [0, R] \subset \mathbb{R}^2$ a subset of our parameter space. We can compute an ε-approximation $\lambda(M)^{(\varepsilon)}$ to the persistence landscape $\lambda(M)$ of M on the region $[0, R] \times [0, R]$ in time $O(m^3 \frac{R}{\varepsilon})$. Our approximation is with respect to the infinity norm, that is $\|\lambda(M)^{(\varepsilon)} - \lambda(M)\|_\infty \leq \varepsilon$.*

Proof. Divide the region $[0, R] \times [0, R]$ into a grid of spacing ε. It suffices to calculate the values of the landscape on this grid since the landscape functions are 1-Lipschitz and so we can extend the grid values to an ε-approximate function on $[0, R] \times [0, R]$. Thus we reduce our computation to the computation of $\frac{2R}{\varepsilon}$ single parameter landscapes corresponding to the collection of $\frac{2R}{\varepsilon}$ slope $\mathbf{1}$ lines passing through the points of the grid.

Given birth-death pairs, (38) provide an algorithm to compute the persistence landscapes in time $O(m^2)$. It is well known from (6) that one can produce birth-death pairs from a filtration of size m in time $O(m^3)$. Hence the result follows. □

It is possible that the above time estimate for the landscape computation could be improved by using vineyard style updates between the single parameter landscapes (39). Moreover it may be that in practical applications, computing the module with RIVET is significantly faster than the worst time bound $O(m^5)$ and thus using the fibered barcode queries will be faster than the computation of a series

of single parameter landscapes. Note also that the $\frac{2R}{\varepsilon}$ single parameter landscape calculations are independent and so can be computed in parallel. We postpone comparisons of different computational algorithms, benchmarking, and efficient implementation to follow up work.

One may want to utilise machine learning algorithms with landscape functions as a collection of features for a data set. Recall that if we consider the persistence landscapes associated with the 2-landscape distance then we are naturally in the setting of a Hilbert Space. The inner product on this space is positive definite on the space of persistence landscapes. As such we may use this kernel to learn non-linear relationships in our data and then apply convex optimisation techniques to an SVM.

Another point to note is that we can discretize the landscape to give an n-dimensional array as a summary of our data to which one could apply a convolutional neural network. This transform from landscape to multidimensional array will satisfy stability with respect to the landscape distance. A similar approach is used by (12) to produce a *persistence image* from a persistence diagram.

We provide three computational examples together with the application of a basic statistical test and standard SVM classifier. Our examples demonstrate that the multiparameter landscape is sensitive to both topology and geometry. We do not claim that the multiparameter landscape is the optimal analytic tool to perform the various tasks in our examples, rather we demonstrate a range of potential applications.

2.6.1 *Concentric Circles*

Our first example will look at points sampled from densities concentrated around a pair of concentric circles with radii 1 and 3 respectively. We colour the points from each circle in two distinct ways. Colouring A assigns the large circle colour parameter 0.5 and the small circle colour parameter 1.5. Colouring B assigns the small circle colour parameter 0.5 and the large circle colour parameter 1.5. We examine how the multiparameter landscapes differ depending on the colouring of the circles. For each colouring we perform 30 samples, each sample consisting of 100 points uniformly sampled from each circle Figure 2.3(a).

We produce a filtration on each pointcloud with the Rips filtration in the first parameter and the colour parameter in the second parameter. Thus at parameter value $(r, c) \in \mathbb{R}^2$, we have the space $X_{(r,c)} = \mathrm{VR}(\mathcal{P}_c, r)$ where $\mathcal{P}_c$ denotes the sampled points with colour parameter no more than c. We apply the degree one homology func-

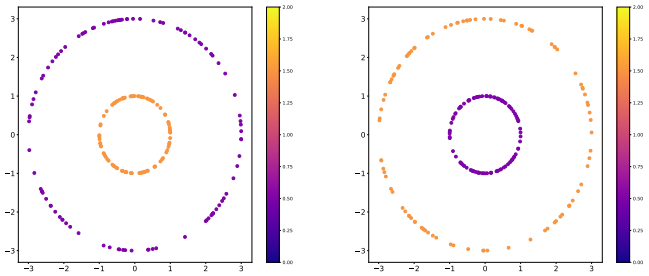
tor H_1 to produce a multiparameter persistence module which detects the loops in our filtered topological space.

We compute the average landscapes of the H_1-modules for the two different colourings, Figure 2.3. When the large circle has the smaller colour parameter value, the first landscape ($k = 1$) can detect the large circle Figure 2.3(b). We see the large circle in the first landscape as the large mountain spanning the parameter subspace $[1, 5.4] \times [0.5, 1.5]$. When the large circle has the higher parameter value, the persistence in the Rips filtration parameter is diminished by the presence of the small circle with smaller colour parameter. In both colourings, the second landscape ($k = 2$) exhibits the range of parameter values for which both circles are detected Figure 2.3(c).

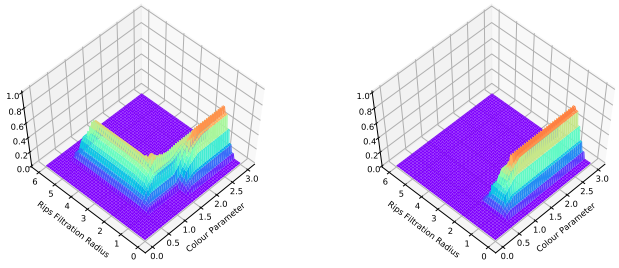
We test the robustness of the landscape by repeating the sampling this time with only 50 points per circle and perturbing both the location and colour of the sampled points with the addition of i.i.d. normals $\mathcal{N}(0, 0.3)$, Figure 2.4(a). We illustrate in Figure 2.4(b) and Figure 2.4(c) the average landscapes taken over 30 noisy samples. The resulting landscapes are similar to those of the larger samples without noise.

Let us perform a statistical test to determine whether the multiparameter landscapes can detect that the noisy samples are drawn from different distributions. Let μ denote the Lebesgue measure on $\mathbb{R}^2$ and consider the functional $f_R(\lambda) = \int_R \lambda d\mu$. Using the results of Section 2.5 we find approximate confidence intervals for $f_R(\lambda)$ with $R = \{1\} \times ([2, 6] \times [0, 1.5]) \subset \mathbb{N} \times \mathbb{R}^2$. We attain approximate 99%-confidence intervals on the noisy samples: for Colouring A $[0.400, 0.474]$, and for Colouring B $[0.00556, 0.00809]$. A two sample t-test on the values of this functional on the two sets of colourings attains a p-value of 0.00629. Thus we reject the null hypothesis that the functional values on the landscapes of the two colourings have the same mean.

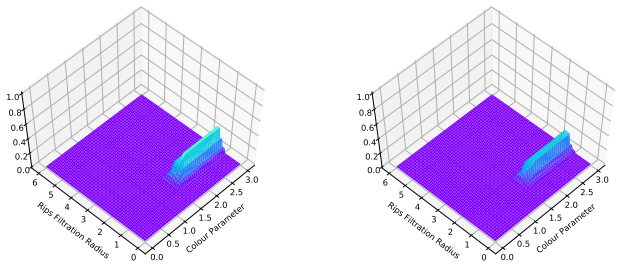
We also perform a permutation test for the test statistic $\|\bar{\lambda}^A - \bar{\lambda}^B\|_4$ (with norm taken for the first landscape over the whole parameter space $[0, 6] \times [0, 3]$). We apply 10,000 random permutations to the collection of landscapes randomly assigning 30 landscapes to group A and 30 landscapes to group B and then computing the test statistic. Each of the sampled permutations attained a smaller test statistic than the observed test statistic, (indicating an approximate p-value of 0), and identifying that the landscapes for Colouring A have been drawn from a distinct distribution than the landscapes for Colouring B.



(a) An example point cloud sample from each colouring.

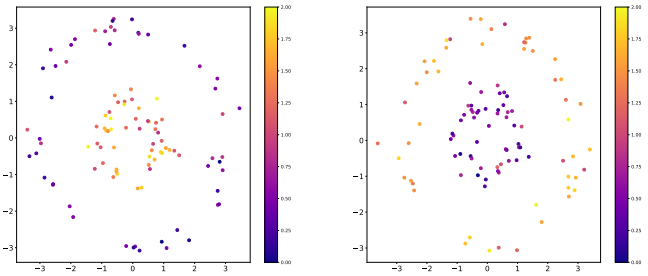


(b) The mean first landscape for each colouring taken over the 30 samples, $\overline{\lambda}(1, \mathbf{x})$.

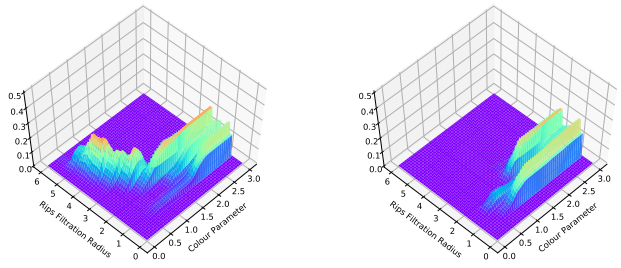


(c) The mean second landscape for each colouring taken over the 30 samples, $\overline{\lambda}(2, \mathbf{x})$.

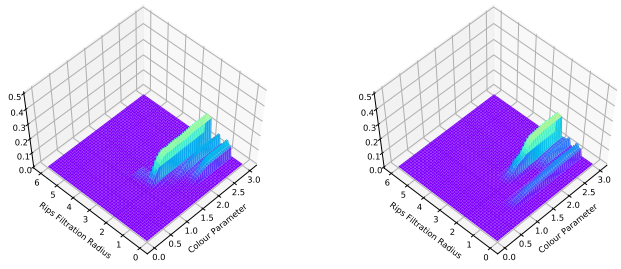
Figure 2.3: The first column shows the plots for Colouring A and the second column Colouring B.



(a) An example point cloud sample from each colouring with noise added.



(b) The mean first landscape $\overline{\lambda}(1, \mathbf{x})$ taken over the 30 noisy samples.



(c) The mean second landscape $\overline{\lambda}(2, \mathbf{x})$ taken over the 30 noisy samples.

Figure 2.4: The first column shows the plots for Colouring A and the second column Colouring B.

2.6.2 *Modal Estimation*

For this example we work on meteorite data which we have lifted from (40). The data set consists of values of the proportion of silica measured in 22 samples. Our task is to infer how many modes there are in the distribution from which this data has been sampled. Various statistical techniques for assessing the number of modes in empirical data have previously been explored (41; 42; 43).

A standard approach to this task is kernel density estimation (KDE). With data $\{x_i\} \subset \mathbb{R}^n$ one estimates the probability density function (pdf) of the distribution using a sum of normalized kernels:

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} K_\sigma(x - x_i)$$

Here K_σ is a density function with mass concentrated about the origin, for example a Gaussian centred at the origin. There are two natural parameters in this KDE setup. The *log bandwidth* parameter σ of the kernel function K_σ, and a *threshold* parameter which dictates how large a peak in the estimated distribution must be to be considered a mode. The choice of these parameters will dramatically alter our inferred number of modes (see Figure 2.5).

Figure 2.6 is a surface plot of the KDEs ranging over various bandwidth parameters, demonstrating the change in the number of modes as we change the bandwidth. The surface has been triangulated using a triangulation subordinate to a regular grid on our parameter space. To each 2-simplex τ in the triangulation we attach two parameters; the mean bandwidth $\sigma(\tau)$, and the mean probability density value $p(\tau)$, (averages taken over the vertices of the simplex). We produce a bifiltration by taking the simplicial closure of the 2-simplices with appropriate parameter values, $X_{(\sigma_0, p_0)} = \mathrm{SC}(\{\tau : \sigma(\tau) \leq \sigma_0, p(\tau) \geq 1 - p_0\})$.

The multiparameter landscape detects that three modes appear in the KDEs for a range of parameter values. Looking at the landscapes associated to the H_0-module we see that the infinity norm of the first three landscapes is constant but decreases significantly between the

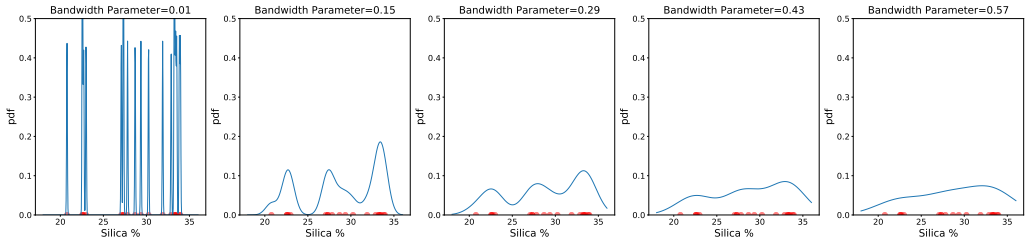


Figure 2.5: We plot kernel density estimates on the meteorite data (red) for a range of bandwidth parameters. As we increase the bandwidth parameter we yield fewer modes in our kernel density estimate.

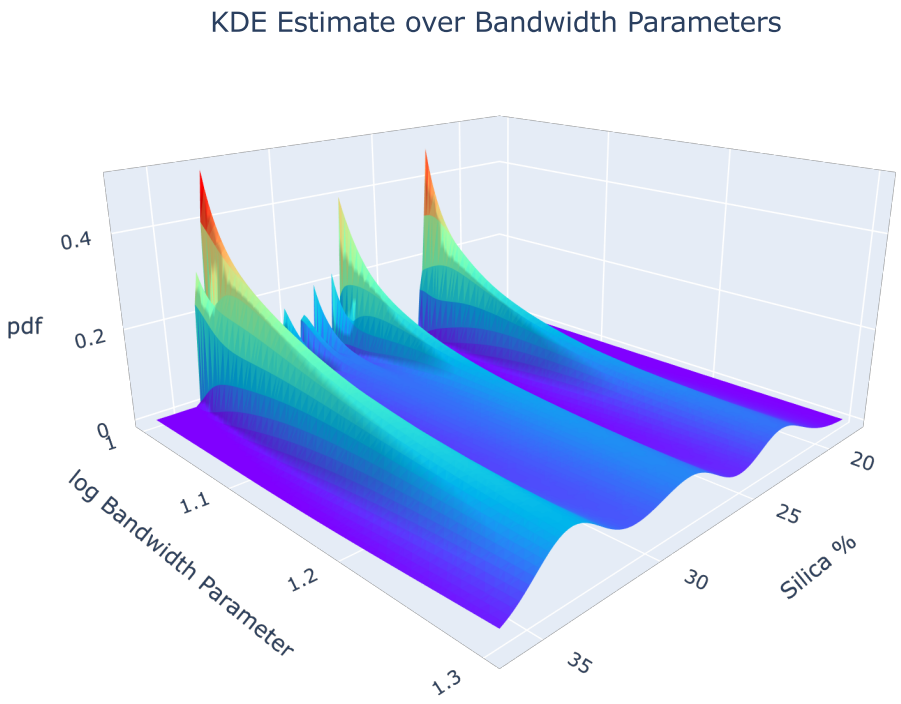


Figure 2.6: A triangulated surface plot of the KDE for a range of bandwidth parameters. We observe three modes in the KDE estimate for a large range of bandwidth values.

third and fourth landscapes, Figure 2.7. This indicates that within this setup, three modes are seen across a significantly wider range of parameter values than four modes, suggesting the data is drawn from a tri-modal distribution which coincides with our expected result. Whilst in this simple example one could suggest there are three modes from inspection, the landscape analysis can equally be applied to higher dimensional data sets for which visualisation is not possible.

This basic example can be generalized to detect other properties of KDEs robust to changes in parameter values. For example one could detect significant i-dimensional holes in the distribution by considering the H_i module in a similar setup. For related work see Persistence Terraces (44).

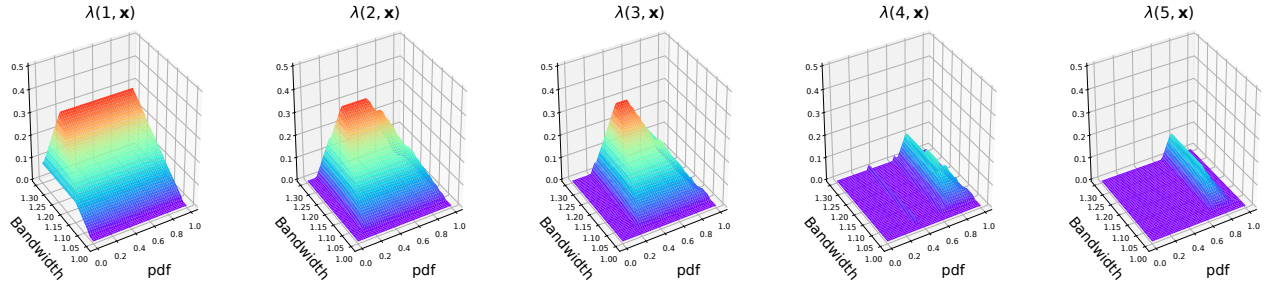


Figure 2.7: The first to fifth multiparameter persistence landscapes associated to the H_0 module for the KDE surface. The infinity norm of the first three landscapes is constant $\|\lambda(1,\mathbf{x})\|_\infty = \|\lambda(2,\mathbf{x})\|_\infty = \|\lambda(3,\mathbf{x})\|_\infty = 0.283$ but drops significantly between the third and fourth landscape $\|\lambda(4,\mathbf{x})\|_\infty = 0.120$, $\|\lambda(5,\mathbf{x})\|_\infty = 0.106$. This indicates that there are three modes in the KDE for a wide range of bandwidth parameters.

2.6.3 *Curvature*

In this subsection we shall work with a synthetic data set sampled from spaces of different curvature. Single parameter persistence landscapes have been used to detect curvature (8). This example is used to emphasise the ability of the multiparameter landscapes to detect geometric differences between point samples. The samples consist of 100 points chosen uniformly with respect to the volume measure from discs of radius 1 in the hyperbolic plane, the surface of the unit sphere and Euclidean space so that the spaces have constant curvature of $-1, 1, 0$ respectively. Topologically these disks are all trivial, our landscapes are detecting geometric differences induced by the distribution of points.

We would like to show that the multiparameter landscape is able to detect the curvature of the space from which a sample is drawn given only the pairwise distances between points. A multifiltered complex is built on the sampled points by filtering the Rips complex with the third nearest neighbour density function ρ on the points.

Explicitly, if $\mathcal{P}$ denotes our sampled points and $(r, \rho_0) \in \mathbb{R}^2$ then $X_{(r,\rho_0)} = \mathrm{VR}(\mathcal{P}_{\rho_0}, r)$ where $\mathcal{P}_{\rho_0} = \{p \in \mathcal{P} : \rho(p) \leq \rho_0\}$ for the third nearest neighbour density function ρ. We take 100 samples of 100 points in each space and investigate the resulting multiparameter landscapes for dimension 1 homology.

We plot the average first multiparameter landscapes in Figure 2.8(a) and the differences between the average landscapes in Figure 2.8(b). We observe that the persistence of cycles is affected by the curvature of the space. The more negative the curvature the longer the one dimensional cycles persist.

Let us now apply a simple machine learning algorithm to the multiparameter landscapes to see if we can reliably distinguish the curvature of the space from which our small samples have been drawn.

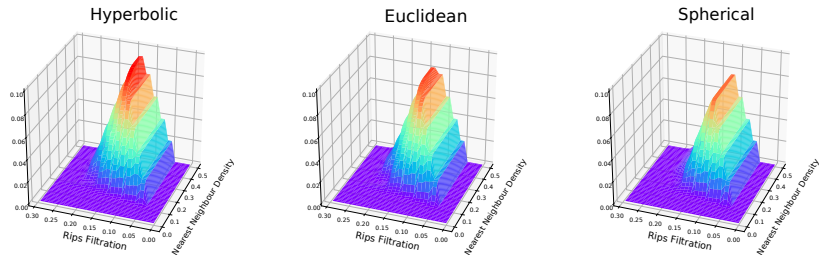
Using the Python package LinearSVC, we train a Support Vector Machine (SVM) with linear kernel on discretizations of the first 10 landscapes for the samples of the hyperbolic discs and elliptic discs, using l^2 penalty and squared hinge loss function. We randomly partition our samples into 160 training samples and 40 test samples and evaluate the accuracy by the proportion of test samples correctly classified. Repeating this process 100 times we attain an average classification score of 85.78% (3.70% standard deviation). Thus we see that the multiparameter landscapes are able to reliably detect curvature given a relatively small local sample. It is possible that alternative choices of filtration parameters may be better suited to detecting curvature.

2.7 *Discussion*

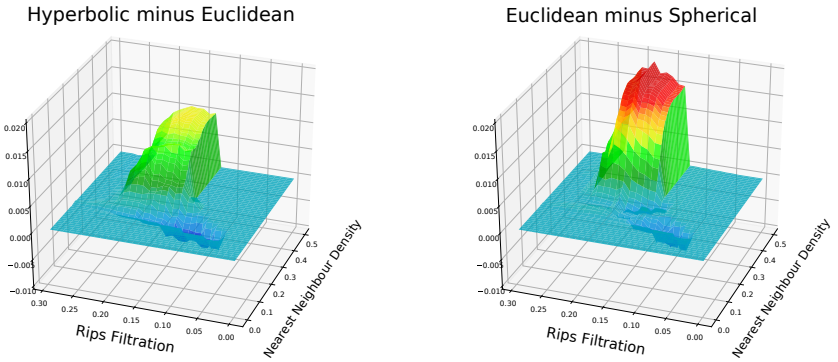
Multiparameter persistence landscapes provide a stable representation of the rank invariant of a persistence module whilst retaining the discriminating power of the rank invariant. Moreover the landscape distance provides a computable lower bound for the optimal stable distance on persistence modules, the interleaving distance.

The multiparameter landscape also offers a bridge from topological data analysis to machine learning and statistical analysis of multiparameter modules. The multiparameter landscapes, although hard to visualize in dimensions higher than 2, are interpretable in any dimension with large landscape values indicating features robust to changes in the filtration parameters, and non-zero landscapes for large k indicating a large number of homological features.

An important question to address in practical applications of multiparameter persistence landscapes is how to determine the optimal weighting $\mathbf{w}$ of the parameter space to be used in any given applica-



(a) The mean first landscape $\bar{\lambda}(1, \mathbf{x})$ of the H_1 module for the hyperbolic, Euclidean and elliptic discs taken over 100 samples.



(b) The pointwise difference between the mean landscapes $\bar{\lambda}(1, \mathbf{x})$.

Figure 2.8: Multiparameter persistence landscapes for uniform samples of discs of unit radius from spaces of different curvature.

tion. The optimal choice of weighting is likely to be highly dependent on the application in mind. The need to choose a hyperparameter is not unique to multiparameter persistence and similar challenges are presented in choosing hyperparameters for various vectorizations of single parameter persistence applications (11; 12; 34). We note that multiparameter persistence landscapes depend continuously on the choice of weighting vector $\mathbf{w}$, and so given sufficient training data one could tune this parameter to maximize discriminatory power.

The multiparameter landscapes highlight several open questions and challenges in the development of the theory and applications of multiparameter persistent homology that we would be interested to see addressed.

1. We would like to understand the relationship between the interleaving distance and landscape distance associated to modules to understand when the landscape distance provides a good lower bound estimate.

2. We have restricted our invariant to the discriminating power of the rank invariant. We would be interested to see if we could combine our landscapes with invariants that capture the more subtle relationships between features born at incomparable parameter values.

3. With respect to applications there is a question as to whether the increased complexity of multiparameter persistence is necessary, or whether single parameter persistence may be sufficient. One could possibly substitute a multiparameter persistence landscape for a vectorization of a well-chosen single parameter persistence module without significant loss of information.

4. The Bootstrap Method has been used to compute confidence bands for single parameter persistence landscapes (34). We would be interested in applying similar analysis for multiparameter landscapes.

Finally it is worth remarking that the construction of multiparameter persistence landscapes from multiparameter persistence modules can be generalized to produce stable invariants of generalized persistence modules indexed over other posets. Providing the indexing poset P is equipped with a superlinear family of translations Ω, one can derive a landscape function $\lambda : \mathbb{N} \times P \to \overline{\mathbb{R}}$ from the rank function $\mathrm{rk} : P \times P \to \mathbb{N}$. This landscape equipped with the supremum norm is stable with respect to the interleaving distance induced by the superlinear family, and provides an interpretable, stable representation of the rank function. This vectorization may prove a useful invariant should the computation of generalized persistence modules be developed in future work.

Acknowledgements

The author would like to give recognition to the *Theory and Foundations of TGDA* workshop hosted at Ohio State University which facilitated useful conversations with experts in TGDA. The author wishes to thank his supervisors Ulrike Tillmann and Vidit Nanda for their guidance and support with this project, and Peter Bubenik for helpful suggestions. The author would like to thank the anonymous reviewers for their detailed feedback on an earlier draft of this paper. The author gratefully acknowledges support from EPSRC studentship EP/N509711/1 and EPSRC grant EP/R018472/1.

Appendix

2.8 Multiparameter Persistence Theory

For completeness we present the multiparameter persistence theory, defining presentations and interleavings of multiparameter persistence modules.

2.8.1 Presentations

Definition 2.8.1 (Translation Endofunctors (28))**.** *Let* $\mathbf{P}$ *be the category associated to a preordered set (proset) and let* $\Gamma : \mathbf{P} \to \mathbf{P}$ *be an endofunctor. We say that* Γ *is a translation. Since* Γ *is a functor* Γ *is monotone* $x \leq y$ *implies that* $\Gamma(x) \leq \Gamma(y)$*. We say that* Γ *is increasing if* $x \leq \Gamma(x)$ *for all* $x \in P$*.*

Let $\mathbf{Trans_P}$ *denote the set of increasing translations of* $\mathbf{P}$ *and observe that* $\mathbf{Trans_P}$ *is a monoid with respect to composition.*

It is straight forward to see that $\mathbf{Trans_P}$ also has a natural proset structure with preorder $\Gamma \leq K \Leftrightarrow \Gamma(x) \leq K(x)$ for all x . This preorder is compatible with the monoid structure of $\mathbf{Trans_P}$ and $\Gamma \leq K$ implies there is a unique natural transformation $\eta_K^\Gamma : \Gamma \Rightarrow K$. If $\mathbf{P}$ is a poset then so is $\mathbf{Trans_P}$.

Definition 2.8.2 (Shift Functor)**.** *Let* $\mathbf{C}$ *be a category,* F *an element of the functor category* $\mathbf{C^P}$*, and* Γ *a translation endofunctor. Let* $F(\Gamma)$ *denote* $F \circ \Gamma \in \mathbf{C^P}$*, we call this functor the* Γ *shift of* F*.*

For a multiparameter persistence module $M \in \mathbf{Vect}^{\mathbf{R}^n}$ we shall write $M(\mathbf{a})$ to denote the shift by the translation in $\mathbb{R}^n$, $\Gamma_{\mathbf{a}}(\mathbf{x}) = \mathbf{x} + \mathbf{a}$. We define an $\mathbb{R}^n$-graded set to be some set $\mathcal{X}$ together with a map $\mathrm{gr} : \mathcal{X} \to \mathbb{R}^n$. For an element $j \in \mathcal{X}$, we shall refer to $\mathrm{gr}(j) = \mathbf{a} \in \mathbb{R}^n$ as the grade of j. Let P_n denote the $\mathbb{R}^n$-graded monoid ring of the monoid $([0,\infty)^n, +)$ over a field $\mathbb{F}$.

Definition 2.8.3 (Free Module)**.** *Let* $\mathcal{X}$ *be an* $\mathbb{R}^n$*-graded set. The free module on* $\mathcal{X}$ *is denoted as* $Free[\mathcal{X}]$*, and defined to be:*

$$Free[\mathcal{X}] = \bigoplus_{j \in \mathcal{X}} P_n(-gr(j)).$$

Note that each element of the graded set gives rise to an independent copy of the $\mathbb{R}^n$-graded ring P_n, and the negative shift determines that the elements of grade $\mathbf{a}$ in P_n are shifted to be elements of grade $\mathbf{a} + \mathrm{gr}(j)$ in the shifted copy $P_n(-\mathrm{gr}(j))$. The notion of a free module on a graded set can equivalently be defined using a univeral property characterisation.

We say a subset $\mathcal{R} \subset M$, of a persistence module is homogeneous if $\mathcal{R} \subset \cup_{\mathbf{a} \in \mathbb{R}^n} M_{\mathbf{a}}$, that is to say each element has a well defined grade.

Definition 2.8.4 (Presentations)**.** *Let $\mathcal{X}$ be a graded set and $\mathcal{R}$ a homogeneous subset of the free module on $\mathcal{X}$ generating the submodule $\langle\mathcal{R}\rangle$. We say that a persistence module M has presentation $\langle\mathcal{X}|\mathcal{R}\rangle$ if:*

$$M \cong \frac{Free[\mathcal{X}]}{\langle\mathcal{R}\rangle}.$$

We say that a presentation is finite if both $\mathcal{X}$ and $\mathcal{R}$ are finite. Let I denote the ideal of P_n generated by the elements $\{\mathbf{x}^{\mathbf{a}} : \mathbf{a} > 0\}$ and let $\Phi_{\langle\mathcal{X}|\mathcal{R}\rangle} : Free[\mathcal{R}] \to Free[\mathcal{X}]$ be the map induced by the inclusion $\mathcal{R} \hookrightarrow Free[\mathcal{X}]$. We say that a presentation of M is minimal if $\mathcal{R} \subset I \cdot Free[\mathcal{X}]$ and $\ker_{\Phi_{\langle\mathcal{X}|\mathcal{R}\rangle}} \subset I \cdot Free[\mathcal{R}]$.

Definition 2.8.5 (Multigraded Betti Numbers)**.** *Let M be a persistence module. The associated multigraded Betti numbers are maps $\xi_i(M) : \mathbb{R}^n \to \mathbb{N}$ defined by:*

$$\xi_i(M)(\mathbf{a}) = \dim_{\mathbb{F}}(Tor_i^{P_n}(M, P_n/IP_n)_{\mathbf{a}}).$$

Standard homological algebra arguments establish that the multigraded Betti numbers are well defined (see (23) for details). If $\langle\mathcal{X}|\mathcal{R}\rangle$ is a minimal presentation for M then $\xi_0(M)(\mathbf{a}) = |gr_{\mathcal{X}}^{-1}(\mathbf{a})|$ and $\xi_1(M)(\mathbf{a}) = |gr_{\mathcal{R}}^{-1}(\mathbf{a})|$, where $gr_{\mathcal{X}} : J \to \mathbb{R}^n$, so that $|gr_{\mathcal{X}}^{-1}(\mathbf{a})|$ gives the cardinality of the collection of elements of $\mathcal{X}$ with grading $\mathbf{a}$.

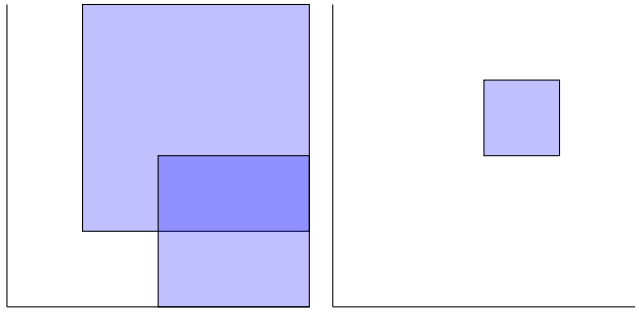
The multigraded Betti numbers are related to the initial topological space with $\xi_0(M)$ marking the filtration values for the birth of homological features, and $\xi_1(M)$ marking the filtration values for relations between features.

2.8.2 *Generalized Interleavings*

We adopt the notion of a generalized interleaving from (28) to define interleaving distances on multiparameter persistence modules.

Definition 2.8.6 (Interleaving (28))**.** *Let $\mathbf{P}$ be a proset and $\mathbf{C}$ be a category. Let $F, G \in \boldsymbol{C^P}$ be modules and $\Gamma, K \in \boldsymbol{Trans_P}$. We say that F, G are (Γ, K)-interleaved if there exist natural transformations $\varphi : F \Rightarrow G\Gamma$, $\psi : G \Rightarrow FK$ satisfying the coherence criteria that $(\psi\Gamma)\varphi = F\eta^{\mathrm{id}}_{K\Gamma}$, $(\varphi K)\psi = G\eta^{\mathrm{id}}_{\Gamma K}$ where $\eta^{\mathrm{id}}_{\alpha}$ denotes the unique natural transformation between the translations $\mathrm{id} \leq \alpha$.*

An interleaving may be thought of as an approximate isomorphism. Indeed if we take $\Gamma = K = \mathrm{id}$ then F, G are (Γ, K)-interleaved if and only if F, G are isomorphic. By warping the proset with translations Γ, K we admit flexibility to the rigid notion of isomorphism. In order to introduce an associated distance we must assign a weight to the translations to quantify how close the interleaving is to an isomorphism.



$M = [(1,1),(4,4)] \oplus [(2,0),(4,2)]$ $\quad$ $N = [(2,2),(3,3)]$

Figure 2.9: A pair of interval decomposable modules M, N with interleaving distance $d_I(M,N) = 1$.

Definition 2.8.7 (Superlinear Family (28))**.** *Let $\Omega : [0,\infty) \to \boldsymbol{Trans_P}$ be a superlinear function: $\Omega_{\varepsilon_1+\varepsilon_2} \geq \Omega_{\varepsilon_1}\Omega_{\varepsilon_2}$. We say that Ω is a superlinear family.*

Definition 2.8.8 (ε-Interleaving (28))**.** *Let $F, G \in \boldsymbol{C^P}$ be modules and Ω a superlinear family. We say F and G are ε-interleaved with respect to Ω if they are $(\Omega_\varepsilon, \Omega_\varepsilon)$-interleaved.*

Proposition 2.8.9 (Induced Interleaving Distance (28))**.** *Given a superlinear family Ω and modules $F, G \in \boldsymbol{C^P}$, we have an induced interleaving distance given by:*

$$d^{\Omega}(F,G) = \inf\{\varepsilon \geq 0 : F, G \text{ are } \varepsilon\text{-interleaved with respect to } \Omega\}.$$

The most common superlinear family to consider for persistence modules indexed by $\mathbb{R}^n$ is given by the translation in the diagonal direction, $\Omega_\varepsilon(\mathbf{x}) = \mathbf{x} + \varepsilon\mathbf{1}$. We refer to the interleaving distance induced by this superlinear family as simply *the interleaving distance* and denote this distance by d_I.

Example 2.8.10. *Let $M, N \in \mathbf{vect}^{\mathbf{R}^2}$ be interval decomposable modules with $M = [(1,1),(4,4)] \oplus [(2,0),(4,2)]$ and $N = [(2,2),(3,3)]$. There is a $1+\delta$-interleaving between M and N for all $\delta > 0$ obtained by matching the first summand of M with the summand of N. Explicitly, the natural transformation $\varphi : M \to N\Omega_{1+\delta}$ is zero everywhere except at grades $\mathbf{a} \in \mathbb{R}^2$ such that $\mathbf{a}$ lies in the support of the first summand of M and $\Omega_{1+\delta}(\mathbf{a})$ lies in the support of N, for which it is an isomorphism. The natural transformation $\psi : N \to M\Omega_{1+\delta}$ is defined similarly. There is no 1-interleaving between this pair of modules. For example, the morphism $M_{(2,0)} \to M_{(4,2)}$ is non-zero and so cannot factor through $N_{\Omega_1(2,0)} = N_{(3,1)} = 0$. Hence the interleaving distance between these modules is $d_I(M,N) = 1$.*

2.8.3 Discretization and Continuous Extension

In Section 2.6 our computations were simplified by restricting a continuous module to a finite grid and then dealing with the continuous extension of this discretization. We will show that restricting to a finite grid gives us a suitable approximation to our module with respect to the interleaving distance between modules.

Definition 2.8.11 (Grid Function)**.** *Let $\mathcal{G} : \mathbb{Z}^n \to \mathbb{R}^n$ be defined by component-wise strictly increasing functions $\mathcal{G}_i : \mathbb{Z} \to \mathbb{R}$ with $\sup \mathcal{G}_i = \sup -\mathcal{G}_i = \infty$. We say $\mathcal{G}$ is a grid function. Let us define the size of $\mathcal{G}$ to be*

$$|\mathcal{G}| = \max_{i:1\leq i\leq n} \sup_{z\in\mathbb{Z}} |\mathcal{G}_i(z) - \mathcal{G}_i(z+1)|.$$

Definition 2.8.12 (Discretization)**.** *Let $M \in \mathbf{Vect}^{\mathbf{R}^n}$ be a persistence module and $\mathcal{G}$ a grid function. We say that $M \circ \mathcal{G} \in \mathbf{Vect}^{\mathbf{Z}^n}$ is the $\mathcal{G}$-discretization of M.*

Definition 2.8.13 (Continuous Extension (23))**.** *Let $Q \in \mathbf{Vect}^{\mathbf{Z}^n}$ be a discrete persistence module and $\mathcal{G}$ a grid function. For $\mathbf{x} \in \mathbb{R}^n$ let us define floor and ceiling functions:*

$$\lfloor \mathbf{x} \rfloor_{\mathcal{G}} = \max\{\mathbf{z} \in \operatorname{Im} \mathcal{G} : \mathbf{z} \leq \mathbf{x}\} \text{ and } \lceil \mathbf{x} \rceil_{\mathcal{G}} = \min\{\mathbf{z} \in \operatorname{Im} \mathcal{G} : \mathbf{z} \geq \mathbf{x}\}.$$

We define the continuous extension $E_{\mathcal{G}}(Q) \in \mathbf{Vect}^{\mathbf{R}^n}$ to be the persistence module with:

$$E_{\mathcal{G}}(Q)_{\mathbf{a}} = Q_{\mathcal{G}^{-1}(\lfloor \mathbf{a} \rfloor_{\mathcal{G}})} \text{ and } E_{\mathcal{G}}(Q)(\mathbf{a} \leq \mathbf{b}) = Q(\mathcal{G}^{-1}(\lfloor \mathbf{a} \rfloor_{\mathcal{G}}) \leq \mathcal{G}^{-1}(\lfloor \mathbf{b} \rfloor_{\mathcal{G}}))$$

with the obvious action on the morphisms of $\mathbf{Vect}^{\mathbf{Z}^n}$. We have defined a functor $E_{\mathcal{G}} : \mathbf{Vect}^{\mathbf{Z}^n} \to \mathbf{Vect}^{\mathbf{R}^n}$.

The following proposition shows that discretization is stable with respect to the interleaving distance, and so we may produce an arbitrarily close approximation to a persistence module by restricting the module to a grid of sufficiently small size.

Proposition 2.8.14. *Let $M \in \mathbf{Vect}^{\mathbf{R}^n}$ be a persistence module and $\mathcal{G}$ a grid function, we have that:*

$$d_I(M, E_{\mathcal{G}}(M \circ \mathcal{G})) \leq |\mathcal{G}|.$$

Proof. The modules $M, E_{\mathcal{G}}(M \circ \mathcal{G})$ are $(\lceil \cdot \rceil_{\mathcal{G}}, \mathrm{id})$-interleaved with natural transformations given by the appropriate internal morphisms of M. □

2.9 Probability in Banach Spaces

We shall present the general theory of Probability for Banach space valued random variables from which we derive the convergence

results of Section 2.5. Let us begin by defining some notation. Let $(\mathcal{B}, \|\cdot\|)$ denote a real, separable Banach Space with topological dual space $\mathcal{B}^*$. Let $V : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{B}$ denote a Borel measurable random variable. The covariance structure of such a random variable is defined to be the set of expectations

$$\{\mathbb{E}[(f(V) - \mathbb{E}[f(V)])(g(V) - \mathbb{E}[g(V)])] : f, g \in \mathcal{B}^*\}.$$

In order to take expected values of Banach valued random variables we require the notion of the Pettis Integral, which is an extension of the Lebesgue integral to functions on measure spaces taking values in normed spaces. We shall briefly introduce the properties of this integral and existence criteria, the essence of which is built upon reducing the problem to integrability of $\mathbb{R}$-valued functions.

Definition 2.9.1 (Scalarly Integrable (45))**.** *A function $V : (\Omega, \mathcal{F}, \mu) \to \mathcal{B}$ is scalarly integrable if for all $f \in \mathcal{B}^*$ we have that $f(V)$ is measurable and $f(V) \in L^1(\mu)$.*

Definition 2.9.2 (Pettis Integrable (45))**.** *A scalarly integrable function $V : (\Omega, \mathcal{F}, \mu) \to \mathcal{B}$ is Pettis integrable if for all $E \in \mathcal{F}$ there is an element $I_V(E) \in \mathcal{B}$ such that:*

$$\int_E f(V) d\mu = f(I_V(E)) \text{ for all } f \in \mathcal{B}^*.$$

The set function $I_V : \mathcal{F} \to \mathcal{B}$ is called the Pettis Integral of V with respect to μ. We may also refer to $I_V(\Omega)$ as the Pettis Integral of V and denote this by I_V.

Definition 2.9.3 (Strong and Weak Measurability (46))**.** *A function $V : (\Omega, \mathcal{F}, \mu) \to \mathcal{B}$ is simple if there exist $v_1, ..., v_m \in \mathcal{B}$ and $E_1, ..., E_m \in \mathcal{F}$ such that $f = \sum_{i=1}^m v_i \mathbb{1}_{E_i}$.*

A function $V : (\Omega, \mathcal{F}, \mu) \to \mathcal{B}$ is strongly μ-measurable if there exists a sequence of simple functions $V_n : (\Omega, \mathcal{F}, \mu) \to \mathcal{B}$ such that $\lim_{n\to\infty} \|V_n(\omega) - V(\omega)\| = 0$ μ-almost everywhere.

A function $V : (\Omega, \mathcal{F}, \mu) \to \mathcal{B}$ is weakly μ-measurable if for all functionals in the dual space $v^* \in \mathcal{B}^*$ the real-valued function $v^*V : (\Omega, \mathcal{F}, \mu) \to \mathbb{R}$ is μ-measurable.

We shall suppress the prefix μ if the measure is clear from context.

Theorem 2.9.4 (Pettis Measurability Theorem (46, Theorem 3.1))**.** *A function $V : (\Omega, \mathcal{F}, \mu) \to \mathcal{B}$ is strongly measurable if and only if:*

1. *V is weakly measurable.*
2. *There exists a μ-null set E such that $V(\Omega \setminus E)$ is a separable subset of $\mathcal{B}$.*

Proposition 2.9.5 ((46, Proposition 5.1)). *A strongly measurable $\mathcal{B}$-valued function $V : (\Omega, \mathcal{F}, \mu) \to \mathcal{B}$ with $\mathbb{E}^{\mu}[\|V\|] < \infty$ is Pettis Integrable.*

A consequence of the Pettis Measurablilty Theorem is that if the codomain is a separable Banach Space then the notions of weak and strong measurability coincide. Thus Proposition 2.9.5 gives a sufficient criterion for Pettis Integrability in the setting of multiparameter persistence landscapes endowed with the p-norm for $p \in [1, \infty)$ for which the underlying Banach space is separable.

Corollary 2.9.6. *Let $V : (\Omega, \mathcal{F}, \mu) \to \mathcal{B}$ be measurable with $\mathcal{B}$ real and separable. If $\mathbb{E}^{\mu}[\|V\|] < \infty$, then V is Pettis Integrable and $\|I_V(\Omega)\| \leq \mathbb{E}^{\mu}[\|V\|]$.*

Proof. By assumption V is measurable. Moreover, $\mathbb{E}^{\mu}[\|V\|] < \infty$ implies that for any linear functional $v^* \in \mathcal{B}^*$ we have $\mathbb{E}^{\mu}[\|v^* V\|] \leq \|v^*\| \mathbb{E}^{\mu}[\|V\|] < \infty$ and so V is scalarly integrable. Hence by Proposition 2.9.5 V is Pettis Integrable. A straight forward application of the Hahn-Banach theorem establishes that $\|I_V(\Omega)\| \leq \mathbb{E}^{\mu}[\|V\|]$. □

Theorem 2.9.7 (Strong Law of Large Numbers (33, Corollary 7.10))**.** *Let V_i be i.i.d copies of $V : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{B}$ and let $S_n = \sum_{i=1}^{n} V_i$. It follows that $I_V = 0$ and $\mathbb{E}[\|V\|] < \infty$ if and only if:*

$$\frac{S_n}{n} \to 0 \text{ almost surely as } n \to \infty.$$

Corollary 2.9.8. *Let V_i be i.i.d copies of $V : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{B}$. If V has Pettis integral I_V and $\mathbb{E}[\|V\|] < \infty$ and $S_n = \sum_{i=1}^{n}(V_i)$ then:*

$$\frac{S_n}{n} \to I_V \text{ almost surely as } n \to \infty.$$

Proof. Apply Theorem 2.9.7 to the random variable $V - I_V$. □

Definition 2.9.9. *We say a $\mathcal{B}$-valued random variable X is Gaussian if for all $f \in \mathcal{B}^*$ the real valued random variable $f(X)$ is Gaussian with mean zero. Note that such a Gaussian random variable is determined by its covariance structure (33).*

The next result only applies for a certain class of Banach spaces. The type and cotype of a Banach space can be thought loosely of as a measure of how close that Banach space is to a Hilbert space, see (33) for more details. For $p \in [1, 2]$ the Lebesgue space L^p has type p and cotype 2, and for $p \in [2, \infty)$ the Lebesgue space L^p has type 2 and cotype p.

Proposition 2.9.10 ((33, Proposition 9.24))**.** *Let $\mathcal{B}$ be a Banach space. If $\mathcal{B}$ is of type 2 then every mean zero Radon random variable V with values in $\mathcal{B}$ such that $\mathbb{E}[\|V\|^2] < \infty$ is pregaussian.*

Theorem 2.9.11 (Central Limit Theorem (47))**.** *Let $\mathcal{B}$ be a Banach space of type 2 and $V : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{B}$ a Radon random variable. If $I_V = 0$, $\mathbb{E}[\|V\|^2] < \infty$ and V is pregaussian then $\frac{1}{\sqrt{n}} S_n$ converges weakly to a Gaussian random variable with the same covariance structure as V.*

Corollary 2.9.12. *Let $\mathcal{B}$ be a separable Banach space of type 2 and $V : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{B}$. If $I_V = 0$, $\mathbb{E}[\|V\|^2] < \infty$ then $\frac{1}{\sqrt{n}} S_n$ converges weakly to a Gaussian random variable with the same covariance structure as V.*

Proof. Note that since $\mathcal{B}$ is a Polish space any probability measure on $\mathcal{B}$ is tight and hence V is a Radon random variable. Moreover Proposition 2.9.10 implies that V is pregaussian and hence the hypotheses of Theorem 2.9.11 are satisfied and the result follows. □

3
Local Equivalence of Metrics for Multiparameter Persistence Modules

Contents

Abstract

An ideal invariant for multiparameter persistence would be *discriminative*, *computable* and *stable*. In this work we analyse the discriminative power of a stable, computable invariant of multiparameter persistence modules: the fibered barcode. The fibered barcode is equivalent to the rank invariant and encodes the barcodes of the 1-parameter restrictions of a multiparameter module. This invariant is well known to be globally incomplete. However in this work we show that the fibered barcode is locally complete for finitely presented modules by showing a local equivalence of metrics between the interleaving distance (which is complete on finitely-presented

modules) and the matching distance on fibered barcodes. More precisely, we show that: for a finitely-presented multiparameter module M there is a neighbourhood of M, in the interleaving distance d_I, for which the matching distance, d_0, satisfies the following bi-Lipschitz inequalities $\frac{1}{34} d_I(M,N) \leq d_0(M,N) \leq d_I(M,N)$ for all N in this neighbourhood about M. As a consequence no other module in this neighbourhood has the same fibered barcode as M.

3.1 *Introduction*

The theory and application of multiparameter persistent homology is a topic of significant interest in the field of Topological Data Analysis. Current work has studied invariants and metrics of multiparameter persistence modules, together with algorithms for their efficient computation (48; 49; 50; 51).

Compared to single parameter persistence, studying data sets filtered over multiple parameters yields a richer class of topological invariants: multiparameter persistence modules. A distinct difference between multiparameter and single parameter persistence modules is that multiparameter modules do not admit a discrete complete invariant analogous to the barcode for single parameter persistence modules (25). Another upshot of the increased complexity of multiparameter modules, is that the analogous interleaving distance is not equivalent to a matching distance as in the single parameter case and is thus harder to compute (52). An ideal invariant for multiparameter persistence would be *discriminative*, *computable* and *stable*. In earlier work studying invariants for multiparameter persistence, it is common to sacrifice the discriminating power of an invariant in order to achieve computability (48; 53; 54). Also it is common to study subclasses of persistence modules to tame the wild behaviour of arbitrary multiparameter modules. Such subclasses include: exact (55), interval decomposable (32; 50), block decomposable (56; 55; 52) and rectangular interval decomposable (57) multiparameter persistence modules. For data science applications, stability is an essential property of a multiparameter module invariant; indeed most data is susceptible to noise and so we desire that our invariants be robust to small perturbations.

A discriminative, stable metric one could choose to study multiparameter persistence modules is the interleaving distance (58). The theoretical properties of the interleaving distance are well-behaved. Lesnick showed that the interleaving distance is universal on the space of multiparameter modules over prime fields, and thus the most discriminative stable metric (58). In recent work, it has been shown that this approach presents computational challenges. Bjerke-

vik, Botnan and Kerber showed that the interleaving distance is NP-hard to compute and approximate for multiparameter modules (52). We therefore must find a compromise between discriminating power and computability of multiparameter modules invariants and metrics.

Recently a wealth of computable invariants and metrics have been proposed and studied: (59; 16; 48; 53; 60; 54; 61). Several of these proposed invariants and metrics have the same discriminating power as the fibered barcode (48; 53; 54). The fibered barcode, introduced in (16), encodes all of the single parameter modules realised as restrictions of the multiparameter module to a positively sloped line, and discards higher order interactions. The fibered barcode is well known to be an incomplete invariant, what's more, there exist modules arbitrarily distant in the interleaving distance which have the same fibered barcode. However the fibered barcode and derived invariants are computable, stable and admit a range of desired features for data analysis: L^p-norms, averaging, statistical analysis (48; 62; 54).

Our contributions

This work relates the local behaviour of the theoretically well-behaved interleaving distance d_I to the computable yet incomplete matching distance d_0 (59; 16; 49). We provide a positive result that shows a local equivalence between these two metrics.

A significant strength of our work is that the results apply to a constructible subclass of multiparameter modules, those which are finitely presented. This restriction is relevant to data analysis applications, since the sublevel-set persistence module of a finite multifiltered simplicial complex arising from finite data will give rise to such a constructible module. Hence our results apply to the multiparameter modules one is likely to encounter in applications.

The main technical result of our paper shows that the matching distance d_0 distinguishes a finitely presented ($\mathbb{R}^n$-indexed) multiparameter persistence module M from all modules N in a d_I-neighbourhood of M (Theorem 3.4.1). We show that there is a d_I-open ball, B_M, centred at M in the space of finitely presented modules such that for all $N \in B_M$:

$$\frac{1}{34} d_I(M,N) \leq d_0(M,N) \leq d_I(M,N) \tag{3.2}$$

The fraction $\frac{1}{34}$ is likely not optimal and is a consequence of the proof methodology.

The radius of the open ball for which (3.2) is valid is dependent on M. Moreover the statement is anchored about M, the inequalities of (3.2) hold in the ball B_M when the first argument of the distance

functions is fixed to be M. Indeed there exist $N, N' \in B_M$ such that $\frac{1}{34} d_I(N', N) \not\leq d_0(N', N)$.

The upper bound $d_0 \leq d_I$ is not new and is a constructed property of the matching distance d_0. The local lower bound $\frac{1}{34} d_I \leq d_0$ is proven in this work and gives rise to a global equivalence of intrinsic metrics Theorem 3.5.6.

We observe that the space of $\mathbb{R}^n$-indexed finitely presented multiparameter persistence modules is a path metric space when equipped with the interleaving distance. This property follows from the characterisation of interleavings proven by Lesnick (58). However, to the best of the author's knowledge, the fact that d_I is an intrinsic metric has not been explored nor formally stated in the multiparameter persistence literature to date.

The local result (3.2) extends to a global result for the induced path metric $\hat{d}_0$ (Definition 3.5.2). This says that $\hat{d}_0$ and d_I are bi-Lipschitz equivalent as metrics on the whole space of finitely presented multiparameter persistence modules (Theorem 3.5.6). For any pair of finitely presented multiparameter modules M, N we attain:

$$\frac{1}{34} d_I(M, N) \leq \hat{d}_0(M, N) \leq d_I(M, N) \tag{3.3}$$

Whilst d_I is an intrinsic metric (Corollary 3.5.5), the collection of geodesics between any pair of distinct modules is highly non-unique. This is a consequence of the interleaving distance behaving as an L^∞-style norm. The matching distance d_0 also behaves like an L^∞-style norm, however there are L^p-style norms defined on faithful invariants derived from the fibered barcode (48; 54). These L^p-style metrics constrain the collection of geodesics between modules. This local equivalence result is the first step towards a sensible notion of interpolation between multiparameter modules compatible with the interleaving distance.

As a simple corollary, we attain equivalences of metrics on constructible spaces which can be embedded into the space of finitely presented modules. In particular we show how interlevel set persistence modules give rise to finitely presented 2-parameter persistence modules.

Related work

A major inspiration for this work was the article of Carriere and Oudot: *Local Equivalence and Intrinsic Metrics Between Reeb Graphs* (63). In this work Carriere and Oudot conjecture that similar results to their local equivalence of metrics on Reeb graphs apply to "more general classes of metric spaces than Reeb graphs". Indeed in Section 3.6 we show that our local equivalence result for multiparameter

modules induces a similar but weaker result to the main result of (63) for Reeb graphs (induced by embedding Reeb graphs into multiparameter persistence modules). In our proof of Theorem 3.4.1 we employ similar techniques to those used in (63).

The work of Lesnick and Wright has been fundamental to the author's understanding of finitely presented multiparameter persistence modules and the interleaving distance (58; 51; 23). In particular we have used in this work that a free resolution of a multiparameter module induces a free resolution of the 1-parameter restrictions, the characterisation of interleavings (Theorem 3.5.3), and the push function of (23).

In extending our local equivalence result Theorem 3.4.1 to a global equivalence Theorem 3.5.6 we consider paths in the space of multiparameter persistence modules in order to induce intrinsic metrics. The resulting intrinsic metric is similar in nature to the Wasserstein distance in (64), which is defined via a path of simple morphisms.

Outline of Content

In Section 3.2 we introduce multiparameter persistence theory including the two metrics which we wish to compare: the interleaving distance (Definition 3.2.5) and the matching distance (Definition 3.2.8). In Section 3.3 we define Merge and Simplification Functors which we shall use to manipulate finitely presented multiparameter persistence modules and establish their effect on the presentation of a finitely presented multiparameter persistence module. Section 3.4 is the technical heart of the paper within which we prove the local equivalence of the interleaving distance and matching distance (Theorem 3.4.1). The proof of this result makes heavy use of the functors defined and established in Section 3.3. In Section 3.5 we recall the definition of intrinsic metrics. We show that the space of finitely presented multiparameter persistence modules has geodesic paths with respect to the interleaving distance, and thus our local equivalence result induces a global equivalence between the interleaving distance and the intrinsic matching distance. Finally in Section 3.6 we show that our local equivalence of metrics result for multiparameter persistence modules induces a local equivalence of metrics for Reeb graphs analogous to the result attained in (63).

3.2 *Preliminaries*

In this section we tersely introduce the key multiparameter persistence definitions we require to develop the functors in Section 3.3 and prove the main results in Section 3.4 and Section 3.5. For a more

complete introduction see (25; 23).

3.2.1 Multiparameter Persistence Modules

Notation

Let P_n denote the monoid ring of the monoid $([0,\infty)^n, +)$ over a field $\mathbb{F}$. One can think of P_n as a pseudo-polynomial ring $\mathbb{F}[x_1, ..., x_n]$ in which exponents are only required to be non-negative and can be non-integral. We shall denote the monomial $\Pi_{i=1}^n x_i^{a_i} \in P_n$ as $\mathbf{x}^{\mathbf{a}}$. Let $\mathbf{R}^n$ denotes the category associated to the poset $(\mathbb{R}^n, \leq)$ under the standard coordinate-wise partial order, and more generally let $\mathbf{P}$ denote the category associated to the poset P. Let **Vect** denote the category of vector spaces and linear maps over $\mathbb{F}$, and **vect** denote the subcategory of finite dimensional vector spaces. For a category $\mathbf{C}$ and a poset category $\mathbf{P}$ let us denote the functor category of $\mathbf{C}$-valued functors on $\mathbf{P}$ by $\mathbf{C}^{\mathbf{P}}$. For a poset P we shall use $\vee$ to denote the join of elements.

Definition 3.2.1 (Multiparameter Persistence Module)**.** *A multiparameter persistence module is an $\mathbb{R}^n$-graded P_n-module, normally denoted by M. That is to say M has a decomposition as a $\mathbb{F}$-vector space $M = \bigoplus_{\mathbf{a}\in\mathbb{R}^n} M_{\mathbf{a}}$ compatible with the action of P_n: if $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{x}^{\mathbf{b}} \in P_n$ and $m \in M_{\mathbf{a}}$ then $\mathbf{x}^{\mathbf{b}} \cdot m \in M_{\mathbf{a}+\mathbf{b}}$. A morphism of graded modules is required to respect the grading and be compatible with the module structure i.e. if $f : M \to N$, $r \in P_n$ and $m \in M_{\mathbf{a}}$ then $f(m) \in N_{\mathbf{a}}$ and $r \cdot f(m) = f(r \cdot m)$.*

For the statement of some results it is more convenient to use the following equivalent category-theoretic definition.

Definition 3.2.2 (Multiparameter Persistence Module)**.** *A multiparameter persistence module is an element of the functor category $\mathbf{Vect}^{\mathbf{R}^n}$. A morphism of multiparameter persistence modules is a natural transformation $M \Rightarrow M'$.*

The equivalence of the two perspectives is simply saying that we have an equivalence of categories between $\mathbb{R}^n$-graded P_n-**Mod** and $\mathbf{Vect}^{\mathbf{R}^n}$ (58). We shall freely switch between these equivalent perspectives throughout.

Let $T_\varepsilon : \mathbb{R}^n \to \mathbb{R}^n$ via $T_\varepsilon(\mathbf{a}) = \mathbf{a} + \varepsilon(1, ..., 1)$ and $T_\varepsilon^* : \mathbf{Vect}^{\mathbf{R}^n} \to \mathbf{Vect}^{\mathbf{R}^n}$ denote the pullback ($T_\varepsilon^*(M) = M \circ T_\varepsilon$).

Definition 3.2.3 (Internal Translation)**.** *The internal translation of a module M is the morphism denote by $\varphi_\varepsilon^M : M \to T_\varepsilon^*(M)$ given by the action of the ring element $\mathbf{x}^{\varepsilon\mathbf{1}} \in P_n$ on M, $\varphi_\varepsilon^M(m) = \mathbf{x}^{\varepsilon\mathbf{1}} \cdot m$.*

Definition 3.2.4 (Interval Decomposable Modules)**.** *Let $I \subset \mathbb{R}^n$ be a connected subposet[3] such that $\mathbf{a}, \mathbf{b} \in I$ and $\mathbf{a} \leq \mathbf{p} \leq \mathbf{b}$ implies $\mathbf{p} \in I$,*

[3] Connected as a poset so that for any $\mathbf{a}, \mathbf{b} \in I$ there is a finite sequence of comparable elements *connecting* $\mathbf{a}$ and $\mathbf{b}$. More precisely there exists $k \in \mathbb{N}$ and $\mathbf{x}_1, ..., \mathbf{x}_k \in I$ such that $\mathbf{a} \leq \mathbf{x}_1 \geq \mathbf{x}_2 \leq \mathbf{x}_3 \geq ... \leq \mathbf{x}_{k-1} \geq \mathbf{x}_k \leq \mathbf{b}$.

then we say I is an interval. Let $\mathbb{1}^I$ denote the **vect***-valued functor with domain $\mathbf{R}^n$ for which $\dim \mathbb{1}^I(\mathbf{a}) = \mathbb{1}\{\mathbf{a} \in I\}$ and such that the internal morphisms $\mathbb{1}^I(\mathbf{a} \leq \mathbf{b})$ are isomorphisms wherever possible. We say that $\mathbb{1}^I$ is an interval module, and any module M which is isomorphic to the direct sum of interval modules is interval decomposable.*

We shall regularly use interval decomposable modules in our examples and figures. However one should keep in mind that our results are not constrained to this subclass of modules.

Let us now define the interleaving distance, the most discriminative stable metric one can consider on multiparameter persistence modules alluded to in the introduction. We shall adopt the notation of (24). Let $T_\varepsilon : \mathbf{R}^n \to \mathbf{R}^n$ denote the translation endofunctor given by $T_\varepsilon(\mathbf{a}) = \mathbf{a} + \varepsilon\mathbf{1}$ where $\mathbf{1} = (1, 1, ..., 1)$. Observe that this translation induces an endofunctor $T_\varepsilon^* : \mathbf{Vect}^{\mathbf{R}^n} \to \mathbf{Vect}^{\mathbf{R}^n}$ where $T_\varepsilon^*(M) = M \circ T_\varepsilon$ with the obvious action on morphisms.

Definition 3.2.5 (Interleaving Distance)**.** *Let $M, N \in \mathbf{Vect}^{\mathbf{R}^n}$ be multiparameter persistence modules and $\varepsilon \geq 0$. We say that M, N are ε-interleaved if there exist natural transformations $f : M \Rightarrow NT_\varepsilon$, $g : N \Rightarrow MT_\varepsilon$ satisfying the coherence criteria that $T_\varepsilon^*(g)f = \varphi_{2\varepsilon}^M$, $T_\varepsilon^*(f)g = \varphi_{2\varepsilon}^N$. We define the interleaving distance to be the infimum of ε for which M, N are ε-interleaved:*

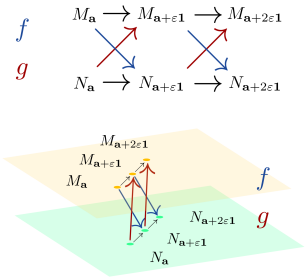
$$d_I(M, N) = \inf\{\varepsilon \geq 0 : M, N \text{ are } \varepsilon\text{-interleaved}\}$$

An interleaving may be thought of as an approximate isomorphism. Modules are 0-interleaved if and only if they are isomorphic. By *blurring* the poset with translations T_ε we admit flexibility to the rigid notion of isomorphism. The interleaving distance is universal amongst stable distances on persistence modules over prime fields, that is to say any other stable distance is bounded above by the interleaving distance (58). However, the interleaving distance has been shown to be NP-hard to compute and approximate (52).

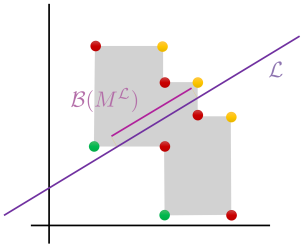
Associated to a multiparameter module is a family of single parameter modules whose collection of barcodes is known as the fibered barcode. A line in $\mathbb{R}^n$ with equation $\mathcal{L}(t) = t\mathbf{m} + \mathbf{c}$ is said to be positively sloped if each coordinate of the gradient $\mathbf{m}$ is strictly positive ($m_i > 0$ for all i).

Definition 3.2.6 (Fibered Barcode (16))**.** *Let $\mathbf{L}$ denote the subposet of $\mathbf{R}^n$ corresponding to a positively sloped line $\mathcal{L} \subset \mathbb{R}^n$. Let $\iota_\mathcal{L} : (\mathbb{R}, \|\cdot\|_\infty) \to (\mathbb{R}^n, \|\cdot\|_\infty)$ denote the order preserving isometric embedding with $\iota_\mathcal{L}(\mathbf{R}) = \mathbf{L}$ and $\iota_\mathcal{L}(0) \in \{\mathbf{x} \in \mathbb{R}^n : x_n = 0\}$. Then for $M \in \mathbf{Vect}^{\mathbf{R}^n}$ the composite $M^\mathcal{L} = M \circ \iota_\mathcal{L}$ is a single parameter persistence module, and thus has an associated barcode $\mathcal{B}(M^\mathcal{L})$. The fibered barcode of M is the collection $\{\mathcal{B}(M^\mathcal{L}) : \mathcal{L} \in \Lambda\}$ where Λ denotes the set of positively sloped lines.*

In contrast to the interleaving distance, the fibered barcode is computable for a (finitely-presented) multiparameter persistence module. The RIVET software (62) efficiently computes the fibered barcode for a 2-parameter finitely presented module, and has been used in (65; 48).



(a) A visualisation of the morphisms comprising an ε-interleaving for 2-parameter persistence modules. The interleaving morphisms $f : M \to NT_\varepsilon$ and $g : N \to MT_\varepsilon$ knit together the persistence modules M and N so that the diagram commutes.



(b) A barcode $\mathcal{B}(M^{\mathcal{L}})$ in the fibered barcode of a 2-parameter interval module M. We depict the multiparameter Betti numbers of the 2-parameter interval module, using green for generators, red for relations and amber for second order relations.

Figure 3.1: An illustration of an ε-interleaving and the fibered barcode, for 2-parameter persistence modules.

Definition 3.2.7 (Push Function (23)). *Let $\mathcal{L} : \mathbb{R} \to \mathbb{R}^n$ be a strictly positively sloped line in $\mathbb{R}^n$ isometrically embedded with respect to $\|\cdot\|_\infty$. The map $push_{\mathcal{L}} : \mathbb{R}^n \to \operatorname{Im}\mathcal{L}$ is defined as $push_{\mathcal{L}}(\mathbf{p}) = \min\{\mathbf{a} \in \operatorname{Im}\mathcal{L} : \mathbf{a} \geq \mathbf{p}\}$, and is a partial order preserving map, pushing every element of $\mathbb{R}^n$ to the line $\mathcal{L}$.*

Using the bottleneck distance for 1-parameter persistence modules on each 1-parameter restriction we induce a matching distance on the fibered barcode. The matching distance was defined for multiparameter persistence modules in (16), and the stability of the matching distance with respect to the interleaving distance explicitly shown in (59).

Definition 3.2.8 (Matching Distance (16)). *Let $M, N \in \mathbf{vect}^{\mathbf{R}^2}$ the matching distance is taken to be the supremum of the bottleneck distance over 1-dimensional restrictions:*

$$d_0(M,N) = \sup_{\mathcal{L}\in\Lambda} w(\mathcal{L}) \cdot d_B(M^{\mathcal{L}}, N^{\mathcal{L}})$$

where the weighting is $w(\mathcal{L}) := (\|push_{\mathcal{L}}(\mathcal{L}(0) + \mathbf{1}) - \mathcal{L}(0)\|_\infty)^{-1}$

An ε-interleaving of multiparameter persistence modules M, N induces an ε-interleaving between the single parameter modules along the line $\mathcal{L}$, (hence a ε-matching between their barcodes), and thus

$d_B(M^{\mathcal{L}}, N^{\mathcal{L}}) \leq d_I(M, N)$ (59; 16) (see Figure 3.2). We can generalise the matching distance defined on $\mathbf{vect}^{\mathbf{R}^2}$ to a matching distance on $\mathbf{vect}^{\mathbf{R}^n}$. That is, for $M, N \in \mathbf{vect}^{\mathbf{R}^n}$ we define the matching distance between M and N to be:

$$d_0(M, N) = \sup_{\mathcal{L} \in \Lambda} w(\mathcal{L}) \cdot d_B(M^{\mathcal{L}}, N^{\mathcal{L}})$$

where the weighting $w(\mathcal{L})$ of the line $\mathcal{L}$ is such that an ε-interleaving of the module M, N induces a $\frac{\varepsilon}{w(\mathcal{L})}$-interleaving between the modules restricted to the line $\mathcal{L}$. For example, a line $\mathcal{L}$ of slope $\mathbf{1}$ has weighting $w(\mathcal{L}) = 1$, and the weighting for the line $\mathcal{L} : x_2 = mx_1 + c \subset \mathbb{R}^2$ is given by:

$$w(\mathcal{L}) = \begin{cases} m & \text{for } m \geq 1 \\ \frac{1}{m} & \text{for } m < 1 \end{cases}$$

In the proof of our main result we only use the barcodes along slope $\mathbf{1}$ lines to bound the matching distance from below and so our weighting on the bottleneck distance is always equal to 1. Exact computation of the matching distance for bimodules has been studied and proven to be computable in polynomial time (49; 66).

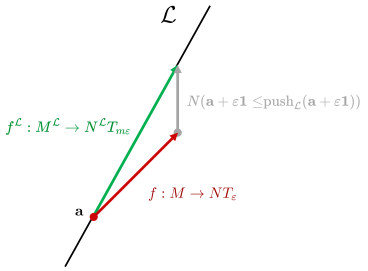


Figure 3.2: An ε-interleaving morphism $f : M \to NT_{\varepsilon}$ for 2-parameter persistence modules M and N induces an $m\varepsilon$-interleaving morphism $f^{\mathcal{L}} : M^{\mathcal{L}} \to N^{\mathcal{L}}T_{m\varepsilon}$ for the line $\mathcal{L} : x_2 = m \cdot x_1 + c$ with slope $m \geq 1$. The interleaving morphism $f^{\mathcal{L}}$ is realised as a composition of the interleaving morphism f and an internal morphism of N: $f^{\mathcal{L}}_{\mathbf{a}} = N(\mathbf{a} + \varepsilon\mathbf{1} \leq \text{push}_{\mathcal{L}}(\mathbf{a} + \varepsilon\mathbf{1})) \circ f_{\mathbf{a}}$.

3.2.2 *Presentations*

Let us now develop the theory required to define presentations of multiparameter persistence modules.

We define an $\mathbb{R}^n$-graded set to be an indexing set $\mathcal{X}$ together with a grading map $\text{gr} : \mathcal{X} \to \mathbb{R}^n$. For an element j of a graded set with $\text{gr}(j) = \mathbf{a}$, we shall refer to $\mathbf{a}$ as the grade of j. For a graded set $(\mathcal{X}, \text{gr})$ we will use $\mathcal{X}(\varepsilon)$ to denote the graded set $(\mathcal{X}, T_{-\varepsilon} \circ \text{gr})$. We will use $T^*_{-\text{gr}(j)}P_n$ to denote the the ring P_n with grading shifted by $\text{gr}(j)$ so that the multiplicative identity of $T^*_{-\text{gr}(j)}P_n$ lives at grade $\text{gr}(j)$.

Definition 3.2.9 (Free Module)**.** *The free module on $\mathcal{X}$, an $\mathbb{R}^n$-graded set, is defined to be:*

$$Free[\mathcal{X}] = \bigoplus_{j \in \mathcal{X}} T^*_{-gr(j)}P_n$$

A free module on a graded set can equivalently be defined using a universal property characterisation (25), and it is clear that a morphism from a free module is determined by its image on the generating set.

We say a subset $\mathcal{R} \subset M$ of a persistence module is homogeneous if $\mathcal{R} \subset \bigcup_{\mathbf{a} \in \mathbb{R}^n} M_{\mathbf{a}}$. That is to say each element has a well-defined grade.

Definition 3.2.10 (Presentations). *Let $\mathcal{X}$ be a graded set and $\mathcal{R}$ a homogeneous subset of the free module on $\mathcal{X}$ generating the submodule $\langle\mathcal{R}\rangle$. We say that a persistence module M has presentation $\langle\mathcal{X}|\mathcal{R}\rangle$ if:*

$$M \cong \frac{Free[\mathcal{X}]}{\langle\mathcal{R}\rangle}$$

A presentation is finite if both $\mathcal{X}$ and $\mathcal{R}$ are finite. We shall denote the subspace of finitely presented multiparameter modules as $\text{vect}_{\text{fin}}^{\mathbf{R}^n}$. Let $I_{\max}$ denote the maximal ideal of P_n generated by the elements $\{\mathbf{x}^{\mathbf{a}} \mid \mathbf{a} > 0\}$ and let $\Phi_{\langle\mathcal{X}|\mathcal{R}\rangle} : \text{Free}[\mathcal{R}] \to \text{Free}[\mathcal{X}]$ be the map induced by the inclusion $\mathcal{R} \hookrightarrow \text{Free}[\mathcal{X}]$. We say that a presentation of M is minimal if $\mathcal{R} \subset I_{\max} \cdot \text{Free}[\mathcal{X}]$ and $\ker_{\Phi_{\langle\mathcal{X}|\mathcal{R}\rangle}} \subset I_{\max} \cdot \text{Free}[\mathcal{R}]$.

More generally we say that a free resolution $(F_\bullet, \partial_\bullet)$ of M is minimal if $F_i = \text{Free}[\mathcal{X}_i]$ and $\ker(\partial_i) \subset I \cdot \text{Free}[\mathcal{X}_i]$ for all $i \in \mathbb{N}$. If M is finitely generated, then a minimal resolution exists and is unique up to isomorphism (67). A minimal resolution can be constructed sequentially by choosing a minimal set of generators to surject onto the kernel of the previous map in the resolution. Further results about minimal free resolutions can be found in (67).

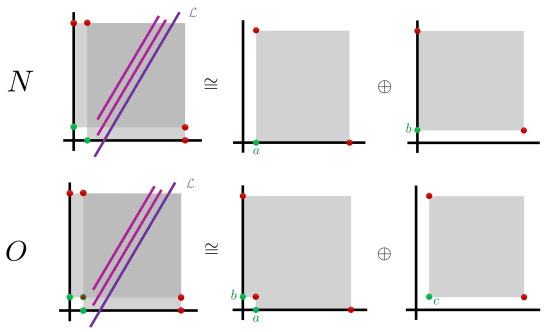


Figure 3.3: Non-isomorphic interval decomposable 2-parameter persistence modules N and O, $d_I(N,O) = \varepsilon > 0$ Example 3.2.11. The fibered barcodes of N and O are identical and thus the matching distance between N and O is zero, $d_0(N,O) = 0$.

Example 3.2.11 (Incompleteness of the Matching Distance). *Let us define the following graded sets and homogeneous subset:*

$$\begin{aligned}
\mathcal{X}_N &= \{(a,(\varepsilon,0)),(b,(0,\varepsilon))\} \\
\mathcal{X}_O &= \mathcal{X}_N \cup \{(c,(\varepsilon,\varepsilon))\} \\
\mathcal{R}_N &= \{x_1^{9\varepsilon}\cdot a, x_2^{9\varepsilon}\cdot b, x_2^{10\varepsilon}\cdot a, x_1^{10\varepsilon}\cdot b\} \subset Free[\mathcal{X}_N] \\
\mathcal{R}_O &= \mathcal{R}_N \cup \{x_1^{\varepsilon}\cdot b - x_2^{\varepsilon}\cdot a, x_1^{9\varepsilon}\cdot c, x_2^{9\varepsilon}\cdot c\} \subset Free[\mathcal{X}_O]
\end{aligned}$$

Let N, O denote the 2-parameter persistence modules with presentations $\langle\mathcal{X}_N|\mathcal{R}_N\rangle, \langle\mathcal{X}_O|\mathcal{R}_O\rangle$ respectively (see Figure 3.3). The fibered barcode of N

and O are indistinguishable, $d_0(N,O) = 0$, yet $d_I(N,O) = \varepsilon$. We also have that $d_I(N,0), d_I(O,0) \leq 5\varepsilon$. Hence for any finitely presented 2-parameter module M, $d_0(M \oplus N, M \oplus O) = 0$, $d_I(M, M \oplus N), d_I(M, M \oplus O) \leq 5\varepsilon$. Moreover since $M \oplus N$ and $M \oplus O$ are not isomorphic $d_I(M \oplus N, M \oplus O) > 0$. This example witnesses the incompleteness of the matching distance and shows that our local equivalence of metrics between d_I and d_0 (3.2) is necessarily anchored about M.

Definition 3.2.12 (Multiparameter Betti Numbers)**.** *Let M be a multiparameter persistence module. The multiparameter Betti numbers are maps $\xi_i(M) : \mathbb{R}^n \to \mathbb{N}$ defined by:*

$$\xi_i(M)(\mathbf{a}) = \dim_{\mathbb{F}}(Tor_i^{P_n}(M, P_n/IP_n)_{\mathbf{a}})$$

If M is finitely generated and $\langle \mathcal{X} | \mathcal{R} \rangle$ is a minimal presentation of M, then $\xi_0(M)(\mathbf{a}) = |\mathrm{gr}_{\mathcal{X}}^{-1}(\mathbf{a})|$ and $\xi_1(M)(\mathbf{a}) = |\mathrm{gr}_{\mathcal{R}}^{-1}(\mathbf{a})|$. More generally the support of the maps $\xi_i(M)$ are the locations of the generators of F_i in a minimal free resolution $(F_\bullet, \partial_\bullet)$ of M (up to multiplicity) (67).

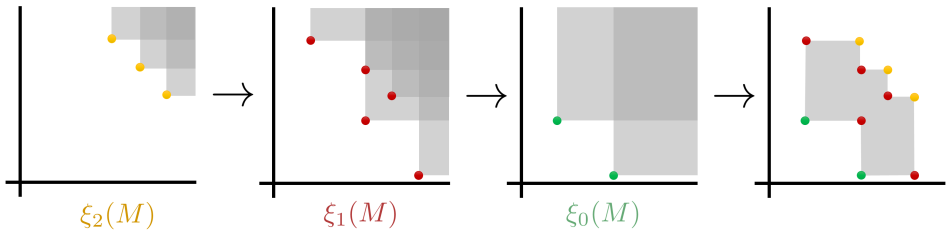


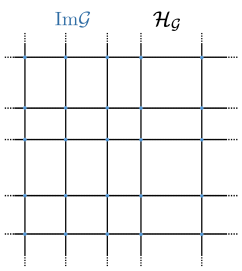
Figure 3.4: A minimal free resolution for a 2 dimensional interval module $M \in \mathbf{vect}_{\mathrm{fin}}^{\mathbf{R}^2}$.

See Figure 3.1(b) for an illustration of the multiparameter Betti numbers of a 2-parameter interval module, and Figure 3.4 for a minimal resolution realising the multiparameter Betti numbers.

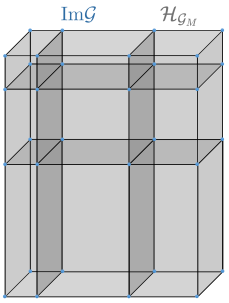
For convenience we will refer to $\xi_i(M)$ as a multiset, by which we mean the support of the map $\xi_i(M)$. That is, when we refer to a multiparameter Betti number $\mathbf{a} \in \xi_i(M)$ we mean that $\mathbf{a}$ is in the support of $\xi_i(M)$ and thus there is a generator at grade $\mathbf{a}$ in any minimal free resolution.

For a finitely presented module we can break up $\mathbb{R}^n$ into cells within which the module is constant (the internal morphisms are isomorphisms). The boundaries of these cells are determined by the positions of the multiparameter Betti numbers.

Definition 3.2.13 (Grid Function)**.** *Let $[k]$ denote the set $\{1, ..., k\}$ and $\pi_j : \mathbb{R}^n \to \mathbb{R}$ the canonical projection onto the j^{th} coordinate. A grid function is a map $\mathcal{G} = \Pi_{j=1}^n \mathcal{G}^j : \Pi_{j=1}^n [k_j] \to \mathbb{R}^n$. We shall use $\mathcal{H}_\mathcal{G}$ to*



(a) A grid function $\mathcal{G} : [k_1] \times [k_2] \to \mathbb{R}^2$ together with the associated collection of axis aligned lines $\mathcal{H}_\mathcal{G}$.



(b) A grid function $\mathcal{G} : [k_1] \times [k_2] \times [k_3] \to \mathbb{R}^3$ together with the associated collection of axis aligned hyperplanes $\mathcal{H}_\mathcal{G}$.

Figure 3.5: Example grid functions with their associated collection of axis aligned hyperplanes.

denote the collection of axis aligned hyperplanes which passes through the image of this map:

$$\mathcal{H}_\mathcal{G} := \{\mathbf{a} \in \mathbb{R}^n : \pi_j(\mathbf{a}) \in \operatorname{Im} \mathcal{G}^j \text{ for some } j \in [n]\}$$

Definition 3.2.14 (Controlling Constant)**.** *For a grid function $\mathcal{G} : \Pi_{j=1}^n [k_j] \to \mathbb{R}^n$ we define the controlling constant $c(\mathcal{G})$ to be:*

$$c(\mathcal{G}) := \min\{\|\mathbf{a} - \mathbf{b}\|_\infty : \mathbf{a} \neq \mathbf{b}, \mathbf{a}, \mathbf{b} \in \operatorname{Im} \mathcal{G}\}$$

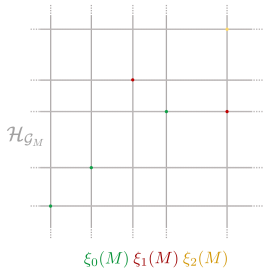
In the edge cases for which $\operatorname{Im} \mathcal{G}$ *is empty or a singleton, we define the controlling constant to be infinite.*

Definition 3.2.15 (Multiparameter Betti Grid)**.** *Suppose $M \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$ is a finitely presented multiparameter persistence module and that the multiparameter Betti numbers $\bigcup_i \xi_i(M)$ have k_j distinct gradings in the j^{th}-coordinate; (note $|\xi_0(M) \cup \xi_1(M)| < \infty$ in conjunction with Lemma 3.2.16 guarantees that k_j are indeed finite). Let $\mathcal{G}_M = \Pi_{j=1}^n \mathcal{G}_M^j : \Pi_{j=1}^n [k_j] \to \mathbb{R}^n$ be the map such that $\bigcup_i \xi_i(M) \subset \operatorname{Im} \mathcal{G}_M$. The map $\mathcal{G}_M$ shall be referred to as the multiparameter Betti grid of M and the image of this map denoted* $\operatorname{Im} \mathcal{G}_M$.

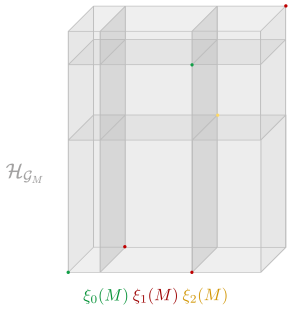
The hyperplanes $\mathcal{H}_{\mathcal{G}_M}$ mark the boundaries of the cells within which the module M does not change.

For a finitely presented module $M \in \mathbf{vect}_{\text{fin}}^{\mathbf{R}^n}$ we define the controlling constant c_M to be the controlling constant of the multiparameter Betti grid of M, $c_M = c(\mathcal{G}_M)$. This is the minimum non-zero difference between coordinates of the collection of multiparameter Betti numbers $\bigcup_i \xi_i(M)$.

Lemma 3.2.16 (Betti Grid is Determined by Generators and Relations)**.** *Let $M \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$ be a finitely presented multiparameter persistence module. For all $j \geq 1$ if $\mathbf{s} \in \xi_{j+1}(M)$ then there exist $\mathbf{r}_i \in \xi_j(M)$ such*



(a) Multiparameter Betti numbers of some 2-parameter module $M \in \mathbf{vect}^{\mathbf{R}^2}_{\text{fin}}$ together with the collection of axis aligned lines $\mathcal{H}_{\mathcal{G}_M}$.



(b) Multiparameter Betti numbers of some 3-parameter module $M \in \mathbf{vect}^{\mathbf{R}^3}_{\text{fin}}$ together with the collection of axis aligned hyperplanes $\mathcal{H}_{\mathcal{G}_M}$.

Figure 3.6: Multiparameter Betti numbers of multiparameter persistence modules together with their associated collection of axis aligned hyperplanes.

that $\mathbf{s} = \bigvee_i \mathbf{r}_i$. *Thus the multiparameter Betti grid of* M *is determined by* $\xi_0(M) \cup \xi_1(M)$.

Proof. We will just prove the result for $j = 1$, since the result for $j > 1$ follows identically. Let $F_\bullet$ be a minimal resolution of $M \in \mathbf{vect}^{\mathbf{R}^n}_{\text{fin}}$. So that in particular $F_2 \twoheadrightarrow \ker(\partial_1) = \ker(F_1 \to F_0)$. Let F_1 have generating set $\{r_i\}_{i=1}^{k_1}$ with $\text{gr}(r_i) = \mathbf{r}_i$. Suppose s is a generator of F_2 with grade $\mathbf{s}$. Since $F_\bullet$ is minimal the s has non-zero image under ∂_2. Suppose $\partial_2(s) = \sum_i \mu_i \mathbf{x}^{\mathbf{s}-\mathbf{r}_i} \cdot r_i \in \ker(\partial_1)$. Note that we must have $\mathbf{s} \geq \bigvee_{k:\mu_k \neq 0} \mathbf{r}_k$. Assume seeking contradiction that $\mathbf{s} > \bigvee_{k:\mu_k \neq 0} \mathbf{r}_k$. We observe that:

$$\begin{aligned} 0 = \partial_1(\partial_2(s)) &= \partial_1(\mathbf{x}^{\mathbf{s}-\bigvee_{k:\mu_k\neq 0}\mathbf{r}_k} \cdot \sum_i \mu_i \mathbf{x}^{\bigvee_{k:\mu_k\neq 0}\mathbf{r}_k - \mathbf{r}_i} \cdot r_i) \\ &= \mathbf{x}^{\mathbf{s}-\bigvee_{k:\mu_k\neq 0}\mathbf{r}_k} \cdot \partial_1(\sum_i \mu_i \mathbf{x}^{\bigvee_{k:\mu_k\neq 0}\mathbf{r}_k - \mathbf{r}_i} \cdot r_i) \end{aligned}$$

Since F_0 is free it follows that $\partial_1(\sum_i \mu_i \mathbf{x}^{\bigvee_{k:\mu_k\neq 0}\mathbf{r}_k - \mathbf{r}_i} \cdot r_i) = 0$ and hence (by exactness) there is some element $m \in F_2$ with $\partial_2(m) = \sum_i \mu_i \mathbf{x}^{\bigvee_{k:\mu_k\neq 0}\mathbf{r}_k - \mathbf{r}_i} \cdot r_i$. Thus $s - \mathbf{x}^{\mathbf{s}-\bigvee_{k:\mu_k\neq 0}\mathbf{r}_k} \cdot m \in \ker(\partial_2)$ and we contradict minimality of the resolution $F_\bullet$. Therefore $\mathbf{s} = \bigvee_{k:\mu_k\neq 0} \mathbf{r}_k$ and the multiparameter Betti grid of M is determined by $\xi_0(M) \cup \xi_1(M)$. □

The following Lemma 3.2.17 gives a simple criterion for identifying redundant generators in a free resolution. For convenience let us define the following notation. If $(F_\bullet, \partial_\bullet)$ is a free resolution with F_i generated by the graded set $\{b_j\}$ and F_{i+1} generated by the graded set $\{r_k\}$, let $\langle \partial_{i+1}(r_k), b_j \rangle \in \mathbb{F}$ denote the scalar such that $\partial_{i+1}(r_k) = \sum_j \langle \partial_{i+1}(r_k), b_j \rangle \mathbf{x}^{\mathbf{r}_k - \mathbf{b}_j} \cdot b_j$.

If we are working over a field for which inner products exist, $\langle \cdot, \cdot \rangle$ denotes the point-wise vector space inner product of a finitely generated free module which makes its generating set an orthonormal set.

Lemma 3.2.17 (Cancellation Lemma). *Let $M \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$ be a finitely presented multiparameter persistence module. Suppose $F_\bullet \to M$ is a free resolution such that there exists a generator b of F_i, and a generator r of F_{i+1} with $gr(b) = \mathbf{b} = \mathbf{r} = gr(r)$ and $\langle \partial_{i+1}(r), b\rangle \neq 0$, then $F_\bullet$ is not a minimal resolution, and there is a free resolution $F'_\bullet \to M$ with $\xi_0(F'_{i+1}) = \xi_0(F_{i+1}) \setminus \{\mathbf{r}\}$, $\xi_0(F'_i) = \xi_0(F_i) \setminus \{\mathbf{b}\}$ and $\xi_0(F'_j) = \xi_0(F_j)$ for all $j < i$.*

Proof. We will explicitly construct the free resolution $(F'_\bullet, d)$ satisfying $\xi_0(F'_{i+1}) = \xi_0(F_{i+1}) \setminus \{\mathbf{r}\}$, $\xi_0(F'_i) = \xi_0(F_i) \setminus \{\mathbf{b}\}$ and $\xi_0(F'_j) = \xi_0(F_j)$ for all $j < i$. For $j < i$ set $F'_j = F_j$ and $d_j = \partial_j$. Suppose F_i is generated by the graded set $\{b_j\} \cup \{b\}$ and F_{i+1} is generated by the graded set $\{r_k\} \cup \{r\}$. Set F'_i to be the submodule of F_i generated by $\{b_j\}$ and $d_i = \partial_i|_{F'_i}$. Since $\mathrm{gr}(b) = \mathbf{b} = \mathbf{r} = \mathrm{gr}(r)$:

$$\partial_{i+1}(r) = \langle \partial_{i+1}(r), b\rangle \cdot b + \sum_j \langle \partial_{i+1}(r), b_j\rangle \mathbf{x}^{\mathbf{r}-\mathbf{b}_j} \cdot b_j.$$

Moreover $0 = \partial_i \partial_{i+1}(r)$ implies

$$\begin{aligned}\partial_i(b) &= \frac{-1}{\langle \partial_{i+1}(r), b\rangle} \sum_j \langle \partial_{i+1}(r), b_j\rangle \mathbf{x}^{\mathbf{b}-\mathbf{b}_j} \cdot \partial_i(b_j) \\ &= \frac{-1}{\langle \partial_{i+1}(r), b\rangle} \sum_j \langle \partial_{i+1}(r), b_j\rangle \mathbf{x}^{\mathbf{b}-\mathbf{b}_j} \cdot d_i(b_j) \in \operatorname{Im} d_i\end{aligned}$$

and so d_i surjects onto $\ker d_{i-1} = \ker \partial_{i-1}$. Further we may define F'_{i+1} to be the submodule of F_{i+1} generated by $\{r_k\}$ and define d_{i+1} on generators of F'_{i+1} by

$$d_{i+1}(r_k) = \sum_j \langle \partial_{i+1}(r_k), b_j\rangle \mathbf{x}^{\mathbf{r}_k-\mathbf{b}_j} \cdot b_j - \frac{\langle \partial_{i+1}(r_k), b\rangle}{\langle \partial_{i+1}(r), b\rangle} \sum_j \langle \partial_{i+1}(r), b_j\rangle \mathbf{x}^{\mathbf{r}_k-\mathbf{b}_j} \cdot b_j.$$

By construction $d_i d_{i+1}(r_k) = 0$ and d_{i+1} surjects onto $\ker d_i$. We may complete the free resolution $F'_\bullet$ arbitrarily, to achieve free resolution with the desired properties. □

3.3 *Merge and Simplification Functors*

In this section we define two families of endofunctors acting on $\mathbf{vect}_{\mathrm{fin}}^{\mathbf{R}^n}$. These functors will be used in the proof of our main result Theorem 3.4.1. We shall show that these functors are well-defined and explain the effect of these functors on the presentation of a module.

Notation

Let us establish some notation conventions for this section. We shall denote free multigraded module resolutions as $(F_\bullet, \partial_\bullet)$ where ∂_i :

$F_i \to F_{i-1}$. We shall denote generators of F_0 using the letter b for "births" and we shall denote generators of F_1 using the letter r for "relations". We shall use $\mathrm{gr}(b) = \mathbf{b}$, $\mathrm{gr}(r) = \mathbf{r}$ to denote the grading of generators. Given an element $\mathbf{a} \in \mathbb{R}^n$ we will denote the slope $\mathbf{1}$ line through $\mathbf{a}$ as $\mathcal{L}_\mathbf{a}$ with $\mathcal{L}_\mathbf{a}(0) = \mathbf{a}$ and the isometric embedding with respect to $\|\cdot\|_\infty$. For a set $S \subset \mathbb{R}^n$ and $\mathbf{a} \in \mathbb{R}^n$ we use the shorthand $\|\mathbf{a} - S\|_\infty := \inf_{\mathbf{s}\in S} \|\mathbf{a} - \mathbf{s}\|_\infty$, and say $\mathbf{a}$ is δ-close to S if $\|\mathbf{a} - S\|_\infty \leq \delta$.

3.3.1 *Merge Functor*

The first family of functors we define we call merge functors. These are analogous to the merge operations defined in (68) for Reeb graphs which are heavily used in (63). Figure 3.7 gives a pictorial description of the action of a merge functor on a multiparameter module. The merge functor moves the grades of multiparameter Betti numbers lying close to a grid to lie on that grid. Moving the multiparameter Betti numbers may cause cancellations (Lemma 3.2.17).

Definition 3.3.1 (Merge Function)**.** *Let $\mathcal{G} : [k] \to \mathbb{R}$ be a finite grid with controlling constant c, and let $\delta < \frac{c}{2}$. Define the merge function $M_\delta^\mathcal{G} : \mathbb{R} \to \mathbb{R}$ as follows:*

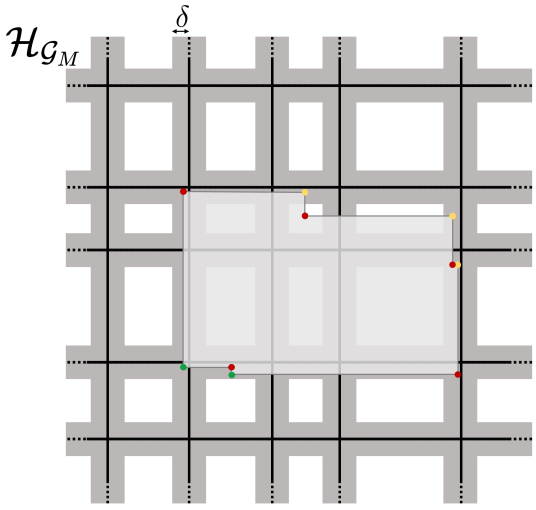
$$M_\delta^\mathcal{G}(x) = \begin{cases} \mathcal{G}(i) & \text{if } x \in [\mathcal{G}(i) - \delta, \mathcal{G}(i) + \delta] \\ x & \text{if } x \in \mathbb{R} \setminus \bigcup_i [\mathcal{G}(i) - \delta, \mathcal{G}(i) + \delta] \end{cases}$$

For a multiparameter grid function $\mathcal{G} = \mathcal{G}^1 \times ... \times \mathcal{G}^n$ with controlling constant c and $\delta < \frac{c}{2}$, the function $M_\delta^\mathcal{G} : \mathbb{R}^n \to \mathbb{R}^n$ is defined using the merge functions $M_\delta^{\mathcal{G}_i}$ coordinate-wise.

Note that the merge function is a morphism of posets and so preserves the partial order of a pair of elements in $\mathbb{R}^n$. That is to say $\mathbf{a} \leq \mathbf{b}$ implies $\mathsf{M}_\delta^\mathcal{G}(\mathbf{a}) \leq \mathsf{M}_\delta^\mathcal{G}(\mathbf{b})$. We first define the action of the merge functor on free modules.

Definition 3.3.2 (Merge Functor for Free Modules)**.** *Let $\mathcal{G} = \Pi_{i=1}^n \mathcal{G}^i : \Pi_{i=1}^n [k_i] \to \mathbb{R}^n$ be a grid with separation constant c, and let $\delta < \frac{c}{2}$. We define the δ-merge of the free module $F = Free[\mathcal{X}]$ with respect to the grid $\mathcal{G}$ to be the free module with grading set regraded via the merge function: $\mathsf{Merge}_\delta^\mathcal{G}(F) = Free[M_\delta^\mathcal{G}(\mathcal{X})]$, together with the obvious action on morphisms of free modules.*

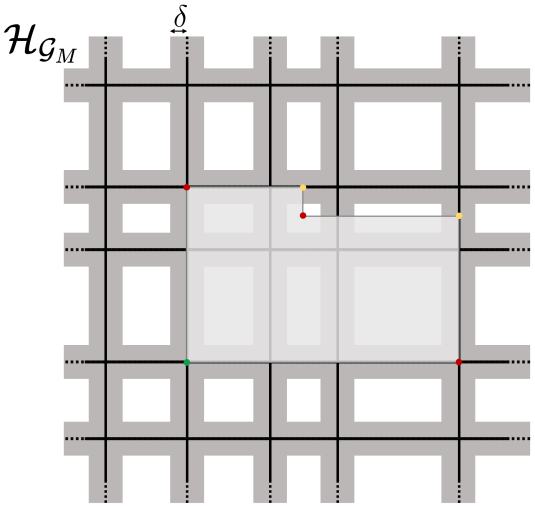
The action of $\mathsf{Merge}_\delta^\mathcal{G}$ on morphisms is well defined on free modules since the merge function is a poset endomorphism of $(\mathbb{R}^n, \leq)$. It is straight forward to check that $\mathsf{Merge}_\delta^\mathcal{G}$ is an additive and exact functor since it acts independently on the summands of free modules.



$N \in \mathbf{vect}^{\mathbf{R}^2}_{\mathrm{fin}}$

$\xi_0(N)\ \xi_1(N)\ \xi_2(N)$

(a)



$N' = \mathsf{M}_\delta^{\mathcal{G}_M}(N) \in \mathbf{vect}^{\mathbf{R}^2}_{\mathrm{fin}}$

$\xi_0(N')\ \xi_1(N')\ \xi_2(N')$

(b)

Figure 3.7: An example of a module under a Merge functor. In (a) we depict an interval module N overlayed on the δ-neighbourhood of the grid $\mathcal{H}_{\mathcal{G}_M}$, with the multiparameter Betti numbers of N marked. We observe cancellations of the multiparameter Betti numbers under the Merge functor $\mathsf{M}_\delta^{\mathcal{G}_M}$, and the multiparameter Betti numbers of N lying δ-close to the grid are merged to lie on the grid.

We extend the merge functor to arbitrary finitely presented multiparameter persistence modules.[4] Let $\mathsf{ProjResol} : \mathbf{vect}_{\text{fin}}^{\mathbf{R}^n} \to K^b(\mathbf{vect}_{\text{fin}}^{\mathbf{R}^n})$ denote the functor from finitely presented multiparameter persistence modules to the bounded homotopy category, taking a finitely presented multiparameter persistence module M to a complex of free modules $(F_\bullet, \partial_\bullet)$ with homology concentrated in degree 0 isomorphic to M, ($H_0(F_\bullet) \cong M$ and $H_i(F_\bullet) \cong 0$ for $i \neq 0$). Let $K^b(\mathsf{Merge}_\delta^{\mathcal{G}})$ denote the degree-wise application of $\mathsf{Merge}_\delta^{\mathcal{G}}$ to a complex of free modules, which is well defined since $\mathsf{Merge}_\delta^{\mathcal{G}}$ is additive and exact.

[4] The definition of the merge functor, whilst mathematically unchanged, was substantially improved in clarity based on the suggestion of an anonymous reviewer. The author would like to explicitly thank and credit the anonymous reviewer here.

Definition 3.3.3 (Merge Functor)**.** *Let $M \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$ and $\mathcal{G} = \Pi_{i=1}^n \mathcal{G}^i : \Pi_{i=1}^n [k_i] \to \mathbb{R}^n$ a grid with separation constant c, and let $\delta < \frac{c}{2}$. We define the δ-merge functor with respect to the grid $\mathcal{G}$ to be the composition:*

$$M_\delta^{\mathcal{G}} = H_0 \circ K^b(\mathit{Merge}_\delta^{\mathcal{G}}) \circ \mathit{ProjResol}$$

Note that if $F_1 \to F_0 \to M \to 0$ is a presentation of M then $\mathsf{M}_\delta^{\mathcal{G}}(M)$ has presentation: $\mathsf{Merge}_\delta^{\mathcal{G}}(F_1) \to \mathsf{Merge}_\delta^{\mathcal{G}}(F_0) \to \mathsf{M}_\delta^{\mathcal{G}}(M) \to 0$.

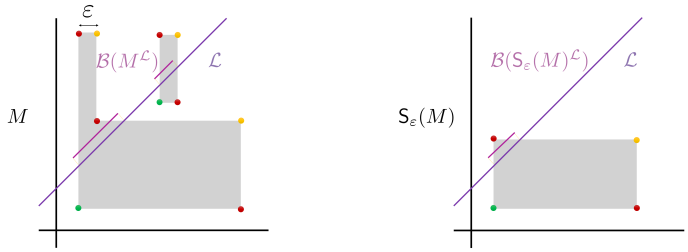
Proposition 3.3.4 ($\mathsf{M}_\delta^{\mathcal{G}}$ is Exact)**.** *The merge functor $M_\delta^{\mathcal{G}}$ is an exact functor on $\mathbf{vect}_{fin}^{\mathbf{R}^n}$.*

Proof. Let $A, B, C \in \mathbf{vect}_{\text{fin}}^{\mathbf{R}^n}$ and $0 \to A \to B \to C \to 0$ be a short exact sequence. The horseshoe lemma implies that $\mathsf{ProjResol}(0 \to A \to B \to C \to 0)$ is a short exact sequence of free resolutions. Exactness of the functor $\mathsf{Merge}_\delta^{\mathcal{G}}$ implies that $K^b(\mathsf{Merge}_\delta^{\mathcal{G}}) \circ \mathsf{ProjResol}(0 \to A \to B \to C \to 0)$ is also a short exact sequence. The homology long exact sequence applied to this short exact sequence yields the short exact sequence: $0 \to \mathsf{M}_\delta^{\mathcal{G}}(A) \to \mathsf{M}_\delta^{\mathcal{G}}(B) \to \mathsf{M}_\delta^{\mathcal{G}}(C) \to 0$ since the homology of each complex is concentrated in degree zero. □

Another important property of the family of merge endofunctors is that $\mathsf{M}_\delta^{\mathcal{G}}$ perturbs multiparameter persistence modules by no more than δ in the interleaving distance.

Proposition 3.3.5 ($d_I(\mathsf{M}_\delta^{\mathcal{G}}(M), M) \leq \delta$ (69, Lemma 5.13))**.** *For any grid $\mathcal{G}$ and $\delta < \frac{c(\mathcal{G})}{2}$, if $M \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$ then $d_I(M_\delta^{\mathcal{G}}(M), M) \leq \delta$.*

Proof. Let $(F_\bullet, \partial_\bullet)$ be a free resolution of M. Note that any free module F is δ-interleaved with $\mathsf{Merge}_\delta^{\mathcal{G}}(F)$ via morphisms which map each summand of F independently to the corresponding shifted summand of $\mathsf{Merge}_\delta^{\mathcal{G}}$ and vice versa. We may therefore construct degree-wise interleavings of $F_\bullet$ and $K^b(\mathsf{Merge}_\delta^{\mathcal{G}})(F_\bullet)$ which are chain maps, and so $F_\bullet$ and $K^b(\mathsf{Merge}_\delta^{\mathcal{G}})(F_\bullet)$ are δ-interleaved. Hence by (69, Lemma 5.13) $\mathsf{M}_\delta^{\mathcal{G}}(M) \cong H_0(K^b(\mathsf{Merge}_\delta^{\mathcal{G}})(F_\bullet))$ and $M \cong H_0(F_\bullet)$ are also δ-interleaved. □



(a) A 2-parameter interval decomposable module M.

(b) The image of the internal morphism $\operatorname{Im}\varphi_\varepsilon^M = \mathsf{S}_\varepsilon(M)$.

Figure 3.8: A 2-parameter interval decomposable module M under the ε-simplification functor S_ε (Definition 3.3.6).

3.3.2 *Simplification Functor*

The second family of functors we define simplify a module by removing "ε-small features" of the module. A pictorial description of a simplification functor applied to an interval decomposable 2-parameter persistence module is depicted in Figure 3.8.

Definition 3.3.6 (Simplification Functor)**.** *If $M \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$ define the ε-simplification of M, denoted $\mathcal{S}_\varepsilon(M)$, to be the image under the internal translation φ_ε^M of M:*

$$\mathcal{S}_\varepsilon(M) = \operatorname{Im}\varphi_\varepsilon^M.$$

Proposition 3.3.7 ($d_I(M, \mathsf{S}_\varepsilon(M)) \leq \varepsilon$)**.** *$\mathcal{S}_\varepsilon$ is an endofunctor of $\mathbf{Vect}^{\mathbf{R}^n}$ and perturbs modules by no more than ε in the interleaving distance, that is $d_I(M, \mathcal{S}_\varepsilon(M)) \leq \varepsilon$.*

Proof. We can explicitly define the interleaving morphisms realising the upper bound $d_I(M, \mathsf{S}_\varepsilon(M)) \leq \varepsilon$: $f : M \Rightarrow \mathsf{S}_\varepsilon(M) \circ T_\varepsilon$, $g : \mathsf{S}_\varepsilon(M) \Rightarrow M \circ T_\varepsilon$. Taking $f = \varphi_{2\varepsilon}^M$ and $g = \operatorname{Im}\varphi_\varepsilon^M \hookrightarrow M \circ T_\varepsilon$ we see that $T_\varepsilon^*(g) \circ f = \varphi_{2\varepsilon}^M$ and $T_\varepsilon^*(f) \circ g = \varphi_{2\varepsilon}^M \circ T_\varepsilon = \varphi_{2\varepsilon}^{\mathsf{S}_\varepsilon(M)}$. □

We explicit the effect of the simplification functor on a multiparameter presentation and on the barcode of a single parameter module.

Lemma 3.3.8 (Presentation Change under Simplification)**.** *Suppose $F_\bullet$ is a free resolution of $M \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$ and let F_1 and F_0 have generating sets $\{r_i\}$ and $\{b_j\}$ respectively. There is a corresponding free resolution $F'_\bullet$ of $\mathcal{S}_\varepsilon(M)$, such that F'_1 and F'_0 have generating sets $\{r'_i\}$ and $\{b'_j\}$ in bijective correspondence with the generating sets of F_0 and F_1, with the following relationship between the gradings:*

$$gr(b') = gr(b)$$

$$gr(r') = \left((gr(r) - \varepsilon\mathbf{1}) \vee \bigvee_{j:\lambda_j \neq 0} gr(b_j) \right) \text{ where } \partial_1(r) = \sum_j \lambda_j \mathbf{x}^{\mathbf{r}-\mathbf{b}_j} \cdot b_j$$

together with the obvious inherited differentials.

Proof. Let F_1' and F_0' have generating sets $\{r_i'\}$ and $\{b_j'\}$. Define $\partial_0'(b') = \mathbf{x}^{\varepsilon\mathbf{1}} \cdot \partial_0(b)$ so that $\partial_0' : F_0' \to \operatorname{Im} \varphi_\varepsilon^M$ is clearly surjective. For every relation r' define $\partial_1'(r') = \sum_j \lambda_j \mathbf{x}^{\mathbf{r}'-\mathbf{b}_j'} \cdot b_j'$. We have that:

$$\begin{aligned}\partial_0' \circ \partial_1'(r') &= \partial_0'\Big(\sum_j \lambda_j \mathbf{x}^{\mathbf{r}'-\mathbf{b}_j'} \cdot b_j'\Big) = \sum_j \lambda_j \mathbf{x}^{\mathbf{r}'+\varepsilon\mathbf{1}-\mathbf{b}_j'} \cdot \partial_0(b_j) \\ &= \mathbf{x}^{\mathbf{r}'-\mathbf{r}} \cdot \sum_j \lambda_j \mathbf{x}^{\mathbf{r}-\mathbf{b}_j} \cdot \partial_0(b_j) = \mathbf{x}^{\mathbf{r}'-\mathbf{r}} \cdot \partial_0 \circ \partial_1(r) = 0\end{aligned}$$

and so $\operatorname{Im} \partial_1' \subset \ker \partial_0'$. Moreover suppose that $\sum_j \lambda_j \mathbf{x}^{\mathbf{s}-\mathbf{b}_j'} \cdot b_j' \in \ker \partial_0'$ then $\sum_j \lambda_j \mathbf{x}^{\mathbf{s}+\varepsilon\mathbf{1}-\mathbf{b}_j} \cdot b_j \in \ker \partial_0 = \operatorname{Im} \partial_1$, hence there exists relations r_j and scalars μ_j such that $\partial_1(\sum_j \mu_j \mathbf{x}^{\mathbf{s}-\mathbf{r}_j} \cdot r_j) = \sum_j \lambda_j \mathbf{x}^{\mathbf{s}+\varepsilon\mathbf{1}-\mathbf{b}_j} \cdot b_j$ and thus $\partial_1'(\sum_j \mu_j \mathbf{x}^{\mathbf{s}-\mathbf{r}_j'} \cdot r_j') = \sum_j \lambda_j \mathbf{x}^{\mathbf{s}-\mathbf{b}_j'} \cdot b_j' \in \operatorname{Im} \partial_1'$. We have shown exactness of $F_1' \to F_0' \to \operatorname{Im} \varphi_\varepsilon^M \to 0$. □

Lemma 3.3.9 (Barcode Simplification (70, Proposition 5.18))**.** *Let $M \in \mathbf{vect}^{\mathbf{R}}$ have interval decomposition $\bigoplus_{j\in\mathcal{J}} \mathbb{1}^{I_j}$, where $I_j = [b_j, d_j)$. Define the ε-simplification of the intervals I_j by:*

$$\mathsf{S}_\varepsilon(I_j) = \begin{cases} [b_j, \infty) & \text{if } d_j = \infty \\ [b_j, d_j - \varepsilon) & \text{if } d_j - b_j > \varepsilon \\ \emptyset & \text{otherwise.} \end{cases}$$

The module $\mathsf{S}_\varepsilon(M) \cong \bigoplus_{j\in\mathcal{J}} \mathbb{1}^{\mathsf{S}_\varepsilon(I_j)}$.

Proof. Consider the minimal free resolution of M where F_0 is free on the graded set $\{b_j\}_{j\in\mathcal{J}}$, F_1 is free on the graded set $\{d_j \neq \infty\}_{j\in\mathcal{J}}$, and $\partial_1(d_j) = x^{d_j - b_j} \cdot b_j$. By Lemma 3.3.8, $\mathsf{S}_\varepsilon(M)$ has a corresponding resolution $F_\bullet'$. Moreover F_0' is free on the graded set $\{b_j'\}_{j\in\mathcal{J}}$ with $\operatorname{gr}(b_j') = b_j$, F_1 is free on the graded set $\{d_j'\}_{j\in\mathcal{J}}$ with $\operatorname{gr}(d_j') = b_j \vee (d_j - \varepsilon)$, and $\partial_1'(d_j') = x^{d_j' - b_j'} \cdot b_j'$. As claimed, we have $\operatorname{coker}(\partial_1') \cong \bigoplus_{j\in\mathcal{J}} \mathbb{1}^{\mathsf{S}_\varepsilon(I_j)}$. □

The simplification functor naturally restricts to the 1-parameter restrictions of a multiparameter persistence modules. For slope $\mathbf{1}$ lines $\mathcal{L}_\mathbf{a}$, the ε-simplification functor commutes with the restriction to the line $\mathcal{L}_\mathbf{a}$:

$$\mathsf{S}_\varepsilon|_{\mathcal{L}_\mathbf{a}}(M^{\mathcal{L}_\mathbf{a}}) = \mathsf{S}_\varepsilon(M)^{\mathcal{L}_\mathbf{a}} \tag{3.4}$$

Thus using Lemma 3.3.9 we can track how the simplification functor S_ε applied to a multiparameter module affects barcodes in the fibered barcode of that module.

3.4 *Local Equivalence*

In this section we prove the main result of this article Theorem 3.4.1. We make use of the merge and simplification functors defined in Section 3.3 and their effect on the presentation and fibered barcode of multiparameter persistence modules. After a series of technical lemmas we establish the following local equivalence result:

Theorem 3.4.1 (Local Equivalence)**.** *Suppose $M \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$ is a finitely presented multiparameter persistence module, M has controlling constant $c_M = c(\mathcal{G}_M)$, and $B(M, \frac{c_M}{4}) \subset (\mathbf{vect}_{fin}^{\mathbf{R}^n}, d_I)$ is the open ball centred at M of radius $\frac{c_M}{4}$. For all $N \in B(M, \frac{c_M}{4})$ we have:*

$$\frac{1}{34} d_I(M,N) \leq d_0(M,N) \leq d_I(M,N)$$

Since the interleaving distance is a metric on isomorphism classes in $\mathbf{vect}_{\text{fin}}^{\mathbf{R}^n}$, Theorem 3.4.1 implies that the fibered barcode of M is distinct from the fibered barcode of any $N \in B(M, \frac{c_M}{4}) \setminus \{M\}$, and thus is a locally complete invariant for M. Theorem 3.4.1 follows immediately from the following proposition:

Proposition 3.4.2. *Suppose $M, N \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$ are finitely presented multiparameter persistence modules, and M has controlling constant $c_M = c(\mathcal{G}_M)$. For all $\kappa \in [0, \frac{1}{34})$, if $0 < d_I(M,N) = \varepsilon < \frac{c_M}{2(34\kappa+1)}$ then the matching distance is bounded below: $d_0(M,N) > \kappa\varepsilon$.*

The proof of Proposition 3.4.2 is built using a series of auxiliary lemmas. The holistic idea for the proof of Proposition 3.4.2 follows the same structure as the proof of the local equivalence result in (63). We assume, seeking contradiction, that two modules M and N have a small matching distance in comparison to one of their controlling constants $d_0(M,N) = \kappa\varepsilon < c_M$ and in comparison to their interleaving distance $d_0(M,N) = \kappa\varepsilon < \varepsilon = d_I(M,N)$. We use the small matching distance $d_0(M,N) = \kappa\varepsilon < c_M$ to deduce that the multiparameter Betti numbers of N either lie close to the grid $\operatorname{Im}\mathcal{G}_M$ or can be cancelled with a nearby multiparameter Betti number. After applying a series of merges and simplifications we attain a nearby module $\tilde{N}$, ($d_I(N,\tilde{N}) \leq \text{const} \cdot \kappa\varepsilon$), such that the multiparameter Betti numbers of $\tilde{N}$ are contained in the grid $\operatorname{Im}\mathcal{G}_M$. Finally we note that if $\tilde{N}$ and M are δ-interleaved for some $\delta < \frac{c_M}{2}$ then Lemma 3.4.12

implies that $\tilde{N}$ and M are isomorphic. Hence if κ is too small these results violate the triangle inequality for the interleaving distance, and thus we deduce a lower bound for $d_0(M, N) = \kappa\varepsilon$.

Let us first recall the push function Definition 3.2.7. Note that the fibre of the push map $\text{push}_{\mathcal{L}}$ for an element $\mathbf{a} \in \mathcal{L}$ is the boundary of the downset of $\mathbf{a}$:

$$\text{push}_{\mathcal{L}}^{-1}(\mathbf{a}) = \partial\{\mathbf{p} \in \mathbb{R}^n : \mathbf{p} \leq \mathbf{a}\} \tag{3.5}$$

Lemma 3.4.3 (Induced 1-Parameter Resolution)**.** *Let $\mathcal{L} : \mathbb{R} \to \mathbb{R}^n$ be a positively sloped line and suppose $M \in \mathbf{Vect}^{\mathbf{R}^n}$ has a free resolution $F_\bullet \to M$. The 1-parameter restriction $M^{\mathcal{L}} := M \circ \mathcal{L} \in \mathbf{Vect}^{\mathbf{R}}$ has a corresponding free resolution $F_\bullet^{\mathcal{L}}$, where the generators of F_i are in bijection with the generators of $F_i^{\mathcal{L}}$; in particular a generator $\mathbf{a}$ of F_i corresponds to a generator $\mathbf{a}^{\mathcal{L}}$ of $F_i^{\mathcal{L}}$ with $\mathbb{R}$-grading given by $\mathcal{L}^{-1} \circ \text{push}_{\mathcal{L}}(\mathbf{a})$.*

Proof. For any $\mathbf{p} \in \mathbb{R}^n$ at least one coordinate of $\mathbf{p}$ is preserved by the push function i.e. there is some $i \in [n]$ such that $\text{push}_{\mathcal{L}}(\mathbf{p})_i = p_i$. The push function $\text{push}_{\mathcal{L}}$ collapses $\mathbb{R}^n$ to the line $\mathcal{L}$ whilst preserving the partial order. It is straightforward to check that the pullback functor $\mathcal{L}^* : \mathbf{Vect}^{\mathbf{R}^n} \to \mathbf{Vect}^{\mathbf{R}}$ preserves free modules, and preserves exactness since exactness of a sequence of persistence modules may be checked pointwise. Thus given a resolution of a multiparameter module M the push function induces a resolution of $M^{\mathcal{L}}$. □

Free Resolutions to Barcodes

Let $M \in \mathbf{vect}_{\text{fin}}^{\mathbf{R}}$ be a single parameter persistence module with finite free resolution $F_\bullet \to M$. Suppose F_0 has generating set $\{b_j\}_{j \in [i_0]}$ and F_1 has generating set $\{r_k\}_{k \in [i_1]}$ with the indexing respecting the grading ($j_1 < j_2$ implies $\text{gr}(b_{j_1}) \leq \text{gr}(b_{j_2})$). We may represent the map $\partial_1 : F_1 \to F_0$ of the free resolution using an $i_0 \times i_1$ matrix D with $D_{j,k} = \langle \partial_1(r_k), b_j \rangle$.

Algorithm 1: Graded Smith Normal Form (71)

Result: Given a matrix D representing a map between single parameter free modules $\partial_1 : F_1 \to F_0$ (where the indexing of the bases of F_0 and F_1 respects the grading), reduce D into Graded Smith Normal Form.

while *There exists an untreated row or column* **do**
  if *There exists i, j such that $D_{i,j} \neq 0$ and $D_{k,l} \neq 0$ implies $i < k$ or $j < l$* **then**
    Let i, j satisfy the if statement with j minimal and row i or column j untreated;
    Clear the non-zero entries to the right of $D_{i,j}$ in row i with elementary column operations;
    Clear the non-zero entries in column j above $D_{i,j}$ with elementary row operations;
    Mark row i and column j as treated;
  else
    break
  end
end

Using Algorithm 1 (detailed in (71)) we may perform grade-respecting column and row operations clearing non-zero entries upwards and to the right, to reduce the matrix D into graded Smith Normal Form $\tilde{D}$. In this form, each row of the matrix has at most one non-zero entry and moreover $\tilde{D}_{j,k} \neq 0$ implies that $D_{j,k'} \neq 0$ for some $k' \leq k$. We can read off the isomorphism class of $\operatorname{coker}(\partial_1)$ as:

$$\operatorname{coker}(\partial_1) \cong \bigoplus_{j \in [i_0]} \mathbb{1}^{I_j}$$

where $I_j = [\operatorname{gr}(b_j), \operatorname{gr}(r_k))$ if $\tilde{D}_{j,k} \neq 0$ and $I_j = [\operatorname{gr}(b_j), \infty)$ if the j^{th} row of $\tilde{D}$ is zero. In either case we say the bar b_j generates the bar I_j, and in the former case we say that the bar r_k kills the bar I_j. Note that the graded Smith Normal Form and the correspondence between bars and generators depends only on the choice of grade respecting indexing and the entries of the matrix D.

For $M \in \mathbf{vect}_{\text{fin}}^{\mathbf{R}^n}$ a multiparameter persistence module with free resolution $F_\bullet \to M$, recall that there is a corresponding free resolution $F_\bullet^{\mathcal{L}}$ of $M^{\mathcal{L}}$ for each positively sloped line $\mathcal{L}$. Given grade respecting indexings of the generating sets of $F_0^{\mathcal{L}}$ and $F_1^{\mathcal{L}}$ for each positively sloped lines $\mathcal{L}$, we say that b (a generator of F_0) generates the bar I_j in $M^{\mathcal{L}}$ if $b^{\mathcal{L}}$ generates I_j in $M^{\mathcal{L}}$. Similarly we say that the relation r (a generator of F_1) kills the bar I_j, if $r^{\mathcal{L}}$ kills the bar I_j in $M^{\mathcal{L}}$.

Recall the merge and simplification functors from Section 3.3. A key property of these functors is that: given a free resolution for

$M \in \mathbf{vect}_{\text{fin}}^{\mathbf{R}^n}$, we yield free resolutions of the image of M under these functors by appropriately shifting the gradings of the generators in the free modules of the free resolution of M. Moreover the grading shifts respect the partial order on $\mathbf{R}^n$. Thus any grade respecting indexing along a positively sloped line $\mathcal{L}$ remains a grade respecting indexing under these grading shifts, and thus the graded Smith Normal Form along the line $\mathcal{L}$ is preserved under these grade shiftings. Hence we can track the bar generated by a generator b as the grade of b is shifted by the merge and simplification functors and similarly track a bar killed by a relation r.

Our next lemma identifies the fibre of the push map.

Lemma 3.4.4 (Fibre of the Push Map)**.** *If* $\mathbf{a} = (a_1, ..., a_n)$, $\mathbf{b} = (b_1, ..., b_n)$ *are such that* $\|push_{\mathcal{L}_\mathbf{b}}(\mathbf{a}) - \mathbf{b}\|_\infty \leq \delta$ *then* $\min_{i\in[n]} |a_i - b_i| \leq \delta$*. In particular, if* $\mathbf{a} \in \operatorname{Im} \mathcal{G}$ *then* $\mathbf{b}$ *is* δ*-close to* $\mathcal{H}_\mathcal{G}$*.*

Proof. At least one of the coordinates a_i is preserved by the push map. □

Suppose modules M and N satisfy the assumptions of Proposition 3.4.2. Lemma 3.4.4 will be used to identify that the multiparameter Betti numbers inducing births of long bars in the fibered barcode of N must lie close to the grid of M.

In Lemma 3.4.9 we use the positions of births of bars in the fibered barcode of M to limit the positions of generators of N. Lemma 3.4.5 establishes that every non trivial generator of N generates some non-empty bar in the fibered barcode of N.

Lemma 3.4.5 (Non-trivial Generators Generate Non-zero Bars)**.** *Let* $M \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$ *be a finitely presented multiparameter persistence module with minimal free resolution* $(F_\bullet, \partial_\bullet)$*. If* b *is a generator of* F_0 *then for any choice of grade respecting indexing of generating sets,* $b^{\mathcal{L}_\mathfrak{b}}$ *generates a non-trivial bar in* $M^{\mathcal{L}_\mathfrak{b}}$*.*

Proof. Suppose that M has minimal free resolution $(F_\bullet, \partial_\bullet)$ the generators of $F_0^{\mathcal{L}_\mathfrak{b}}$ and $F_1^{\mathcal{L}_\mathfrak{b}}$ are given any grade respecting indexing, and $\tilde{D}$ is the graded Smith Normal Form of the matrix representing the map $\partial_1^{\mathcal{L}_\mathfrak{b}}$. If $b^{\mathcal{L}_\mathfrak{b}}$ generates a trivial bar and is assigned index j then $\tilde{D}_{j,k} \neq 0$ for some index k for which $\operatorname{gr}(r_k^{\mathcal{L}_\mathfrak{b}}) = \operatorname{gr}(b^{\mathcal{L}_\mathfrak{b}})$. $\tilde{D}_{j,k} \neq 0$ implies that $D_{j,k'} = \langle \partial_1(r_{k'}^{\mathcal{L}_\mathfrak{b}}), b^{\mathcal{L}_\mathfrak{b}} \rangle \neq 0$ for some $k' < k$, and 3.4.3 implies $\langle \partial_1(r_{k'}), b \rangle \neq 0$. Since the indexing respects the grading $\operatorname{gr}(r_{k'}^{\mathcal{L}_\mathfrak{b}}) \leq \operatorname{gr}(r_k^{\mathcal{L}_\mathfrak{b}}) = \operatorname{gr}(b^{\mathcal{L}_\mathfrak{b}})$. The non-zero coefficient $\langle \partial_1(r_{k'}), b \rangle \neq 0$ implies that $\operatorname{gr}(r_{k'}) \geq \operatorname{gr}(b)$ and in combination: $\operatorname{gr}(r_{k'}) \geq \operatorname{gr}(b)$, $\operatorname{gr}(r_{k'}^{\mathcal{L}_\mathfrak{b}}) \leq \operatorname{gr}(b^{\mathcal{L}_\mathfrak{b}})$ and Equation (3.5) imply $\operatorname{gr}(r_{k'}) = \operatorname{gr}(b)$. Hence $r_{k'}$ and b satisfy the hypotheses of Lemma 3.2.17, contradicting the minimality of $(F_\bullet, \partial_\bullet)$. □

Similarly, in Lemma 3.4.10 we use the positions of deaths of bars in the fibered barcode of M to determine the positions of relations of N. Lemma 3.4.6 establishes that every non-trivial relation of N kills some (possibly trivial) bar in the fibered barcode of N.

Lemma 3.4.6 (Non-trivial Relations Kill). *Let $M \in \mathbf{vect}_{fin}^{\mathbb{R}^n}$ be a finitely presented multiparameter persistence module with minimal free resolution $(F_\bullet, \partial_\bullet)$. If r is a generator of F_1 then for any choice of grade respecting indexing of generating sets, r kills a (possibly trivial) bar in $M^{\mathcal{L}_r}$.*

Proof. Suppose that M has minimal free resolution $(F_\bullet, \partial_\bullet)$, the generators of $F_0^{\mathcal{L}_r}$ and $F_1^{\mathcal{L}_r}$ are given any grade respecting indexing, and $\tilde{D}$ is the graded Smith Normal Form of the matrix representing the map $\partial_1^{\mathcal{L}_r}$. Suppose $r^{\mathcal{L}_r}$ is given index k. If r does not kill a bar then the k^{th} column of $\tilde{D}$ is zero. Hence $\partial_1(r) \in \operatorname{Im}\langle \partial_1(r_k) \rangle_{k'<k}$ and $\ker \partial_1 \not\subset I_{\max} \cdot F_1$ contradicting minimality of the free resolution. □

Through the proof of Lemma 3.4.6, we observe that if a relation r does not kill a bar in the barcode along the line $\mathcal{L}$ then there must be a second order relation s for which $\operatorname{gr}(s^{\mathcal{L}}) = \operatorname{gr}(r^{\mathcal{L}})$ and $\langle \partial_2(s), r \rangle \neq 0$. This fact will be used in the proof of Lemma 3.4.11.

Definition 3.4.7 (Unmerge Function). *Let $\mathcal{G} = \Pi \mathcal{G}^i$ be a grid, $\delta < \frac{c(\mathcal{G})}{2}$, and let $\mathbf{e}_i$ denote the standard basis vector of $\mathbb{R}^n$. The unmerge function $U_\delta^{\mathcal{G}}$ maps points lying on $\mathcal{H}_{\mathcal{G}}$ to lie δ far away from the grid: $U_\delta^{\mathcal{G}} : \mathcal{H}_{\mathcal{G}} \to \mathbb{R}^n$. For $\mathbf{a} \in \mathcal{H}_{\mathcal{G}}$ let $\mathcal{I}_{\mathbf{a}}$ be the indices of the coordinates of $\mathbf{a}$ such that $|a_i - \operatorname{Im} \mathcal{G}^i| \leq \delta$ and let $\mathbf{e}_i$ denote the i^{th} standard basis vector. We define the δ-unmerge of $\mathbf{a}$ as follows:*

$$U_\delta^{\mathcal{G}}(\mathbf{a}) = M_\delta^{\mathcal{G}}(\mathbf{a}) + \delta \sum_{i \in \mathcal{I}_{\mathbf{a}}} \mathbf{e}_i$$

By construction the unmerge function takes $\mathbf{a}$ and returns the maximum element in the poset $\mathbb{R}^n$ which merges to same point as $\mathbf{a}$, that is:

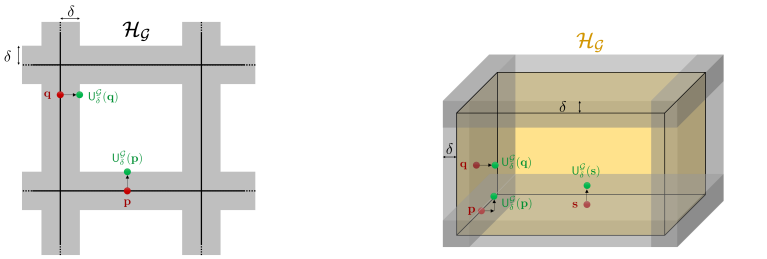
$$\mathsf{U}_\delta^{\mathcal{G}}(\mathbf{a}) = \max\{\mathbf{p} \in \mathbb{R}^n : \mathsf{M}_\delta^{\mathcal{G}}(\mathbf{p}) = \mathsf{M}_\delta^{\mathcal{G}}(\mathbf{a})\}$$

Thus for all $\mathbf{a} \in \mathcal{H}_{\mathcal{G}}$ we have $\|\mathsf{U}_\delta^{\mathcal{G}}(\mathbf{a}) - \mathcal{H}_{\mathcal{G}}\|_\infty = \delta$. Another property of the unmerge function is that if a coordinate of $\mathbf{a} \in \mathbb{R}^n$ is preserved by the merge $\mathsf{M}_\delta^{\mathcal{G}}$ (say the i^{th}-coordinate), then:

$$\operatorname{push}_{\mathcal{L}_{\mathsf{U}_\delta^{\mathcal{G}} \mathsf{M}_\delta^{\mathcal{G}}(\mathbf{a})}}(\mathbf{a}) = \mathsf{U}_\delta^{\mathcal{G}} \mathsf{M}_\delta^{\mathcal{G}}(\mathbf{a}) \tag{3.6}$$

This follows since $\mathsf{U}_\delta^{\mathcal{G}} \mathsf{M}_\delta^{\mathcal{G}}(\mathbf{a}) \geq \mathbf{a}$, and $\mathsf{U}_\delta^{\mathcal{G}} \mathsf{M}_\delta^{\mathcal{G}}(\mathbf{a})$ and $\mathbf{a}$ share the same i^{th}-coordinate. Thus $\mathsf{U}_\delta^{\mathcal{G}} \mathsf{M}_\delta^{\mathcal{G}}(\mathbf{a})$ the minimal element of and $\mathcal{L}_{\mathsf{U}_\delta^{\mathcal{G}} \mathsf{M}_\delta^{\mathcal{G}}(\mathbf{a})}$ greater than $\mathbf{a}$. See Figure 3.9 for examples of the unmerge function.

The following result, Lemma 3.4.8, will be used to show that the grades of multiparameter Betti numbers of N merge to the same



(a) $\mathsf{U}^\mathcal{G}_\delta$ applied to points $\mathbf{p}, \mathbf{q} \in \mathbb{R}^2$.

(b) $\mathsf{U}^\mathcal{G}_\delta$ applied to points $\mathbf{p}, \mathbf{q}, \mathbf{s} \in \mathbb{R}^3$.

Figure 3.9: Examples of the unmerge function $\mathsf{U}^\mathcal{G}_\delta$ moving points from $\mathcal{H}_\mathcal{G}$ to lie δ-far away from $\mathcal{H}_\mathcal{G}$.

grade and satisfy the hypotheses of Lemma 3.2.17 and thus cancel each other out.

Lemma 3.4.8 (Sufficient Conditions for Grades to Merge)**.** *Let* $\mathbf{a}, \mathbf{b} \in (\mathbf{R}^n, \leq)$ *and* $\mathcal{G}$ *be a grid function such that:* $\|\mathbf{a} - \mathcal{H}_\mathcal{G}\|_\infty \leq \delta$ *and* $\mathbf{b} \geq \mathbf{a}$*. If* $push_{\mathcal{L}_{\mathsf{U}^\mathcal{G}_\delta \mathsf{M}^\mathcal{G}_\delta(\mathbf{a})}}(\mathbf{a}) = push_{\mathcal{L}_{\mathsf{U}^\mathcal{G}_\delta \mathsf{M}^\mathcal{G}_\delta(\mathbf{a})}}(\mathbf{b})$*, then* $\mathsf{M}^\mathcal{G}_\delta(\mathbf{a}) = \mathsf{M}^\mathcal{G}_\delta(\mathbf{b})$*.*

Proof. Note that by construction of the unmerge function $\mathsf{U}^\mathcal{G}_\delta \mathsf{M}^\mathcal{G}_\delta(\mathbf{a}) \geq \mathbf{a}$ and $\mathrm{push}_{\mathcal{L}_{\mathsf{U}^\mathcal{G}_\delta \mathsf{M}^\mathcal{G}_\delta(\mathbf{a})}}(\mathbf{a}) = \mathsf{U}^\mathcal{G}_\delta \mathsf{M}^\mathcal{G}_\delta(\mathbf{a})$. The assumption that $\mathrm{push}_{\mathcal{L}_{\mathsf{U}^\mathcal{G}_\delta \mathsf{M}^\mathcal{G}_\delta(\mathbf{a})}}(\mathbf{b}) = \mathrm{push}_{\mathcal{L}_{\mathsf{U}^\mathcal{G}_\delta \mathsf{M}^\mathcal{G}_\delta(\mathbf{a})}}(\mathbf{a})$ thus implies $\mathbf{b} \leq \mathsf{U}^\mathcal{G}_\delta \mathsf{M}^\mathcal{G}_\delta(\mathbf{a})$. In combination with the assumption $\mathbf{b} \geq \mathbf{a}$, we have $\mathbf{a} \leq \mathbf{b} \leq \mathsf{U}^\mathcal{G}_\delta \mathsf{M}^\mathcal{G}_\delta(\mathbf{a})$. By definition of the unmerge function, the coordinate values of $\mathbf{a}$ which are unchanged by $\mathsf{M}^\mathcal{G}_\delta$ are also unchanged when applying $\mathsf{U}^\mathcal{G}_\delta$ to $\mathsf{M}^\mathcal{G}_\delta(\mathbf{a})$. Thus $\mathbf{a}$ and $\mathbf{b}$ share coordinate values for all the coordinate values of $\mathbf{a}$ which are unchanged by $\mathsf{M}^\mathcal{G}_\delta$. Since $\|\mathsf{U}^\mathcal{G}_\delta \mathsf{M}^\mathcal{G}_\delta(\mathbf{a}) - \mathcal{H}_\mathcal{G}\|_\infty = \delta$, then $\|\mathbf{b} - \mathcal{H}_\mathcal{G}\|_\infty \leq \delta$, and so the remaining coordinates of $\mathbf{b}$ are merged to the same coordinate values of $\mathsf{M}^\mathcal{G}_\delta(\mathbf{a})$, thus it follows that $\mathsf{M}^\mathcal{G}_\delta(\mathbf{a}) = \mathsf{M}^\mathcal{G}_\delta(\mathbf{b})$. □

In the following Lemma we assume that finitely presented modules M, N are closer in the matching distance than half the controlling constant of M. For each minimal generator b of N we consider the bars generated by b in the slope $\mathbf{1}$ lines nearby to $\mathrm{gr}(b) = \mathbf{b}$, and deduce that each generator either lies close to $\mathrm{Im}\,\mathcal{G}_M$ or generates a small bar. See Figure 3.10(a) for an illustration.

Lemma 3.4.9 (Properties of Generators)**.** *Suppose that* $M, N \in \mathbf{vect}^{\mathbf{R}^n}_{fin}$*,* $d_0(M, N) < \kappa\varepsilon$ *and* $c_M > 2\kappa\varepsilon$*. Let* $(F_\bullet, \partial_\bullet)$ *be a minimal free resolution of* N*. For any generator* b *of* F_0 *with corresponding grade* $gr(b) = \mathbf{b} \in \xi_0(N)$*, at least one of the following properties holds:*

B.1 $\mathbf{b}$ *lies* $2\kappa\varepsilon$*-close to* $\mathrm{Im}\,\mathcal{G}_M$

B.2 $\mathbf{b}$ *lies* $\kappa\varepsilon$*-close to* $\mathcal{H}_{\mathcal{G}_M}$ *and* b *generates a bar of length less than* $2\kappa\varepsilon$ *in the module* $N^{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon} \mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}}$

B.3 b generates a bar of length less than $2\kappa\varepsilon$ in the module $N^{\mathcal{L}_{\mathbf{b}}}$

Proof. Fix any generator b and consider the module $N^{\mathcal{L}_{\mathbf{b}}}$. Since b is a non-trivial generator, b generates a non-trivial bar in $N^{\mathcal{L}_{\mathbf{b}}}$ (Lemma 3.4.5). Suppose the bar in $N^{\mathcal{L}_{\mathbf{b}}}$ generated by b has length $< 2\kappa\varepsilon$, then we satisfy property B.3.

Else the bar generated by b in $N^{\mathcal{L}_{\mathbf{b}}}$ has length $\geq 2\kappa\varepsilon$ and must be $\kappa\varepsilon$-matched with a bar in the module $M^{\mathcal{L}_{\mathbf{b}}}$ since we assumed that $d_0(M,N) < \kappa\varepsilon$. Hence by Lemma 3.4.4, $\mathbf{b}$ is $\kappa\varepsilon$-close to $\mathcal{H}_{\mathcal{G}_M}$. If $\mathbf{b}$ is $2\kappa\varepsilon$-close to $\operatorname{Im}\mathcal{G}_M$ we satisfy B.1, else $\mathbf{b}$ is $> 2\kappa\varepsilon$ from all grid points of M.

In the latter case, $\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})$ lies on $\mathcal{H}_{\mathcal{G}_M}$ but is not in the image of $\mathcal{G}_M$. Using the fact that $c_M > 2\kappa\varepsilon$ and at least one of the coordinates of $\mathbf{b}$ is unchanged by $\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}$, we have that the $2\kappa\varepsilon$-close line $\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}$ is such that:

$$\|\operatorname{push}_{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}}(\mathbf{b}) - \operatorname{push}_{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}}(\operatorname{Im}\mathcal{G}_M)\|_\infty \geq \kappa\varepsilon.$$

To see this claim is true, note that Equation (3.6) implies that $\operatorname{push}_{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}}(\mathbf{b}) = \mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})$, $\operatorname{push}_{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}}(\operatorname{Im}\mathcal{G}_M) \subset \mathcal{H}_{\mathcal{G}_M}$ since the push map preserves at least one coordinate, and by construction $\|\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b}) - \mathcal{H}_{\mathcal{G}_M}\| = \kappa\varepsilon$.

Since $\operatorname{push}_{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}}(\mathbf{b})$ is $\kappa\varepsilon$ far from all points of $\operatorname{push}_{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}}(\operatorname{Im}\mathcal{G}_M)$ and $d_0(M,N) < \kappa\varepsilon$, then the bar generated by b in $N^{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}}$ cannot be matched to bars in $M^{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}}$ and so must be length $< 2\kappa\varepsilon$. □

The next Lemma assumes that a module $\hat{N}$ is close to M in the matching distance, and the generators of $\hat{N}$ lie on the grid points $\operatorname{Im}\mathcal{G}_M$. For each minimal relation r of $\hat{N}$ we consider the bars killed by r in the slope $\mathbf{1}$ lines nearby to $\operatorname{gr}(r) = \mathbf{r}$. We deduce that each relation lies close to $\mathcal{H}_M$, and if a relation is not close to $\operatorname{Im}\mathcal{G}_M$ then there is a nearby line such that if the relation kills a bar in the module along this line, then it is a small bar. See Figure 3.10(b) for an illustration.

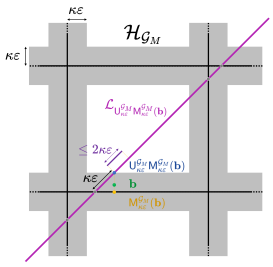
Lemma 3.4.10 (Properties of Relations). *Suppose $M, \hat{N} \in \mathbf{vect}^{\mathbb{R}^n}_{fin}$ are such that $d_0(M,\hat{N}) < \kappa\varepsilon$, $\xi_0(\hat{N}) \subset \operatorname{Im}\mathcal{G}_M$ and $c_M > 8\kappa\varepsilon$. Let $(\hat{F}_\bullet, \hat{\partial}_\bullet)$ be a minimal free resolution for $\hat{N}$ and $\mathcal{S}_{2\kappa\varepsilon}(r)$, $\mathcal{S}_{2\kappa\varepsilon}(\mathbf{r})$ denote the relation and grade of the relation corresponding to r in the free resolution of $\mathcal{S}_{2\kappa\varepsilon}(\hat{N})$ induced by $(\hat{F}_\bullet, \hat{\partial}_\bullet)$ as outlined in Lemma 3.3.8. For any relation r, (a generator of $\hat{F}_1$), with corresponding grade $gr(r) = \mathbf{r} \in \xi_1(\hat{N})$, one of the following properties holds:*

R.1 $\mathbf{r}$ *is* $2\kappa\varepsilon$*-close to* $\operatorname{Im}\mathcal{G}_M$ *or* $\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r})$ *is* $4\kappa\varepsilon$*-close to* $\operatorname{Im}\mathcal{G}_M$

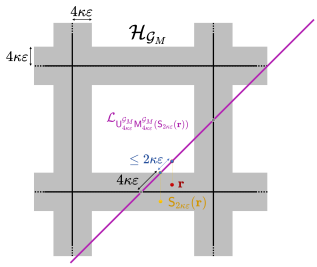
R.2 $\mathbf{r}$ *is* $2\kappa\varepsilon$*-close to* $\mathcal{H}_{\mathcal{G}_M}$ *and if* r *kills a bar in the module* $\hat{N}^{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{4\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{4\kappa\varepsilon}(\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r}))}}}$ *that bar is of length less than* $2\kappa\varepsilon$

Proof. Fix any relation r and consider the module $\hat{N}^{\mathcal{L}_\mathbf{r}}$. Since r is a non-trivial relation, there is a (possibly empty) bar killed by r in the module $\hat{N}^{\mathcal{L}_\mathbf{r}}$ (Lemma 3.4.6). If the bar killed by r has length $> 2\kappa\varepsilon$, as $d_0(M,\hat{N}) < \kappa\varepsilon$, the bar killed by r is $\kappa\varepsilon$-matched to a bar of $M^{\mathcal{L}_\mathbf{r}}$ and thus r is $\kappa\varepsilon$-close to $\mathcal{H}_{\mathcal{G}_M}$ (by Lemma 3.4.4). Else the bar is $2\kappa\varepsilon$-trivial and unmatched, in which case it is $2\kappa\varepsilon$-close to a coordinate of a generator of $\hat{N}$. By assumption $\xi_0(\hat{N}) \subset \operatorname{Im}\mathcal{G}_M$ and so $\mathbf{r}$ is $2\kappa\varepsilon$-close to $\mathcal{H}_{\mathcal{G}_M}$. In either case $\mathbf{r}$ is $2\kappa\varepsilon$-close to $\mathcal{H}_{\mathcal{G}_M}$.

If $\|\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r}) - \operatorname{Im}\mathcal{G}_M\|_\infty > 4\kappa\varepsilon$ then at least one coordinate of $\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r})$ is unchanged by $\mathsf{M}^{\mathcal{G}_M}_{4\kappa\varepsilon}$. Consider the line $\mathcal{L}' = \mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{4\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{4\kappa\varepsilon}(\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r}))}$ and note that by Equation (3.6) $\operatorname{push}_{\mathcal{L}'}(\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r})) = \mathsf{U}^{\mathcal{G}_M}_{4\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{4\kappa\varepsilon}(\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r}))$. Moreover $\|\mathsf{U}^{\mathcal{G}_M}_{4\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{4\kappa\varepsilon}(\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r})) - \operatorname{push}_{\mathcal{L}'}(\operatorname{Im}\mathcal{G}_M)\|_\infty \geq 4\kappa\varepsilon$ since $\mathsf{U}^{\mathcal{G}_M}_{4\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{4\kappa\varepsilon}(\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r}))$ is at least $4\kappa\varepsilon$ far in each coordinate from every grid point of M. Furthermore $\|\mathbf{r} - \mathsf{S}_{2\kappa\varepsilon}(\mathbf{r})\|_\infty \leq 2\kappa\varepsilon$ implies that $\|\operatorname{push}_{\mathcal{L}'}(\mathbf{r}) - \operatorname{push}_{\mathcal{L}'}(\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r}))\|_\infty \leq 2\kappa\varepsilon$ and so: $\|\operatorname{push}_{\mathcal{L}'}(\mathbf{r}) - \operatorname{push}_{\mathcal{L}'}(\operatorname{Im}\mathcal{G}_M)\|_\infty \geq 2\kappa\varepsilon$. Hence any bar killed by r in the module $\hat{N}^{\mathcal{L}'}$ is unmatched and so of length less than $2\kappa\varepsilon$ as $d_0(M,\hat{N}) < \kappa\varepsilon$. □



(a) A generator at grade $\mathbf{b}$ which is $\kappa\varepsilon$ far from $\operatorname{Im}\mathcal{G}_M$ but $\kappa\varepsilon$ close to $\mathcal{H}_{\mathcal{G}_M}$ must generate an unmatched bar in the nearby line $\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{4\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{4\kappa\varepsilon}(\mathbf{b})}$ of length $\leq 2\kappa\varepsilon$.



(b) A relation at grade $\mathbf{r}$ which is such that $\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r})$ is not $4\kappa\varepsilon$-close to $\operatorname{Im}\mathcal{G}_M$ is $2\kappa\varepsilon$ far from all points of $\operatorname{Im}\mathcal{G}_M$ when pushed to the line $\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{4\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{4\kappa\varepsilon}(\mathsf{S}_{2\kappa\varepsilon}(\mathbf{r}))}$, and so if it kills a bar in this line it must be unmatched and of length $\leq 2\kappa\varepsilon$.

Figure 3.10: Illustrations to support Lemma 3.4.9 and Lemma 3.4.10.

Using information about the location of the generators and relations of N together with which generators and relations birth or kill small bars, we find a nearby module $\tilde{N}$ whose multiparameter Betti Numbers all lie on the grid points of M.

Lemma 3.4.11 ($\bigcup_i \xi_i(\tilde{N}) \subset \operatorname{Im} \mathcal{G}_M$)**.** *Suppose $M, N \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$, $d_0(M,N) < \kappa\varepsilon$, $d_I(M,N) = \varepsilon$ and $c_M > 80\kappa\varepsilon$. There exists a module $\tilde{N}$ such that $\bigcup_i \xi_i(\tilde{N}) \subset \operatorname{Im} \mathcal{G}_M$, and $d_I(N,\tilde{N}) \leq 34\kappa\varepsilon$.*

Proof. We first give an overview of the strategy of the proof. The nearby module $\tilde{N}$ is attained by a series of merges and simplifications applied to N. We first use Lemma 3.4.9 to perturb N to a module $\hat{N}$ whose generators all lie on the grid points of M. Next we use Lemma 3.4.10 to perturb $\hat{N}$ to a module $\tilde{N}$ whose generators and relations all lie on the grid points of M. Finally, Lemma 3.2.16 implies that all the multiparameter Betti numbers of $\tilde{N}$ must now lie on the grid points of M.

Let $(F_\bullet, \partial_\bullet)$ be a minimal free resolution of N. Lemma 3.4.9 establishes that for any generator b of F_0 with grade $\mathbf{b} \in \xi_0(N)$ at least one of the following holds:

B.1 $\mathbf{b}$ lies $2\kappa\varepsilon$-close to $\operatorname{Im} \mathcal{G}_M$

B.2 $\mathbf{b}$ lies $\kappa\varepsilon$-close to $\mathcal{H}_{\mathcal{G}_M}$ and b generates a bar of length less than $2\kappa\varepsilon$ in the module $N^{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon} \mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b})}}$

B.3 b generates a bar of length less than $2\kappa\varepsilon$ in the module $N^{\mathcal{L}_{\mathfrak{b}}}$

Consider $N' = \mathsf{S}_{2\kappa\varepsilon}(N)$. By Lemma 3.3.9 every generator b which satisfied B.3 now generates a bar of zero length in the module $N'^{\mathcal{L}_{\mathfrak{b}}}$. Consider the free resolution $(F'_\bullet, \partial'_\bullet)$ of N' induced by shifting the gradings of $(F_\bullet, \partial_\bullet)$ according to Lemma 3.3.8. For every generator b' of F'_0 with grade $\mathbf{b}'$ at least one of the following holds:

B′.1 $\mathbf{b}'$ lies $2\kappa\varepsilon$-close to $\operatorname{Im} \mathcal{G}_M$

B′.2 $\mathbf{b}'$ lies $\kappa\varepsilon$-close to $\mathcal{H}_{\mathcal{G}_M}$ and generates a bar of zero-length in the module $N'^{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon} \mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b}')}}$

B′.3 b' generates a bar of zero-length in the module $N^{\mathcal{L}_{\mathfrak{b}'}}$

Consider $\hat{N} = \mathsf{M}^{\mathcal{G}_M}_{2\kappa\varepsilon}(N')$. Every generator b' satisfying B′.1 is moved to grade $\mathsf{M}^{\mathcal{G}_M}_{2\kappa\varepsilon}(\mathbf{b}') \in \operatorname{Im} \mathcal{G}_M$. Every generator b' satisfying B′.2 lies $\kappa\varepsilon$-close to $\mathcal{H}_{\mathcal{G}_M}$ and there is a relation r' which kills the zero bar generated by b' in $N'^{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{\kappa\varepsilon} \mathsf{M}^{\mathcal{G}_M}_{\kappa\varepsilon}(\mathbf{b}')}}$. Lemma 3.4.8 implies that under $\mathsf{M}^{\mathcal{G}_M}_{2\kappa\varepsilon}$ the generator b' and relation r' are brought to the same grade. Consider the free resolution $(\hat{F}_\bullet, \hat{\partial}_\bullet)$ of $\hat{N}$, induced by shifting the gradings of $(F'_\bullet, \partial'_\bullet)$ with the merge function. Any generator $\hat{b}$ of $\hat{F}_0$ which corresponds to a generator b' of F'_0 which satisfied B′.2 or B′.3 generates a trivial bar in the line $\mathcal{L}_{\hat{\mathfrak{b}}}$ and thus is not a non-trivial generator of $\hat{N}$ (Lemma 3.4.5). Hence all non-trivial generators of

$\hat{N}$ lie on the grid points of M, $\xi_0(\hat{N}) \subset \operatorname{Im} \mathcal{G}_M$, and by construction $d_I(M, \hat{N}) < \kappa\varepsilon + 2\kappa\varepsilon + 2\kappa\varepsilon \leq 5\kappa\varepsilon$.

Now let $(\hat{F}_\bullet, \hat{\partial}_\bullet)$ be a *minimal* free resolution for $\hat{N}$. For each relation r with grade $\mathbf{r} \in \xi_1(\hat{N})$ the hypotheses of Lemma 3.4.10 hold (replacing $\kappa\varepsilon$ with $5\kappa\varepsilon$), thus $\mathbf{r}$ is $10\kappa\varepsilon$-close to $\mathcal{H}_{\mathcal{G}_M}$. More precisely for each relation r at least one of the following holds:

R.1 $\mathbf{r}$ is $10\kappa\varepsilon$-close to $\operatorname{Im} \mathcal{G}_M$ or $\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r})$ is $20\kappa\varepsilon$-close to $\operatorname{Im} \mathcal{G}_M$

R.2 $\mathbf{r}$ is $10\kappa\varepsilon$-close to $\mathcal{H}_{\mathcal{G}_M}$ and if r kills a bar in the module $\hat{N}^{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{20\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{20\kappa\varepsilon}(\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r}))}}$ that bar is of length less than $10\kappa\varepsilon$

Consider $\tilde{N} = \mathsf{M}^{\mathcal{G}_M}_{20\kappa\varepsilon} \circ \mathsf{S}_{10\kappa\varepsilon}(\hat{N})$. Note that the grades of the generators of $\hat{N}$ remain unchanged under these functors. We will show that the grades of the non-trivial relations of $\tilde{N}$ all lie on $\operatorname{Im} \mathcal{G}_M$.

Any relation satisfying R.1 is shifted in grading by no more than $10\kappa\varepsilon$ by the functor $\mathsf{S}_{10\kappa\varepsilon}$ and thus lies on a grid point of $\mathcal{G}_M$ after applying $\mathsf{M}^{\mathcal{G}_M}_{20\kappa\varepsilon}$. Consider instead a relation r with grade $\mathbf{r} \in \xi_1(\hat{N})$ satisfying R.2 and not R.1. We will denote by $\mathsf{S}_{10\kappa\varepsilon}(r)$ and $\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r})$ the relation and grading of the relation corresponding to r in the free resolution induced by shifting the gradings of $(\hat{F}_\bullet, \hat{\partial}_\bullet)$ according to Lemma 3.3.8. Property R.2 implies that if r kills a bar in the module $\hat{N}^{\mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{20\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{20\kappa\varepsilon}(\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r}))}}$ it is of length less than $10\kappa\varepsilon$. For notational convenience let $\mathcal{L}' = \mathcal{L}_{\mathsf{U}^{\mathcal{G}_M}_{20\kappa\varepsilon}\mathsf{M}^{\mathcal{G}_M}_{20\kappa\varepsilon}(\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r}))}$. The effect of the simplification functor on barcodes (Lemma 3.3.9) implies that if r killed a bar in the module $\hat{N}^{\mathcal{L}'}$ then $\mathsf{S}_{10\kappa\varepsilon}(r)$ kills a bar in the module $\mathsf{S}_{10\kappa\varepsilon}(\hat{N})^{\mathcal{L}'}$ and that bar is of length zero. However, it is not possible for $\mathsf{S}_{10\kappa\varepsilon}(r)$ to kill a zero length bar in $\mathsf{S}_{10\kappa\varepsilon}(\hat{N})^{\mathcal{L}'}$ since

$$\|\operatorname{push}_{\mathcal{L}'}(\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r})) - \operatorname{push}_{\mathcal{L}'}(\operatorname{Im} \mathcal{G}_M)\|_\infty \geq 20\kappa\varepsilon$$

and hence $\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r})$ cannot be pushed to the same value as a generator in the line $\mathcal{L}'$.

By considering the free resolution $(\mathsf{S}_{10\kappa\varepsilon}(\hat{F}^{\mathcal{L}'}_\bullet), \mathsf{S}_{10\kappa\varepsilon}(\hat{\partial}^{\mathcal{L}'}_\bullet))$ of $\mathsf{S}_{10\kappa\varepsilon}(\hat{N})^{\mathcal{L}'}$, we note that there is a second order relation s, a generator of $\mathsf{S}_{10\kappa\varepsilon}(\hat{F}_2)$, with $\operatorname{push}_{\mathcal{L}'}(\mathbf{s}) = \operatorname{push}_{\mathcal{L}'}(\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r}))$ and $\langle \hat{\partial}_2(s), \mathsf{S}_{10\kappa\varepsilon}(r) \rangle \neq 0$. We note that $\mathbf{s} \geq \mathsf{S}_{10\kappa\varepsilon}(\mathbf{r})$, $\|\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r}) - \mathcal{H}_{\mathcal{G}_M}\|_\infty \leq 20\kappa\varepsilon$ and hence Lemma 3.4.8 implies that under the merge $\mathsf{M}^{\mathcal{G}_M}_{20\kappa\varepsilon}$ the relation $\mathsf{S}_{10\kappa\varepsilon}(r)$ and second order relation s are brought to the same grade, $\mathsf{M}^{\mathcal{G}_M}_{20\kappa\varepsilon}\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r}) = \mathsf{M}^{\mathcal{G}_M}_{20\kappa\varepsilon}(\mathbf{s})$. If there are multiple relations of $\mathsf{S}_{10\kappa\varepsilon}(\hat{N})$ brought to the grade $\mathsf{M}^{\mathcal{G}_M}_{20\kappa\varepsilon}\mathsf{S}_{10\kappa\varepsilon}(\mathbf{r})$ under the merge then all of these relations are pushed to the same grade under $\operatorname{push}_{\mathcal{L}'}$ (Equation (3.6)) and all do not kill a bar in the module $\mathsf{S}_{10\kappa\varepsilon}(\hat{N})^{\mathcal{L}'}$. By again considering the free resolution $(\mathsf{S}_{10\kappa\varepsilon}(\hat{F}^{\mathcal{L}'}_\bullet), \mathsf{S}_{10\kappa\varepsilon}(\hat{\partial}^{\mathcal{L}'}_\bullet))$ we see that there must be a collection of second order relations equinumerous with these relations, such that the collection of second order relations surject onto

the span of the relations in $\mathsf{S}_{10\kappa\varepsilon}(\hat{F}_1^{\mathcal{L}'})$. Thus all these relations are trivial after the merge and all non-trivial relations of $\tilde{N}$ lie on $\operatorname{Im}\mathcal{G}_M$.

In total we observe that $\tilde{N} = \mathsf{M}^{\mathcal{G}_M}_{20\kappa\varepsilon} \circ \mathsf{S}_{10\kappa\varepsilon} \circ \mathsf{M}^{\mathcal{G}_M}_{2\kappa\varepsilon} \circ \mathsf{S}_{2\kappa\varepsilon}(N)$ is such that $\xi_0(\tilde{N}) \cup \xi_1(\tilde{N}) \subset \operatorname{Im}\mathcal{G}_M$ and Lemma 3.2.16 implies that in fact $\bigcup_i \xi_i(\tilde{N}) \subset \mathcal{G}_M$. Since the simplification functor S_δ and merge functor $\mathsf{M}^{\mathcal{G}}_\delta$ each perturb a module by no more that δ in the interleaving distance we have that $d_I(N, \tilde{N}) \leq 20\kappa\varepsilon + 10\kappa\varepsilon + 2\kappa\varepsilon + 2\kappa\varepsilon = 34\kappa\varepsilon$. □

The final auxilliary lemma we require, Lemma 3.4.12, is a specialised version of (58, Lemma 6.7).

Lemma 3.4.12 (Sufficient Conditions for a Small Interleaving to Induce an Isomorphism (58, Lemma 6.7))**.** *Suppose $M, N \in \mathbf{vect}^{\mathbf{R}^n}_{fin}$ are ε-interleaved and that for all $\mathbf{a} \in \bigcup_i \xi_i(M) \cup \xi_i(N)$ the transition morphisms $(M_\mathbf{a} \to M_{\mathbf{a}+\varepsilon})$,$(M_\mathbf{a} \to M_{\mathbf{a}+2\varepsilon})$, $(N_\mathbf{a} \to N_{\mathbf{a}+\varepsilon})$ and $(N_\mathbf{a} \to N_{\mathbf{a}+2\varepsilon})$ are isomorphisms of vector spaces, then M, N are isomorphic.*

Combining the results of this section we attain a short proof of our main result.

Proof. (Theorem 3.4.1) Suppose (seeking contradiction) that $d_0(M, N) < \kappa\varepsilon$ while $d_I(M, N) = \varepsilon$. Then by Lemma 3.4.11 there is a module $\tilde{N}$ with $\bigcup_i \xi_i(\tilde{N}) \subset \operatorname{Im}\mathcal{G}_M$, and $d_I(N, \tilde{N}) \leq 34\kappa\varepsilon$. By the triangle inequality $d_I(\tilde{N}, M) \leq \varepsilon + 34\kappa\varepsilon$. Since $c_M > 2\varepsilon + 68\kappa\varepsilon$, the internal morphisms $(\tilde{N}_\mathbf{a} \to \tilde{N}_{\mathbf{a}+\varepsilon+34\kappa\varepsilon})$, $(\tilde{N}_\mathbf{a} \to \tilde{N}_{\mathbf{a}+2\varepsilon+68\kappa\varepsilon})$, $(M_\mathbf{a} \to M_{\mathbf{a}+\varepsilon+34\kappa\varepsilon})$ and $(M_\mathbf{a} \to M_{\mathbf{a}+2\varepsilon+68\kappa\varepsilon})$ are isomorphisms for all $\mathbf{a} \in \bigcup_i \xi_i(\tilde{N}) \cup \bigcup_i \xi_i(M) \subset \operatorname{Im}\mathcal{G}_M$. Hence Lemma 3.4.12 implies that $d_I(M, \tilde{N}) = 0$. This leads to a contradiction for all $\kappa \in [0, \frac{1}{34})$:

$$\varepsilon = d_I(M, N) \leq d_I(M, \tilde{N}) + d_I(\tilde{N}, N) \leq 34\kappa\varepsilon < \varepsilon = d_I(M, N).$$

□

3.5 *Equivalence of Intrinsic Metrics*

In this section we show that the local equivalence result in the previous section induces a global equivalence of intrinsic metrics. A consequence of the Characterisation Theorem in (58, Theorem 4.4) (which characterises the presentations of ε-interleaved multiparameter persistence modules), is that the space of finitely presented multiparameter persistence modules equipped with the interleaving distance is a path metric space. That is to say a pair of finitely presented modules at distance ε in the interleaving distance are joined by an ε-length path. Consequently, we deduce a global equivalence of intrinsic metrics between the induced intrinsic matching distance $\hat{d}_0$ and the interleaving distance (Theorem 3.5.6).

In a metric space (X, d) one can define the length of a continuous path γ. For a more thorough introduction to path metric spaces and induced intrinsic metrics see (72).

Definition 3.5.1 (Length of a Path (72))**.** *Define the length of a path $\gamma : [0,1] \to X$ continuous with respect to the metric d to be given by:*

$$\ell_d(\gamma) = \sup_{\substack{0=t_0<t_1<\ldots<t_n=1 \\ n\in\mathbb{N}}} \sum_{i=1}^{n} d(\gamma(t_{i-1}), \gamma(t_i))$$

From the definition of the length of a path we can construct intrinsic metrics on $(\mathbf{vect}^{\mathbf{R}^n}_{\text{fin}}, d_I)$.

Definition 3.5.2 (Intrinsic Metric)**.** *Given a metric d on the space of multiparameter modules we define an associated intrinsic metric $\hat{d}$ with respect to a collection of admissable paths $\mathcal{C}$ (where $\mathcal{C}$ contains all constant paths and is closed under concatenations of paths).*

$$\hat{d}(M, N) = \inf_{\substack{\gamma\in\mathcal{C} \\ \gamma(0)=M \\ \gamma(1)=N}} \ell_d(\gamma) \tag{3.7}$$

Let us recall the characterisation of interleavings proved by Lesnick:

Theorem 3.5.3 (Interleaving Characterisation (58, Theorem 4.4))**.** *Let $M, N \in \mathbf{Vect}^{\mathbf{R}^n}$ then M, N are ε-interleaved if and only if there exist presentations:*

$$M \cong \langle \mathcal{X}_M, \mathcal{X}_N(-\varepsilon) | \mathcal{R}_M, \mathcal{R}_N(-\varepsilon) \rangle$$
$$N \cong \langle \mathcal{X}_M(-\varepsilon), \mathcal{X}_N | \mathcal{R}_M(-\varepsilon), \mathcal{R}_N \rangle$$

Moreover if M and N are finitely presented these presentations may be chosen to be finite.

Remark 3.5.4. The presentations in the result of Theorem 3.5.3 are constructed using the data of the interleaving morphisms. We note that in this construction each relation in $\mathcal{R}_M$ is a homogeneous element of the free module $\text{Free}[\mathcal{X}_M \sqcup \mathcal{X}_N(-\varepsilon)]$. Similarly, each relation in $\mathcal{R}_N$ is a homogeneous element of the free module $\text{Free}[\mathcal{X}_M(-\varepsilon) \sqcup \mathcal{X}_N]$. This means that $\langle \mathcal{X}_M(-t\varepsilon), \mathcal{X}_N((t-1)\varepsilon) | \mathcal{R}_M(-t\varepsilon), \mathcal{R}_N((t-1)\varepsilon)) \rangle$ is a well-defined presentation for all $t \in [0,1]$.

A straight forward consequence of the characterisation of interleavings is the following result:

Corollary 3.5.5 ($(\mathbf{vect}^{\mathbf{R}^n}_{\text{fin}}, d_I)$ is a Path Metric Space)**.** *The extended metric space of finitely presented multiparameter persistence modules equipped with the interleaving distance, $(\mathbf{vect}^{\mathbf{R}^n}_{fin}, d_I)$, is a path metric space.*

Proof. Given a pair of finitely presented modules M, N with $d_I(M,N) = \varepsilon$, since the modules are finitely presented, the interleaving distance is realised and so they are ε-interleaved (58, Theorem 6.1). Theorem 3.5.3 implies that there exist presentations $M \cong \langle \mathcal{X}_M, \mathcal{X}_N(-\varepsilon) | \mathcal{R}_M, \mathcal{R}_N(-\varepsilon)\rangle$, $N \cong \langle \mathcal{X}_M(-\varepsilon), \mathcal{X}_N | \mathcal{R}_M(-\varepsilon), \mathcal{R}_N\rangle$. Consider the path $\gamma : [0,1] \to \mathbf{vect}_{\text{fin}}^{\mathbf{R}^n}$ defined by:

$$\gamma(t) = \langle \mathcal{X}_M(-t\varepsilon), \mathcal{X}_N((t-1)\varepsilon) | \mathcal{R}_M(-t\varepsilon), \mathcal{R}_N(t-1)\varepsilon)\rangle$$

The path γ is well-defined by Remark 3.5.4. Let $t < s$ and $\mathcal{X}_t = \mathcal{X}_M(-t\varepsilon), \mathcal{X}_s = \mathcal{X}_N((s-1)\varepsilon), \mathcal{R}_t = \mathcal{R}_M(-t\varepsilon)$, and $\mathcal{R}_s = \mathcal{R}_N((s-1)\varepsilon)$. Observe that:

$$\gamma(t) = \langle \mathcal{X}_t, \mathcal{X}_s((t-s)\varepsilon) | \mathcal{R}_t, \mathcal{R}_s((t-s)\varepsilon)\rangle$$
$$\gamma(s) = \langle \mathcal{X}_t((t-s)\varepsilon), \mathcal{X}_s | \mathcal{R}_t((t-s)\varepsilon), \mathcal{R}_s\rangle$$

Hence using Theorem 3.5.3 we see that $d_I(\gamma(t), \gamma(s)) \leq |s-t|\varepsilon$. □

Since the space of finitely presented multiparameter modules equipped with the interleaving distance is a path metric space (Corollary 3.5.5) there exists a distance realising path between each pair of modules and so $\hat{d}_I = d_I$. Thus the local result of Theorem 3.4.1 extends to a global result about intrinsic metrics.

Theorem 3.5.6 (Equivalence of Intrinsic Metrics)**.** *Let $\hat{d}_0$ denote the intrinsic matching distance with respect to the collection of admissable paths $\mathcal{C} = \{\gamma : [0,1] \to (\mathbf{vect}_{fin}^{\mathbf{R}^n}, d_I) : \gamma \text{ is } d_I\text{-continuous}\}$. The intrinsic metrics $\hat{d}_0$ and d_I are globally bi-Lipschitz equivalent so that for all $M, N \in \mathbf{vect}_{fin}^{\mathbf{R}^n}$:*

$$\frac{1}{34} d_I(M,N) \leq \hat{d}_0(M,N) \leq d_I(M,N)$$

Proof. The upper bound $\hat{d}_0 \leq d_I$ follows from $d_0 \leq d_I$. Indeed for any path γ we have that between modules M and N

$$\begin{aligned}\ell_{d_0}(\gamma) &= \sup_{\substack{0=t_0<t_1<\ldots<t_n=1 \\ n\in\mathbb{N}}} \sum_{i=1}^{n} d_0(\gamma(t_{i-1}), \gamma(t_i)) \\ &\leq \sup_{\substack{0=t_0<t_1<\ldots<t_n=1 \\ n\in\mathbb{N}}} \sum_{i=1}^{n} d_I(\gamma(t_{i-1}), \gamma(t_i)) = \ell_{d_I}(\gamma).\end{aligned}$$

$\hat{d}_0 = \inf_\gamma \ell_{d_0}(\gamma)$ and Corollary 3.5.5 implies that $d_I = \inf_\gamma \ell_{d_I}(\gamma)$ thus $\hat{d}_0 \leq d_I$.

For the lower bound we will show the length of any path between M and N is bounded below by $\frac{1}{34} d_I(M,N)$. Let $\gamma \in \mathcal{C}$ with $\gamma(0) = M$

and $\gamma(1) = N$. For every $t \in [0,1]$ let $I_t \subset \gamma^{-1}B(\gamma(t), \frac{c_{\gamma(t)}}{4})$, be the largest interval containing t in the preimage of the radius $\frac{c_{\gamma(t)}}{4}$ ball about the module $\gamma(t)$ in $\mathbf{vect}_{\text{fin}}^{\mathbf{R}^n}$. By compactness of $[0,1]$ there is a minimal finite subcover $\{I_{t_1}, ..., I_{t_{2k+1}}, ..., I_{t_{2n+1}}\}$. Let $t_{2k} \in I_{t_{2k-1}} \cap I_{t_{2k+1}}$. Then by the local result Theorem 3.4.1 $d_0(\gamma(t_{2k}), \gamma(t_{2k\pm1})) \geq \frac{1}{34} d_I(\gamma(t_{2k}), \gamma(t_{2k\pm1}))$. Hence we observe that $\ell(\gamma) \geq \frac{1}{34} d_I(M, N)$. □

3.6 *Application to Interlevel Set Persistence and Reeb Graphs*

In this section we outline an application of Theorem 3.4.1 and Theorem 3.5.6 to interlevel set persistence modules (73; 56; 74) and Reeb Graphs. This section is intended to illustrate the generality of Theorem 3.4.1 and its applicability to data structures which arise in topological data analysis, even those which a priori do not have the structure of finitely presented multiparameter persistence modules.

Reeb graphs are a commonly studied object in data science and can be viewed as a particular instance of interlevel set persistence modules. Metrics on Reeb Graphs have been studied extensively (63; 75; 76; 77) and there is a trade off between discriminating power and computability of these metrics. Theoretically sound but computably infeasible metrics exist for Reeb Graphs, such as the functional distortion distance, d_{FD} and the Reeb Graph interleaving distance d_I^{Reeb}. It is shown in (76) that d_I^{Reeb} and d_{FD} are bi-Lipschitz equivalent. Computable, yet incomplete pseudo-metrics on Reeb Graphs include the bottleneck distance d_B (77). A comparison of the bottleneck distance and functional distortion distance is studied in (63), and a global equivalence of intrinsic metrics is established.

We can associate a 2-parameter persistence module to an interlevel set persistence module. However, the 2-parameter persistence module associated to a non-trivial interlevel set persistence module is not finitely presented. Naively embedding interlevel set persistence modules into $\mathbb{R}^2$-indexed persistence modules gives rise to 2-parameter persistence modules with an uncountable number of generators. However, using the block decomposition of interlevel set persistence modules (56; 55; 52) we may extend the 2-parameter persistence modules interlevel set persistence modules so that they are amenable to the arguments in Section 3.4 which apply to finitely presented multiparameter persistence modules.

Definition 3.6.1 (Interlevel Set Persistence Module)**.** *An interlevel set persistence module is an element of the functor category* $\mathbf{vect}^{\boldsymbol{Int}(\mathbb{R})}$*, where* $Int(\mathbb{R})$ *denotes the poset of non-empty bounded open intervals in* $\mathbb{R}$*.*

Example 3.6.2 (Interlevel Set Persistence Modules)**.** *Interlevel set persistence modules naturally arise from real valued functions on a topological*

space. Let $f : X \to \mathbb{R}$ and let H denote a homology functor with coefficients in a field. The functor $(a,b) \mapsto H(f^{-1}((a,b)))$ is an interlevel set persistence module.

One can define an interleaving distance for interlevel set persistence modules analogously to the interleaving distance for multi-parameter modules Definition 3.2.5. In this setting the translation endofunctors $T_\varepsilon : \mathbf{Int}(\mathbb{R}) \to \mathbf{Int}(\mathbb{R})$ augment the intervals by ε: $T_\varepsilon((b,d)) = (b-\varepsilon, d+\varepsilon)$. It is straight forward to show that interlevel set persistence modules isometrically embed as 2-parameter persistence modules.

Lemma 3.6.3 (Interlevel Set Persistence Module Embedding (56))**.** *Let $\mathbb{U}$ denote the subposet $\{x_2 > -x_1\} \subset \mathbb{R}^2$. The space of interlevel set persistence modules $\mathbf{vect}^{\mathbf{Int}(\mathbb{R})}$ equipped with the interleaving distance embeds isometrically into $\mathbf{vect}^{\mathbf{U}}$, using the identification $\mathbb{U} \cong Int(\mathbb{R})$ via $(x_1,x_2) \leftrightarrow (-x_1,x_2)$*

If we hope to associate a finitely presented module to an interlevel set persistence module we need to control the behaviour of interlevel set persistence modules. Definition 3.6.4 provides the necessary criteria to control interlevel set persistence modules.

Definition 3.6.4 (Finitely Determined)**.** *A module $M \in \mathbf{vect}^{\mathbf{R}^n}$ is finitely determined if there exists a grid function $\mathcal{G} : [k_1] \times [k_2] \to \mathbb{R}^2$ such that for all $\mathbf{a} \leq \mathbf{b}$ with $\mathbf{a},\mathbf{b} \in (\mathcal{G}(i),\mathcal{G}(i+1)) \times (\mathcal{G}(j),\mathcal{G}(j+1)) =: Cell_{i,j}$ the internal morphism $M(\mathbf{a} \leq \mathbf{b})$ is an isomorphism, and for all $\mathbf{a} \not\geq \mathcal{G}((1,1)), M(\mathbf{a}) = 0$. For a subset $A \subset \mathbb{R}^2$ we shall say a module $M \in \mathbf{vect}^{\mathbf{A}}$ is finitely determined if there exists a grid function $\mathcal{G} : [k_1] \times [k_2] \to \mathbb{R}^2$ such that for all $\mathbf{a} \leq \mathbf{b}$ with $\mathbf{a},\mathbf{b} \in (\mathcal{G}(i),\mathcal{G}(i+1)) \times (\mathcal{G}(j),\mathcal{G}(j+1)) \cap A$ the internal morphism $M(\mathbf{a} \leq \mathbf{b})$ is an isomorphism and for all $\mathbf{a} \not\geq \mathcal{G}((1,1)), M(\mathbf{a}) = 0$.*

Following Definition 3.6.4 we will say that an interlevel set persistence modules $M \in \mathbf{vect}^{\mathbf{Int}(\mathbb{R})}$ is finitely determined if the associated module $M' \in \mathbf{vect}^{\mathbf{U}}$ is finitely determined. We will denote the collection of finitely determined interlevel set persistence modules by $\mathbf{vect}^{\mathbf{Int}(\mathbb{R})}_{\text{fin}}$. We believe that this constructibility criterion is not too restrictive. For example, the interlevel set persistence module associated to a scalar function $f : X \to \mathbb{R}$ will be finitely determined if X is a compact manifold and f is Morse (Example 3.6.2); alternatively if X is a compact semi-algebraic subset of $\mathbb{R}^n$ and f is a projection onto one of the coordinate values (75).

Definition 3.6.5 (Block Decomposable (56))**.** *Let us define the following*

notation for intervals of $\mathbf{U}$ *which we shall call* $\mathbf{U}$*-blocks:*

$$
\begin{aligned}
(a,b)_{BL} &= \{(x,y) \in \mathbf{U} : x \in (-b,-a), y \in (a,b)\} \\
[a,b)_{BL} &= \{(x,y) \in \mathbf{U} : y \in [a,b)\} \\
(a,b]_{BL} &= \{(x,y) \in \mathbf{U} : x \in [-b,a)\} \\
[a,b]_{BL} &= \{(x,y) \in \mathbf{U} : x \in [-b,\infty), y \in [a,\infty)\}
\end{aligned}
$$

We say a module $M \in \mathbf{vect}^{\mathbf{R}^2}$ *is* $\mathbf{U}$*-block decomposable is it is interval decomposable and each of the intervals in the decomposition is a* $\mathbf{U}$*-block.*

We want to work in the subclass of finitely presented persistence modules. To this end, we define the extension of $\mathbf{U}$-blocks, which takes each $\mathbf{U}$-block to a finitely presented interval of $\mathbb{R}^2$.

Definition 3.6.6 ($\mathbb{U}$-Block Extension)**.** *We define the following extensions of* $\mathbf{U}$*-blocks to intervals of* $\mathbf{R}^2$*:*

$$
\begin{aligned}
\overline{(a,b)}_{BL} &= \{(x,y) \in \mathbf{R}^2 : x \in [-b,-a), y \in [a,b)\} \\
\overline{[a,b)}_{BL} &= \{(x,y) \in \mathbf{R}^2 : x \in [-b,\infty), y \in [a,b)\} \\
\overline{(a,b]}_{BL} &= \{(x,y) \in \mathbf{R}^2 : x \in [-b,a), y \in [a,\infty)\} \\
\overline{[a,b]}_{BL} &= \{(x,y) \in \mathbf{R}^2 : x \in [-b,\infty), y \in [a,\infty)\}
\end{aligned}
$$

Using this extension we can assign a finitely presented module $\overline{M} \in \mathbf{vect}^{\mathbf{R}^2}_{\text{fin}}$ to every $\mathbf{U}$-block decomposable module $M \in \mathbf{vect}^{\mathbf{R}^2}$. Explicitly if $M \cong \bigoplus_j \mathbb{1}^{I_j}$ (with I_j all $\mathbf{U}$-blocks) we take $\overline{M} \cong \bigoplus_j \mathbb{1}^{\overline{I_j}}$.

The interleaving distance between $\mathbf{U}$-block decomposable modules can be computed by matching the blocks in their decomposition (32). More precisely, given $\mathbf{U}$-block decomposable modules $M \cong \bigoplus_j \mathbb{1}^{I_j}$ and $N \cong \bigoplus_k \mathbb{1}^{J_k}$ if $d_I(M,N) = \varepsilon$ then there is a matching between the barcodes $\{I_j\}$ and $\{J_k\}$ such that if I_j is matched to J_k then I_j and J_k are of the same type, $d_I(\mathbb{1}^{I_j}, \mathbb{1}^{J_k}) \leq \varepsilon$ and any unmatched block is ε-interleaved with the zero module (32).

It is straight forward to check that if I_j and J_k are the same type then:

$$d_I(\mathbb{1}^{I_j}, \mathbb{1}^{J_k}) \leq d_I(\mathbb{1}^{\overline{I_j}}, \mathbb{1}^{\overline{J_k}}) \leq 2d_I(\mathbb{1}^{I_j}, \mathbb{1}^{J_k})$$

and so for $\mathbf{U}$-block decomposable modules M, N:

$$d_I(M,N) \leq d_I(\overline{M}, \overline{N}) \leq 2d_I(M,N)$$

We can derive an interlevel set persistence module from a Reeb Graph. In Figure 3.11 we sketch the 2-parameter $\mathbf{U}$-block decomposable module associated to a simple Reeb Graph together with its extension to a finitely presented 2-parameter module. Since we can extend interlevel set persistence modules to finitely presented 2-parameter persistence modules we yield the following result as an application of Proposition 3.4.2:

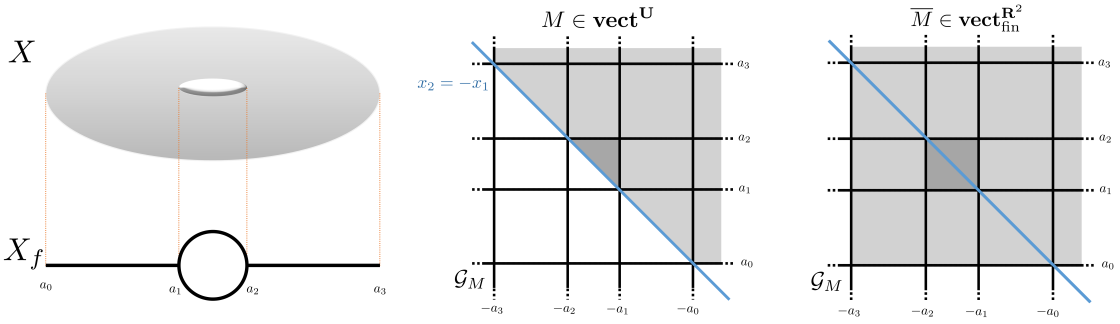


Figure 3.11: An illustration of the Reeb graph X_f, where $X = \mathbb{T}^2$ is the 2-torus equipped with the projection f to the real line. The Reeb graph X_f has an associated **vect** valued-cosheaf on the real line which isometrically embedded into $\mathbf{vect}^{\mathbf{U}}$ (both spaces equipped with the interleaving distance). Using the block decomposition of an interlevel set persistence module $M \in \mathbf{vect}^{\mathbf{U}}$ we may extend M to $\overline{M} \in \mathbf{vect}^{\mathbf{R}^2}_{\text{fin}}$.

Corollary 3.6.7 (Local Equivalence of Interlevel Set Persistence Module Metrics). *Suppose $M, N \in \mathbf{vect}_{fin}^{\boldsymbol{Int}(\mathbb{R})}$ and $\kappa \in [0, \frac{1}{34})$ are such that $d_I(M,N) = \varepsilon$ and $\frac{c_{\overline{M}}}{4(34\kappa+1)} > \varepsilon$ then the matching distance between their extended modules is bounded below $d_0(\overline{M}, \overline{N}) > \kappa\varepsilon$.*

Proof. Apply Proposition 3.4.2 to the modules $\overline{M}$ and $\overline{N}$ using the fact that $d_I(\overline{M}, \overline{N}) \leq 2d_I(M,N)$. □

Corollary 3.6.7 provides an example of a local equivalence result which may be derived from our general result Proposition 3.4.2. We relied on the fact that the poset $\mathbf{Int}(\mathbb{R})$ naturally includes into $\mathbf{R}^2$ and that sensible constructibility criteria for interlevel set persistence modules give rise to finitely presented 2-parameter persistence modules. In this particular instance, interlevel set persistence modules and Reeb Graphs have been well studied and have additional structure which has been exploited to yield sharper results than Corollary 3.6.7 (32; 74; 56; 63; 76). We would be interested to explore other posets $\mathbf{P}$ which naturally include into $\mathbf{R}^n$ for some n, and the corresponding constructibility criteria of modules $\mathbf{vect}^{\mathbf{P}}$ for which this inclusion would give rise to finitely presented modules. This would take advantage of the generality of our main result Theorem 3.4.1 and yield results for more complicated topological data structures than Reeb Graphs. A starting point would be to explore Reeb Spaces.

Acknowledgements

The author would like to thank his supervisors Ulrike Tillmann and Vidit Nanda for their support during his research. In addition, the author would like to thank Jacob Leygonie for many constructive discussions which cemented the technical arguments in this work. The author is grateful to receive support from EPSRC studentship EP/N509711/1 and EPSRC grant EP/R018472/1.

4
Random Čech Complexes on Riemannian Manifolds with Boundary

Contents

Abstract

Let M be a compact, unit volume, Riemannian manifold with boundary. In this paper we study the homology of a random Čech-complex generated by a homogeneous Poisson process in M. Our main results are two asymptotic threshold formulas, an upper threshold above which the Čech complex recovers the k-th homology of M with high probability, and a lower threshold below which it almost certainly does not. These thresholds are close together in the sense that they have the same leading term. Here k is positive and strictly less than the dimension d of the manifold.

This extends work of Bobrowski and Weinberger in (78) and Bobrowski and Oliveira (79) who establish similar formulas when M is a torus and, more generally, is closed and has no boundary. We note that the cases with and without boundary lead to different answers: The corresponding common leading terms for the upper and lower thresholds differ being $\log(n)$ when M is closed and $(2-2/d)\log(n)$ when M has boundary; here n is the expected number of sample points. Our analysis identifies a special type of homological cycle, which we call a Θ-like-cycle, which occur close to the boundary and establish that the first order term of the lower threshold is $(2-2/d)\log(n)$.

4.1 Introduction

Randomly generated simplicial complexes and their topology have recently attracted a lot of attention. The survey article by Bobrowski and Kahle (80) collects together results and provides a wealth of open problems in this field. Here we focus on understanding the homology of random geometric complexes. While this topic was first studied in (81) (see also (82)) our results build directly on (79) and (78).

Much of the current interest in the topology of random simplicial complexes is due to applications to topological data analysis, where random complexes can serve as null models when interpreting the

topology of complexes on data sets. In the context of manifold learning, a non-linear dimension reduction technique, one is interested in recovering the structure of low dimensional manifolds embedded in high dimensional space. Studying topological properties of the underlying manifold, such as homology, can inform the choices of hyperparameters in this dimension reduction technique (83; 84). For example, one can measure the change in the topology of a manifold (represented by a point cloud) under a dimension reduction procedure by measuring the change in the homology of the manifold, and thus tune the dimension reduction procedure to minimise this change (83).

In the specific context of persistent homology, one of the main tools of topological data analysis, one adopts a multiscale approach to the study of the homology of randomly sampled data from a manifold. Understanding the conditions for which the topology of a complex built from a random sample coincides with the homology of the underlying manifold informs which bars in the multiscale barcode invariant relate to inherent features of the underlying manifold. For surveys of persistent homology see (85; 86; 87) and for work in stochastic persistent homology (88).

Various flavours of random simplicial complexes are present in the literature, which can include or exclude geometric considerations. Random simplicial complexes that extend the notion of a random graph in the sense of Erdös-Rényi to higher dimensional complexes are not constrained by an underlying geometry (89). In contrast, we shall work with the random simplicial complex realised by the Čech complex associated to a Poisson point process on a Riemannian manifold and ask the question when the topology of the simplicial complex approximates that of the manifold.

The question of recovering the topology of a space from a finite sample has been studied in (90; 91; 92) and (79; 78) in differing contexts. In (90) the authors consider submanifolds in Euclidean space and use the metric of the ambient Euclidean space when building the Čech complex. They provide explicit conditions to recover the homology of the manifold with high confidence. Naturally these explicit conditions are dependent on the curvature and nearness to self-intersection of the embedded manifold. This theory has been extended to manifolds with boundary in (92) and more general compact subset in (91). In contrast, as in the context of (79; 78), we base the construction of the Čech complex on an intrinsic metric of the manifold independent of any embedding. Our work studies asymptotic properties and the phase transition for which one can recover the homology with high probability when increasing the size of point sample and decreasing the radius over which the associated complex

is constructed. Specifically, we will give an answer to the problem posed in the survey article (80) about extending the homological connectivity theory established for closed Riemannian manifolds in (79), to Riemannian manifolds with boundary.

A principal advantage of studying asymptotics is that results rely on fewer assumptions on the underlying manifold from which the point process is sampled. Our main result like that of (79) thus only has dependence on the dimension of the underlying manifold and the homological dimension one wishes to recover. Our argument follows a similar framework to the argument presented in (79) and (78). We will need to develop completely new arguments to take into account the effect of the boundary.

Our main result is stated with respect to the term $\Lambda := n\omega_d r^d$, the expected number of points of a uniform Poisson process of intensity n lying in a d-dimensional radius r ball. Depending on the asymptotic behaviour of the Λ, the associated Čech complex at scale r built on the point process exhibits different behaviours.

There are three distinct regimes of behaviour for Λ as $n \to \infty, r \to 0$. In the subcritical regime ($\Lambda \to 0$) the connectivity of the Čech complex is very sparse and mostly disconnected, with the number of connected components growing at the same rate as the number of points. In the critical regime ($\Lambda \to \lambda \in (0, \infty)$) the Čech complex is sufficiently connected to exhibit non-trivial homology. However the number of connected components still grows linearly with the number of points. In the supercritical regime ($\Lambda \to \infty$) for sufficiently large Λ the Čech complex is connected, and for even larger Λ the point cloud covers the underlying manifold with high probability.

Analysis in the supercritical regime yields a sequence of increasing thresholds, (*homological connectivity thresholds*), such that if Λ is greater than the k^{th} threshold the Čech complex recovers the k^{th} homology of the underlying closed manifold with high probability. The intermediate homological connectivity thresholds interpolate between the thresholds for more commonly studied properties, from the 0^{th} homology which detects connectivity up to the d^{th} homology which detects coverage. We produce homological connectivity thresholds in the supercritical regime for which the Čech complex recovers the homology of a smooth compact manifold with non-trivial boundary.

The non-trivial boundary has a significant impact on the homological connectivity thresholds. As far as we know, our study of manifolds with boundary has uncovered a previously unobserved phenomenon occuring close to the boundary. Our analysis shows that close to the boundary a large number of spurious k-cycles appear which are not homological cycles inherent to the k^{th} homology of the underlying manifold. This phenomenon determines that the

homological connectivity thresholds for manifolds with boundary are larger than those for a closed manifold. We attain the following result:

Theorem. *(Homological Connectivity of a Riemannian Manifold with Boundary)*

Let M be a unit volume compact Riemannian manifold with smooth non-empty boundary. Let $d \geq 2$ be the dimension of M, $\Lambda = \omega_d n r^d$ and $\mathcal{P}_n$ a Poisson process of intensity n on M. Suppose $w(n)$ is any function with $w(n) \to \infty$ as $n \to \infty$. Then for $1 \leq k \leq d-1$

$$\lim_{n\to\infty} \mathbb{P}(H_k(\mathcal{C}(n,r)) \cong H_k(M)) = \begin{cases} 1 & \Lambda = (2-\frac{2}{d})\log n + 2k \log\log n + w(n), \\ 0 & \Lambda = (2-\frac{2}{d})\log n + 2(k-2-(k+1-\frac{1}{d}))\log\log n - w(n), \end{cases}$$

Note that when simplified the coefficient of the second order term in the lower threshold, $2(\frac{1}{d}-3)$, is independent of the homological dimension k. We state the coefficient of the second order term of the lower threshold of our theorem in unsimplified form to make an easy comparison to the lower thresholds established by Bobrowski and Weinberger in (78) and by Bobrowski and Oliveira in (79), who studied the case when M is a d-dimensional torus and when M is an arbitrary compact closed Riemannian manifold respectively. With the same setup as our Theorem but on a closed Riemannian manifold the corresponding thresholds for $1 \leq k \leq d-1$ are computed to be:

$$\lim_{n\to\infty} \mathbb{P}(H_k(\mathcal{C}(n,r)) \cong H_k(M)) = \begin{cases} 1 & \Lambda = \log n + k\log\log n + w(n), \\ 0 & \Lambda = \log n + (k-2)\log\log n - w(n), \end{cases}$$

We see that the leading term for both the upper and lower bounds are nearly twice as large as in the case of manifolds with boundaries and large dimension d. At the end of the paper we provide an intuitive explanation as to why the presence of a boundary results in differences in the homological connectivity thresholds.

4.1.1 Outline

Our argument follows the same structure as the arguments presented in (79; 78). There are several key ideas in this framework. The first essential idea is to bound the number of homological cycles of a complex by counting the critical points of an associated Morse function. This simplifies the task of counting global phenomena of homological cycles to the purely local considerations which determine critical points of a Morse function.

In order to compute bounds for the number of critical points we require a change of variables integral formula, the Blaschke-Petkantschin formula. This change of variable formula facilitates computing the expected number of critical points induced by the distance function of a Poisson point process.

Within our argument we adapt results which apply to closed manifolds to manifolds with boundary. We use the double manifold as a canonical closed manifold in which our manifold with boundary is embedded. This trick allows us to translate results for closed manifolds to manifolds with boundary.

In Section 4.2 we collate results from the theory of point processes, define our asymptotic notation and also collect Riemannian volume estimates which we use in order to produce bounds in the later sections.

Section 4.3 provides a brief introduction to classical Morse theory. Since the distance function induced by a point process is not necessarily smooth we also provide the necessary results from (93) which describe Morse theory for a wider class of functions to which the distance function belongs, the so-called min-type functions.

In Section 4.4 we derive a coverage result for Riemannian manifolds with boundary. This coverage result is required to establish the upper homological connectivity thresholds.

We introduce in Section 4.5 the change of variables formulae we require in later sections to compute bounds on the number of critical points. These formulae are also used in Section 4.10 in order to bound the variance of the number of critical points.

Section 4.6 is dedicated to computing an upper bound for the expected number of critical points using the tools and results provided in previous sections. This upper bound on the number of critical points is used to produce an upper threshold for the homological connectivity in terms of Λ.

In Section 4.7 we identify a special class of critical point which induce erroneous homological cycles, (homological cycles which are not inherent to the underlying manifold). We produce a lower bound for the expected number of this type of critical points. This lower bound is used to produce a lower threshold for the homological connectivity in terms of Λ.

In Section 4.8 we calculate homological connectivity thresholds given the computed bounds on the number of critical points. Finally in Section 4.9 we compare and contrast the homological connectivity thresholds for Riemannian manifolds with boundary we have calculated to those in (79) for closed manifolds. We indicate how the geometric difference in the two situations inform the difference in the thresholds.

For completeness we include an Appendix containing proofs which are similar in structure to their analogues in the case that the manifold is closed. In Section 4.10 we bound the variance of the number of occurences of the special type of homological cycles established in Section 4.7. We show that the variance is small and so with high probability the number of erroneous cycles behaves like the expected number of erroneous cycles.

Acknowledgements

The authors would like to give recognition to The Alan Turing Institute through which the authors met, and this collaboration was initiated. They would also like to thank Omer Bobrowski for helpful comments on an earlier version of this paper and the anonymous reviewers whose comments helped improve and clarify this manuscript. HdK acknowledges support from the EPSRC studentship EP/L016508/1, and gratefully thanks The Alan Turing Institute for hosting him as an Enrichment student. UT acknowledges support from The Alan Turing Institute through EPSRC grant EP/N510129/1. OV gratefully acknowledges support from EPSRC studentship EP/N509711/1. UT and OV are members of the Oxford based Centre for Topological Data Analysis funded by EPSRC grant EP/R018472/1.

4.2 Background

In this section we shall provide a summary of the background theory on which we build our proofs. Results in this section are well documented so will mostly be stated without proof although references are provided for those who seek further details.

4.2.1 Homology and Complexes

Homology is a measure of complexity of a topological space. It is an algebraic invariant that has proved a powerful tool in the study of geometry. A thorough introduction can be found in (94) far more comprehensive than the brief overview of basic homology theory we present for the unfamiliar reader.

At it's most basic level, the homology of a space can be thought of as a sequence of abelian groups, where the k^{th} group summarises topological information about the k-dimensional subspaces. This algebraic summary is an incomplete invariant for homotopic spaces, that is to say, homotopic spaces have isomorphic homology groups, although it is possible for non-homotopic spaces to have isomorphic

homology groups.

For the purposes of this paper we shall consider homology over coefficients in a field, in which case the algebraic summary is a sequence of vector spaces. Equally in this setting one can consider our algebraic summary as a sequence of integers corresponding to the dimensions of these vector spaces, known as the Betti numbers of the space.

In order to make a topological space amenable to computation, we introduce a purely combinatorial object known as an abstract simplicial complex, or simplicial complex for short. A simplicial complex prescribes the construction of a space out of simplices in which one glues together vertices, edges, triangles, tetrahedra and their higher dimensional analogues. The following definition describes a simplicial complex constructed from a point process on a metric space.

Definition 4.2.1 (Čech Complex)**.** *Let $\mathcal{P}$ be a collection of points in a metric space (M, ρ). Define a one parameter family $\mathcal{C}(\mathcal{P}, r)$ of simplicial complexes on vertex set $\mathcal{P}$ associated to this collection of points as follows: For $r \in [0, \infty)$,*

$$\sigma = [p_0, ..., p_k] \in \mathcal{C}(\mathcal{P}, r) \iff \bigcap_{j=0}^{k} B_r(p_j) \neq \emptyset.$$

Here $B_r(p)$ denotes the open ball in M of radius r and with centre p.

Lemma 4.2.2 (Nerve Lemma)**.** *Let $\mathcal{U} = \{B_r(p) : p \in \mathcal{P}\}$ be an open cover of the metric space (M, ρ). Suppose that the finite intersections of sets in $\mathcal{U}$ are empty or contractible, then the Čech complex $\mathcal{C}(\mathcal{P}, r)$ is homotopy equivalent to M.*

The Nerve Lemma gives us a guarantee that for a suitably dense point process on a metric space and a well chosen radius r we will be able to recover the homology of the underlying space.

4.2.2 Poisson Point Processes

We shall introduce in this section the notion of a general Poisson point process. Let us remark here that although later discussions will use uniformly distributed processes (with respect to the volume measure), this is merely a point of convenience. Since we are considering compact manifolds, more general distributions will only effect our results up to some constant factor.

Definition 4.2.3 (General Poisson Point Process (95))**.** *Let $(M, \mathcal{F}, \mu)$ be a measure space with M a compact metric space, $\mathcal{F}$ a collection of measurable sets, and measure μ which is finite on compact sets and with no atoms.*

The Poisson process on M of intensity measure μ is a point process on M such that:

1. *For every compact set $K \subset M$ the number of points $N(K)$ lying in K follows a Poisson distribution with mean $\mu(K)$;*
2. *If $K_i \subset M$ are disjoint and compact then $N(K_i)$ are independent.*

In our setting we shall consider uniform Poisson point processes on a Riemannian manifold (M, g) with measure $\mu = \mathrm{vol}_g$ the Riemannian volume form. A Poisson process of intensity n on M is thus a Poisson process on M of intensity measure $n\,\mathrm{vol}_g$.

A salient feature of Poisson point processes is the following independence result, Theorem 4.2.4. We shall use this result in our calculations to compute bounds on the number of critical points of the distance function from the point process which correspond to simplices of the associated Čech complex. Intuitively the result follows from the memoryless property of the Poisson distribution and the spatial independence property of Poisson point processes. Conditioning on the position of $k+1$ points of the process $\mathcal{P}$, the remaining points follow the same distribution as $\mathcal{P}$.

Theorem 4.2.4 (Palm Theory (96))**.** *Let (X, ρ) be a metric space[5], $f : X \to \mathbb{R}$ a probability density and $\mathcal{P}_n$ a Poisson process on X with intensity $\lambda_n = nf$. If $h(\mathcal{Y}, \mathcal{X})$ is a measurable function for all finite subsets $\mathcal{Y}, \mathcal{X} \subset X$, with $\mathcal{Y} \subset \mathcal{X}$ and $|\mathcal{Y}| = k+1$ then:*

$$\mathbb{E}_{\mathcal{P}_n}\left[\sum_{\substack{\mathcal{Y}\subset\mathcal{P}_n \\ |\mathcal{Y}|=k+1}} h(\mathcal{Y}, \mathcal{P}_n)\right] = \frac{n^{k+1}}{(k+1)!}\mathbb{E}_{\mathcal{P}_n,\mathcal{Y}'}[h(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)]$$

where $\mathcal{Y}'$ is a set of $k+1$ i.i.d points in X with density f independent of $\mathcal{P}_n$.

[5] This result is stated in (96) for the special case that $X = \mathbb{R}^n$ equipped with the usual Euclidean metric. Provided a Poisson point process exists on the metric space X the proof follows identically. For the existence of a Poisson point process with intensity measure λ it is sufficient for λ to be an s-finite measure (see for example (97, Existence Theorem 3.6)).

4.2.3 *Asymptotic Notation*

We shall use the following set of notation to denote different asymptotic behaviours of functions f, g:

1. $f(n) = O(g(n))$ if there is a constant C and $n_0 \in \mathbb{N}$ such that $|f(n)| \leq C|g(n)|$ for all $n > n_0$.
2. $f(n) = o(g(n))$ if $\lim_{n\to\infty} \frac{|f(n)|}{|g(n)|} = 0$.
3. $f(n) = \Omega(g(n))$ if there is a constant C and $n_0 \in \mathbb{N}$ such that $|f(n)| \geq C|g(n)|$ for all $n > n_0$.
4. $f(n) \sim g(n)$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

4.2.4 *Riemannian Volumes*

A major portion of our later proofs require us to bound various Riemannian volumes by their Euclidean counterparts in order to control the asymptotic behaviour of the Betti numbers. In this section we shall provide these approximations of Riemannian volumes. Let us denote a smooth Riemannian manifold as the pair (M, g) and consider the case when M is compact, of dimension d, and the metric g is smooth. An introduction to Riemannian Geometry can be found for example in (98).

A smooth metric is a smoothly varying inner product on the tangent space $g : T_pM \times T_pM \to \mathbb{R}$ and therefore endows the tangent space at each point with a norm. We define the length of a path using g by integrating over the norm of its velocity. We shall use $\rho(p_1, p_2)$ to denote the shortest path length between two points p_1, p_2 on our manifold.

Let us denote the open ball of radius r about a point p on our manifold as $B_r(p)$ If we are considering a collection of points $\mathcal{P}$ then we denote the union of open balls of radius r centred at each point as $B_r(\mathcal{P}) := \bigcup_{p\in\mathcal{P}} B_r(p)$.

The exponential map $\exp_p : T_pM \to M$ is defined by $\exp_p(\mathbf{v}) = \gamma(1)$ where γ is the unique geodesic in M with $\gamma(0) = p$ and $\dot{\gamma}(0) = \mathbf{v}$. Since $\exp_p$ is a local diffeomorphism, an orthonormal basis of T_pM induces local coordinates about p, which we shall denote as $(x^1, x^2, ..., x^d)$ and are called geodesic normal coordinates. Using the Taylor expansion we can write the metric g in terms of these coordinates where R_{iklj} is known as the Riemann curvature tensor:

$$g = \sum_{i,j} g_{ij} dx^i \otimes dx^j \text{ with } g_{ij} = \delta_{ij} + \frac{1}{3}\sum_{k,l} R_{iklj} x^k x^l + O(|x|^3) \qquad (4.8)$$

The Euclidean metric on $M = \mathbb{R}^d$ is the simple case where $g_{ij} = \delta_{ij}$. Given a point p of a Riemannian manifold (M, g) with local neighbourhood U and geodesic normal coordinates $(x^1, ..., x^d)$ in the neighbourhood U, for sufficiently small radius r we can consider the intrinsic Euclidean ball $B_r^E(p)$ as the radius r ball with respect to the metric g_E on U where $g_E = \sum_{i,j} \delta_{ij} dx^i \otimes dx^j$.

The canonical measure induced by the Riemannian density on the manifold M can be expressed in terms of the Euclidean measure $|\text{dvol}_{g_E}|$ associated to the Euclidean metric:

$$|\text{dvol}_g| = \sqrt{|\det(g_{ij})|}|\text{dvol}_{g_E}|.$$

The Ricci curvature tensor at a point p is given by $Ric_{ij} = -\sum_k R_{ikkj}$,

and we can calculate that:

$$\sqrt{|\det(g_{ij})|} = 1 - \frac{Ric_{ij}}{3}x^i x^j + O(|x|^3).$$

Our first volume approximations apply to balls within a manifold with boundary which are wholly contained within the manifold, that is the centres lie far enough from the boundary. The proofs of these approximations can be found in the Appendix of (79).

Lemma 4.2.5 ((79)). *Let $s(p) = \sum_i Ric_{ii}$ denote the scalar curvature at p and define $s_{min}(\nu) = \inf_{p\in M} \frac{s(p)}{6(d+2)} - \nu$, $s_{max}(\nu) = \sup_{p\in M} \frac{s(p)}{6(d+2)} + \nu$. Then for all $\nu > 0$ there is a continuous choice of $r_\nu > 0$ such that for all $r \leq r_\nu$*

$$\omega_d r^d(1 - s_{max}(\nu)r^2) \leq Vol(B_r(p)) \leq \omega_d r^d(1 - s_{min}(\nu)r^2)$$

Lemma 4.2.6 ((79)). *Let (M, g) be a compact Riemannian manifold of dimension d. Then there is some $\nu > 0$ and $r_\nu > 0$ such that for all $r < r_\nu$ and any two point with $dist(p_1, p_2) < 2r$ we have:*

$$\left(B^E_{(1-\nu r)r}(p_1) \cup B^E_{(1-\nu r)r}(p_2)\right) \subset (B_r(p_1) \cup B_r(p_2)) \subset \left(B^E_{(1+\nu r)r}(p_1) \cup B^E_{(1+\nu r)r}(p_2)\right)$$

Let us now consider how we must adapt these approximations for balls intersecting the boundary. We may consider our manifold to be embedded in $\mathbb{R}^D$, for some D. Indeed, Nash's Imbedding Theorem (99) guarantees existence of an isometric embedding.

Definition 4.2.7 (Reach of a Manifold). *Let the medial axis $Med(M)$ of a manifold M embedded in $\mathbb{R}^D$, be the set of points in $\mathbb{R}^D$ which do not have a unique nearest element in M. The reach τ_M of a manifold is defined to be $\tau_M = \inf_{p\in Med(M)} dist(M, p)$.*

Theorem 4.2.8 ((100, Theorem 4.18)). *Let M be a submanifold of $\mathbb{R}^D$ and τ_M the reach of M. Then the reach is realised as:*

$$\tau_M = \inf_{q\neq p\in M} \frac{\|q - p\|^2}{2dist(q, T_pM)},$$

where $||.||$ denotes the standard Euclidean norm in $\mathbb{R}^D$. See Figure 4.1.

The manifolds we consider are smooth and compact and so in particular the reach of our manifolds is non-zero. We shall frequently consider the double manifold $DM = M \cup_{\partial M} M'$ where M' is a copy of M, and the two copies of M are glued along their common boundary. This gives us a canonical compact closed manifold in which M is embedded. The following result provides an estimate for the volume of a ball centred near the boundary.

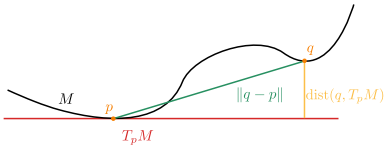


Figure 4.1: We sketch the distances involved in Theorem 4.2.8 for which we can calculate the reach of a manifold embedded in Euclidean space.

Proposition 4.2.9. *Let $B_r^{DM}(c)$ be a ball of radius r centred at $c \in M$ in the double manifold DM, and let δ denote the distance from the centre c to the boundary. Let $B_r^M(c) = B_r^{DM}(c) \cap M$ denote the portion of this ball contained in M. Then the volume of $B_r^M(c)$ can be expressed as:*

$$Vol(B_r^M(c)) = (\frac{1}{2} + \alpha_1 + \alpha_2) Vol(B_r^{DM}(c))$$

where $\alpha_1 = O(\frac{\delta}{r})$, $\alpha_1 \geq 0$ and $\alpha_2 = O(r)$.

Proof. Let us assume that the Riemannian metric is Euclidean, since we have seen above that this will only change the volume of our ball by a factor of order $(1 + O(r^2))$. Let p denote the closest point on the boundary to c. We shall first bound the volume of the ball $B_r^{DM}(c)$ capped by the tangent plane T_pM, so that we are dealing with a Euclidean ball from which we have removed a hemispherical cap. The volume of a d-dimensional ball of radius r capped at height h, B_r^h, can be expressed as a fraction of the volume of the full ball $\omega_d r^d$:

$$\text{Vol}(B_r^h) = \omega_d r^d \frac{1}{2} \left(1 + \frac{G_d(q)}{G_d(1)}\right) =: \omega_d r^d (\frac{1}{2} + \alpha_1),$$

where $G_d(u) = \int_0^u (1 - t^2)^{(d-1)/2} dt$, $q = \frac{\delta}{r}$, and $\delta = r - h$. For sufficiently small q we have the trivial bounds $\frac{1}{2}q \leq G_d(q) \leq q$, and so we see that:

$$\alpha_1 = O\left(\frac{\delta}{r}\right) \text{ and } \alpha_1 \geq 0.$$

Let us now estimate the error volume bounded between the boundary and T_pM. For all $q \in M$ Theorem 4.2.8 implies that $\text{dist}(q, T_pM) \leq \|q - p\|^2 / \tau_M$. Hence we can bound the error volume between T_pM and a parallel hyperplane at distance $O(r^2)$ from T_pM. Thus the error volume is bounded by $\text{Vol}(B_r^{(d-1)}) \cdot r^2 = O(r^{d+1})$ (where $B_r^{(d-1)}$ denotes the radius r, $(d-1)$-dimensional ball). We attain the desired result that:

$$\text{Vol}(B_r^M(c)) = (\frac{1}{2} + \alpha_1 + \alpha_2) \text{Vol}(B_r^{DM}(c))$$

where $\alpha_1 = O(\frac{\delta}{r})$, $\alpha_1 \geq 0$ and $\alpha_2 = O(r)$. □

We can combine this result with Lemma 4.2.6 to similarly estimate the volume of intersecting balls lying close to the boundary.

If M is a smooth manifold with non-empty boundary then we call a neighbourhood of the boundary ∂M a *collar neighbourhood* if it is the image of a smooth embedding $\partial M \times [0, 1) \to M$ and the embedding restricts to the identification $\partial M \times \{0\} \to \partial M$. A standard result (sometimes called the Collar Neighbourhood Theorem (98)) guarantees any smooth manifold with non-empty boundary has a collar

neighbourhood. For small r the r-neighbourhood of the boundary denoted ∂M_r, is a collar neighbourhood. It is straight forward to show that the volume of this collar neighbourhood $\mathrm{Vol}(\partial M_r) \sim r\mathrm{Vol}(\partial M)$.

4.3 Morse Theory

4.3.1 Morse Theory on Manifolds with Boundary

We start with a terse summary of Morse theory for manifolds with boundary. The upshot of this section is that given a suitable smooth function on the manifold we define the Morse complex in the same manner as in (101). This complex will compute relative homology $H_*(M, \partial M)$ if the function attains a minimum on the boundary, and will compute absolute homology $H_*(M)$ if the function attains a maximum on the boundary.

Definition 4.3.1 (Non-Degenerate Critical Point (101))**.** *Let M be a smooth d-manifold with boundary and $f : M \to \mathbb{R}$ a smooth function. A point $p \in M$ is critical if $\nabla f(p) = 0$. A critical point is non-degenerate if the Hessian $H(p)$ is non-singular.*

Definition 4.3.2 (Smooth Manifold Triad (101))**.** *We say $(M; V_0, V_1)$ is a smooth manifold triad if M is a smooth manifold with boundary, and the boundary ∂M is the disjoint union of the open and closed submanifolds[6] V_0, V_1.*

[6] V_0, V_1 are open and closed when viewed as submanifolds of ∂M

Definition 4.3.3 (Morse Function (101))**.** *A smooth function $f : (M; V_0, V_1) \to [a, b]$ is a Morse function on a smooth manifold triad if $f^{-1}(a) = V_0, f^{-1}(b) = V_1$ and all the critical points of f lie in the interior of M, ($int(M) = M \setminus \partial M$).*

In this setting we still have the classical Morse Lemma (102) for closed manifolds that asserts the existence of a coordinate system about each critical point for which the Morse function has a diagonal quadratic form. The index of a critical point is again given by the dimension of the negative eigenspace of the Hessian matrix $H(p)$.

Theorem 4.3.4 ((101, Theorem 7.4))**.** *There is a chain complex C_*, where C_k is the free abelian group on the set of index k critical points of f, and moreover $H_k(C_*) \cong H_k(M, V_0)$.*

In particular we will be interested in the case where we decompose the boundary trivially and so recover either $H_k(M)$ or $H_k(M, \partial M)$. The use of Morse theory in our arguments will not rely on any knowledge of the boundary maps, and will solely be used to bound Betti numbers.

4.3.2 *Morse Theory for Min Type Functions*

Whilst regular Morse theory is concerned with smooth functions, the distance function associated to a point cloud on a Riemannian manifold is generally not smooth. However, the square of such a distance function is a min-type function for which a version of Morse theory has been developed (93).

Definition 4.3.5 (Min-type function (93))**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a germ of a continuous function at $p \in \mathbb{R}^d$. Then f is a germ of a min-type function at p if there exist germs of smooth functions α_i at p such that locally around p we have that $f = \min_{i=1}^m \alpha_i$. A function on a d-manifold $f : M \to \mathbb{R}$ is min-type if for all $p \in M$ the germ at p is a germ of a min-type function.*

A couple of technical Lemmas allow us to show that each germ of a min-type function has an essentially unique minimal representation. We use this canonical minimal representation to build the min-type version of Morse Theory.

Definition 4.3.6 (Non-degenerate regular point (93))**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a germ of a min-type function. The point $p \in \mathbb{R}^d$ is a non-degenerate regular (NDR) point if f has a minimal representation at p, $f = \min_{i=1}^m \alpha_i$, such that:*

1. $\{\nabla(\alpha_i - \alpha_{i+1})(p)\}_{i=1}^{m-1}$ *is linearly independent*
2. $0 \notin \mathrm{Grad}(f) := Conv\{\nabla\alpha_i(p)\}_{i=1}^m$ *where Conv denotes the convex hull*
3. $f|_{\mathcal{G}_f}$ *is a germ of a Morse function on the boundary set* $\mathcal{G}_f := \{x : \alpha_1(x) = ... = \alpha_m(x)\}$
4. *Any $m-1$ gradients among $\{\nabla\alpha_i(p)\}_{i=1}^m$ are linearly independent*

Condition 1 ensures that $\mathcal{G}_f$ is a smooth $(d - m + 1)$-submanifold. If the gradients $\{\nabla\alpha_i(p)\}_{i=1}^m$ are linearly dependent then the convex hull $\mathrm{Conv}\{\nabla\alpha_i(p)\}_{i=1}^m = \mathrm{Grad}(f)$ can be thought of as an $(m-1)$-simplex in the tangent space.

Definition 4.3.7 (Non-degenerate critical point (93))**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a germ of a min-type function. The point $p \in \mathbb{R}^d$ is a non-degenerate critical (NDC) point if f has a minimal representation at p, $f = \min_{i=1}^m \alpha_i$, such that:*

1. $\{\nabla(\alpha_i - \alpha_{i+1})(p)\}_{i=1}^{m-1}$ *is linearly independent*
2. $0 \in \mathrm{Grad}(f) = Conv\{\nabla\alpha_i(p)\}_{i=1}^m$
3. $f|_{\mathcal{G}_f}$ *is a germ of a Morse function on the boundary set* $\mathcal{G}_f = \{x : \alpha_1(x) = ... = \alpha_m(x)\}$

4. *Any $m-1$ gradients among $\{\nabla\alpha_i(p)\}_{i=1}^m$ are linearly independent*

Definition 4.3.8 (Morse min-type function)**.** *We say that a min-type function $f : M \to \mathbb{R}$ is a Morse min-type function if every point is either an NDR or an NDC point.*

If p is an NDC point of $f : M \to \mathbb{R}$ then p is an NDC point of the smooth Morse function $f|_{\mathcal{G}_f}$ in the usual Morse theoretic sense. Consequently we define the index of an NDC point as follows:

Definition 4.3.9 (Index (93))**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Morse min-type function with NDC point p and associated minimal representation $f = \min_{i=1}^m \alpha_i$. The index of p is defined to be $Ind_p(f) = (m-1) + Ind_p(f|_{\mathcal{G}_f})$*

One of the main results from (93) relates Morse min-type functions to smooth Morse functions. In essence the following Theorem says that a Morse min-type function can be ε-approximated by a classical smooth Morse function with the same number of critical points of the same index, and these points are arbitrarily close to the original critical points.

Theorem 4.3.10 (Morse Min-Type Approximation Theorem (93))**.** *Let f be a Morse min-type function on a compact closed Riemannian manifold (M, g) with critical points $y_1, ..., y_m$ and corresponding indices $q_1, ..., q_m$. For any $\varepsilon > 0$ there is a smooth Morse function f_ε satisfying the following properties:*

1. *f_ε is an ε-approximation of f in the C^0- metric*
2. *f_ε is an ε-approximation of f in the C^2- metric for any neighbourhood where f is smooth*
3. *f_ε has the same number of critical points $y_1^\varepsilon, ..., y_m^\varepsilon$ with $Ind_{y_i}(f) = Ind_{y_i^\varepsilon}(f_\varepsilon)$*
4. *$\rho(y_i, y_i^\varepsilon) \leq \varepsilon$*

4.3.3 *Morse Theory for the Distance Function on a Compact Manifold with Boundary*

In this section we seek to recover a Morse function on the triad $(M, \emptyset, \partial M)$ induced by the distance function of a point cloud on M. For $x, y \in M$ and $\mathcal{P}$ a finite subset of M define distance functions

$$\rho_x^2(y) := \rho^2(x, y) \quad \text{and} \quad \rho_{\mathcal{P}}^2(y) := \min_{x\in\mathcal{P}} \rho_x^2(y).$$

Lemma 4.3.11 ((79))**.** *Let (M, g) be a compact Riemannian manifold with boundary, then there exists a positive real number $r_{mt} > 0$ such that for every finite sample of points $\mathcal{P} \subset M \setminus \partial M$ the distance function $\rho_{\mathcal{P}}^2$ is a Morse min-type function on the r_{mt} neighbourhood of these points $B_{r_{mt}}(\mathcal{P})$.*

Proof. Given any point $x \in M$ the distance function ρ_x^2 is smooth, Morse, and strictly convex on some neighbourhood $B_{r_x}(x)$. Since our metric g is smooth, r_x may be chosen continuously. By compactness there is some positive $r_{\text{mt}} \leq r_x$ for all $x \in M$. □

Proposition 4.3.12. *Let (M, g) be a compact Riemannian manifold with boundary and let $\mathcal{P}$ be a finite subset of M such that $M \subset B_r(\mathcal{P})$ for some $r < \frac{r_{mt}}{2}$. Then the critical points of the Morse min-type distance function $\rho^2_{\mathcal{P}}$ are in one to one index preserving correspondence with the critical points of a Morse function on the smooth manifold triad $(M; \emptyset, \partial M)$.*

Proof. Consider the double $DM = M \cup_{\partial M} M'$ of the manifold M. Let $\mathcal{P}' \subset M'$ be a finite set of points in M' such that:

$$\operatorname{dist}(p, \partial M') > \frac{r_{\text{mt}}}{2} \text{ for all } p \in \mathcal{P}' \text{ and } B_{r_{\text{mt}}}(\mathcal{P} \cup \mathcal{P}') = DM.$$

Then the distance function $\rho^2_{\mathcal{P}\cup\mathcal{P}'}$ is of Morse min-type on all of DM and hence, by Theorem 4.3.10, has an ε-approximation by a smooth Morse function f on DM.

Let $h : [0, 1) \times \partial M \to M$ be a diffeomorphism onto a collar neighbourhood of ∂M. Since $\mathcal{P}$ is finite there is some $\delta \in (0, 1)$ such that $h([0, \delta] \times \partial M) \cap \mathcal{P} = \emptyset$. Let us define $M_\delta := M \setminus h([0, \delta) \times \partial M)$ and $\partial M_\delta = h(\{\delta\} \times \partial M)$.

As $\mathcal{P}'$ is bounded away from the common boundary $\partial M = \partial M'$ by $\frac{r_{\text{mt}}}{2}$ and $M \subset B_r(\mathcal{P})$ with $r < \frac{r_{\text{mt}}}{2}$, the distance function $\rho^2_{\mathcal{P}\cup\mathcal{P}'}$, and thus also f, increases on ∂M_δ in the direction of the normal pointing into M'. Hence $f|_{M_\delta}$ can be extended to a smooth Morse function F on M, such that F attains its maximum on the boundary ∂M and has no critical points on the collar $h([0, \delta) \times \partial M)$. F is a smooth Morse function on the smooth manifold triad $(M; \emptyset, \partial M)$ with critical points in one to one correspondence with the critical points of $\rho^2_{\mathcal{P}}$ as required. □

Corollary 4.3.13. *Let $\mathcal{P}_n$ be a Poisson process on M a compact Riemannian manifold with boundary. Then with high probability the distance function $\rho^2_{\mathcal{P}_n}$ induces a Morse function on the smooth manifold triad $(M; \emptyset, \partial M)$. That is to say $\mathbb{P}(\{\rho^2_{\mathcal{P}_n} \text{ induces a Morse function}\}) \to 1$ as $n \to \infty$.*

Proof. Observe that for a Poisson process $\mathcal{P}_n$ on M and fixed $r < \frac{r_{\text{mt}}}{2}$, $\mathbb{P}(M \subset B_r(\mathcal{P}_n)) \to 1$ as $n \to \infty$. Hence using Proposition 4.3.12 the point process $\mathcal{P}_n$ induces a Morse min-type distance function $\rho^2_{\mathcal{P}_n}$ whose critical points recover the homology of M with high probability. □

Let us consider the conditions for which a point of M is a k-critical point or non-degenerate regular point of the distance function induced by a finite set $\mathcal{P} \subset M$.

For $y \in \mathcal{P}$ the local behaviour of $\rho^2_{\mathcal{P}}$ close to y coincides with $\rho^2_y(x) = \rho^2(x,y)$. Thus the minimal representation of $\rho^2_{\mathcal{P}}$ at y is $\rho^2(\cdot, y)$, the submanifold $\mathcal{G}_f$ is just M, and $\nabla\rho^2_y(y) = 0$. So y is an NDC point of index 0.

For any $p \in M \setminus \mathcal{P}$, if there is a point $y \in \mathcal{P}$ strictly closer to p than any other point in $\mathcal{P}$, then locally about p we have that $\rho^2_{\mathcal{P}}(x) = \rho^2_y(x)$ and $\nabla\rho^2_y(p) \neq 0$. So p is an NDR point.

Let $p \in M$ be such a point with $\min_{y\in\mathcal{P}} \rho(p,y)$ achieved precisely at all $y \in \mathcal{Y} = \{y_0, ..., y_k\} \subset \mathcal{P}$. Then locally about p we have the minimal representation $\rho^2_{\mathcal{P}}(x) = \min_{i=0}^k \rho^2_{y_i}(x)$. In order for p to be critical we require linear independence of the set $\{\nabla(\rho^2_{y_i} - \rho^2_{y_{i+1}})(p)\}_{i=0}^{k-1}$ which corresponds to saying that the set $\{y_0, ..., y_k\}$ is generic. Further we require that at p we have $0 \in \mathrm{Grad}(f)$. This corresponds to p lying in $\exp_p(\mathrm{Conv}(\{\exp_p^{-1}(y_0), ..., \exp_p^{-1}(y_k)\}))$, which can be thought of as the convex hull of the set $\{y_0, ..., y_k\}$. If these conditions are met, p is critical and the index of such a critical point will be k, as the point p attains a minimum of the distance function restricted to the submanifold $\mathcal{G}_f$ at p.

In terms of the Čech complex construction built on $\mathcal{P}$, an index k NDC point with critical value r occurs at the point $x \in M$ if x is the point of intersection of closed radius r balls about $k+1$ points of $\mathcal{P}$, and lies in their convex hull. We can thus identify index k critical points of the distance function with the introduction of k simplices to the Čech complex at their respective critical values.

Let $\mathcal{Y} = \{y_0, ..., y_k\} \subset M$ and define the $\mathcal{Y}$-equidistant sets:

$$E(\mathcal{Y}) := \{x \in M \,|\, \rho_{y_0}(x) = ... = \rho_{y_k}(x)\} \quad , \quad E_r(\mathcal{Y}) := E(\mathcal{Y}) \cap B_r(\mathcal{Y}).$$

The following result is a generalisation to Riemannian manifolds of the fact that $k+1$ generic points in Euclidean space lie on a sphere of dimension $k-1$. In particular we can associate a centre and radius to a collection of points which are sufficiently close together.

Lemma 4.3.14 ((79)). *There is a positive $r_{max} < r_{mt}$ such that if $\mathcal{Y}$ is generic with $E_{r_{max}}(\mathcal{Y}) \neq \emptyset$, then the set $\mathcal{Y}$ has a unique point $c(\mathcal{Y}) \in M$ such that for all $p \in \mathcal{Y}$*

$$\rho_p(c(\mathcal{Y})) = \inf_{x\in E(\mathcal{Y})} \rho_{\mathcal{Y}}(x).$$

In this case, we will refer to the point $c(\mathcal{Y})$ as the *centre* of $\mathcal{Y}$ and to $\rho(\mathcal{Y}) := \rho_{\mathcal{Y}}(c(\mathcal{Y}))$ as its *radius*. The set $\mathcal{Y} = \{y_0, ..., y_k\}$ corresponds to an index k critical point of the distance function $\rho^2_{\mathcal{P}}$ with critical value $\rho^2(\mathcal{Y})$.

4.4 *Asymptotic Coverage*

In forming the upper threshold we need to understand the asymptotic coverage of a Riemannian manifold with boundary. We would like to understand the conditions under which $B_r(\mathcal{P})$ covers M w.h.p.

The paper Random Coverings (103) gives a comprehensive treatment of the asymptotic behaviour of the number of randomly chosen radius r balls required to cover a compact *closed* Riemannian manifold m times. The result is translated into a sharp coverage threshold in (78), but again this result applies only to *closed* Riemannian manifolds. We use the coverage result of (78) to derive conditions (not necessarily sharp) for asymptotic coverage of a compact Riemannian manifold with boundary. A sharp coverage result for manifolds with boundary found in (104) has come to our attention. Nevertheless, we include our coverage result which is sufficiently strong for our purposes and arises from a natural geometric argument.

Theorem 4.4.1 (Sharp Coverage Threshold (78))**.** *Let M be a compact, closed, finite volume Riemannian manifold. Let $\mathcal{P}_n$ be a uniform Poisson process on M of intensity n. Let $B_r(\mathcal{P}_n)$ denote the r neighbourhood of the points $\mathcal{P}_n$ and $w(n) \to \infty$, then we yield:*

$$\lim_{n\to\infty} \mathbb{P}(M \subset B_r(\mathcal{P}_n)) = \begin{cases} 1 & \Lambda = \log n + (d-1)\log\log n + w(n), \\ 0 & \Lambda = \log n + (d-1)\log\log n - w(n). \end{cases}$$

Our coverage result below is not proposed as a sharp threshold but merely a threshold that suits our needs in subsequent proofs.

Corollary 4.4.2 (Coverage of Compact Manifold with Boundary)**.** *Let M be a compact Riemannian manifold with boundary and let $\mathcal{Q}_n$ be a uniform Poisson process on M with intensity n. If $\Lambda = \log n + (d-1)\log\log n + w(n)$ with $w(n) \to \infty$, then the $2r$ neighbourhood of the Poisson process covers M w.h.p:*

$$\lim_{n\to\infty} \mathbb{P}(M \subset B_{2r}(\mathcal{Q}_n)) = 1.$$

Proof. Our proof will be an application of the sharp threshold developed in (78). Let us form the double manifold $DM = M \cup_{\partial M} M'$, which has double the volume of M. Now consider a uniform Poisson process $\mathcal{P}_n$ on DM with intensity n. This restricts to a uniform Poisson process $\mathcal{Q}_n = \mathcal{P}_n \cap M$ on M with intensity n.

Then for $\Lambda = \log n + (d-1)\log\log n + w(n)$ Theorem 4.4.1 implies that

$$\lim_{n\to\infty} \mathbb{P}(DM \subset B_r(\mathcal{P}_{Cn})) = 1$$

Let us define the following notation: $M_r = M \setminus (\partial M \times [0, r)) = \{x \in M : \min_{y \in \partial M} \rho_M(x, y) \leq r\}$, and the r-neighbourhoods:

$$B_r^{(M)}(\mathcal{P}) = \{x \in M : \min_{p \in \mathcal{P}} \rho_M(x, p) < r\} \text{ and } B_r^{(DM)}(\mathcal{P}) = \{x \in DM : \min_{p \in \mathcal{P}} \rho_{DM}(x, p) < r\}.$$

For any point $x \in M_r$ the r-neighbourhood of $x \in DM$ coincides with the r-neighbourhood of $x \in M$, $B_r^{(M)}(\{x\}) = B_r^{(DM)}(\{x\})$. For points in the collar $\partial M \times [0, r)$ we cannot make the same statement, however the result follows immediately from the observation that:

$$DM \subset B_r^{(DM)}(\mathcal{P}_n) \implies M_r \subset B_r^{(DM)}(\mathcal{P}_n \cap M) \implies M_r \subset B_r^{(M)}(\mathcal{Q}_n) \implies M \subset B_{2r}^{(M)}(\mathcal{Q}_n).$$

The last implication follows since for any $x \in M$ there is a $y \in M_r$ and $q \in \mathcal{Q}_n$ such that $\rho_M(x, y) < r$ and $y \in B_r^{(M)}(q)$. Hence $x \in B_{2r}^{(M)}(q)$. □

4.5 *Blaschke-Petkantschin Formulae*

A key component of our later arguments and the arguments found in (79) is an integral formula which facilitates calculating bounds on the expected number of critical points of a distance function associated to a Poisson process on a Riemannian manifold. In this section we explain this change of variables formula and make appropriate adaptations to the case for Riemannian manifolds with boundary.

4.5.1 *The Blaschke-Petkantschin formula in the Euclidean case*

We first recall a derivation of the classical Blaschke-Petkantschin formula in the Euclidean case. Our derivation and Proposition 4.5.1, roughly follow Sections 2 and 3 of (105).

Let E_d be a d-dimensional Euclidean space, and let $(e_i)_{i=1}^d$ be an orthonormal moving frame in E_d, where for an infinitesimal rotation of the frame,

$$e_i \cdot de_i = 0, \forall i \in [d];$$

and set

$$\omega_{ij} := e_i \cdot de_j = -\omega_{ji}, \forall i, j \in [d].$$

Given r points $\{x_i : i \in [r]\} \subset E_d$, Miles derives heuristically the associated volume form

$$\bigwedge_{j=1}^{r} dV(x_j) = \bigwedge_{i=1}^{d} \bigwedge_{j=1}^{r} e_i . dx_j.$$

Furthermore, given the Grassmannian manifold $\mathrm{Gr}(r, d)$ with invariant measure $d\mu_{r,d}(V)$, Miles also derives

$$d\mu_{r,d} = \bigwedge_{i=1}^{r} \bigwedge_{j=r+1}^{d} \omega_{ij}.$$

Using the above, the Blaschke-Petkantschin formula expresses the Euclidean volume form $dV(x_i^d)$ on $\{x_i : i \in [r]\}$ in terms of the volume element associated to the r-plane containing $\{x_i : i \in [r]\}$, denoted by $dV(x_i^r)$.

Proposition 4.5.1 (Blaschke-Petkantschin Formula Euclidean Case ((105))). *Let $\{x_i \mid i \in [r]\}$ be a linearly independent set of vectors spanning $V = Span(\{e_i \mid i \in [r]\}) \in \mathrm{Gr}(r,d)$. For each $j \in [r]$, let $(\lambda_{jk})_{k\in[r]} \in E^r$ be such that*

$$x_j = \sum_{k=1}^{r} \lambda_{jk} e_k, \text{ and let } \Upsilon := \left|\det(\lambda_{jk})\right| > 0.$$

Then

$$\bigwedge_{i=1}^{r} dV(x_i^d) = \Upsilon^{d-r} d\mu_{r,d} \bigwedge_{i=1}^{r} dV(x_i^r).$$

Proof. Given $j \in [r]$ and $i \in \{r+1, \ldots, n\}$, we have

$$\begin{aligned}
dx_j &= \sum_{k=1}^{r} (d\lambda_{jk} e_k + \lambda_{jk} de_k) \\
e_i.dx_j &= \sum_{k=1}^{r} \lambda_{jk} \omega_{ik} \\
\bigwedge_{j=1}^{r} e_i.dx_j &= \left|\det(\lambda_{jk})\right| \bigwedge_{k=1}^{r} \omega_{ik} \\
\bigwedge_{i=r+1}^{d} \bigwedge_{j=1}^{r} e_i.dx_j &= \Upsilon^{d-r} \bigwedge_{i=r+1}^{d} \bigwedge_{k=1}^{r} \omega_{ik} \\
\bigwedge_{i=r+1}^{d} \bigwedge_{j=1}^{r} e_i.dx_j &= \Upsilon^{d-r} d\mu_{r,d},
\end{aligned}$$

and multiplying on both sides above by $\bigwedge_{i=1}^{r} \bigwedge_{j=1}^{r} e_i.dx_j = \bigwedge_{j=1}^{r} dV(x_j^r)$, we obtain the desired result. □

4.5.2 *Blaschke-Petkantschin formula for Riemannian Manifolds*

Following (79), we obtain a Riemannian generalization of the Blaschke-Petkantschin formula. The formula is valid for functions with support close to the diagonal of M^{k+1}. It enables us to reparametrise a $(k+1)$-tuple of points near the diagonal of M^{k+1} into local coordinates about their centre. We shall use bold face notation for (k+1)-tuples $\mathbf{y} = (y_0, ..., y_k) \in M^{k+1}$ and as before denote the centre and radius of the corresponding collection of points $\mathcal{Y} = \{y_0, ..., y_k\}$ as $c(\mathbf{y})$ and $\rho(\mathbf{y})$ respectively. The change of variables has the following form:

$$\begin{aligned}
M^{k+1} &\longleftrightarrow M \times \mathbb{R} \times \mathrm{Gr}(k,d) \times (S^{(k-1)})^{k+1} \\
\mathbf{y} &\longleftrightarrow (c(\mathbf{y}), u, V, \mathbf{w})
\end{aligned}$$

A $(k+1)$-tuple of points in $\mathbf{y} \in M^{k+1}$, is reparametrised by the centre of this tuple $c(\mathbf{y})$, the distance of the points from their centre u, the k-plane in which the pre-image of the points lie in the tangent space at the centre V, and the $k+1$ points of the $(k-1)$-sphere upon which they lie $\mathbf{w}$. This change of variables enables us to compute bounds for the number of critical points, since it separates variables and decomposes the integral over M^{k+1} into components which can be estimated independently.

Suppose that $M \subset \mathbb{R}^d$ is a closed Riemannian manifold and let $\mathbf{y} = (y_i)_{i=1}^{k+1} \in M^{k+1}$, with centre $c = c(\mathbf{y})$, with radius $\rho(\mathbf{y}) \leq r$, and local normal coordinates $(x^1, \ldots, x^d)$. For sufficiently small r, we can write for all $i \in [k+1]$

$$y_i = \exp_c(v_i),$$

with

$$v_i = \sum_{j=1}^{d} x^j(y_i) \left(\frac{\partial}{\partial x^i}\right)_c.$$

Let $\mathbf{1}_r(\mathbf{y}) = \mathbf{1}\{E_{r_{\max}}(y_0, ..., y_k) \neq \emptyset \text{ and } \rho(\mathbf{y}) \leq r\}$. Note that this indicator function has support near to the diagonal of M^{k+1} and each tuple in the support of this function has a unique centre. It is shown in (79), that $\{v_i : i \in [k+1]\}$ have linear dependency and span a k-dimensional subspace $V \subset T_{c(\mathbf{y})}M$ when $c(\mathbf{y})$ is a critical point. We yield the following change of variable formula.

Lemma 4.5.2 ((79)). *Let M be a compact closed Riemannian manifold with $M' \subset M$ a submanifold with or without boundary. Let $r_{\max}$ be as in Lemma 4.3.14, and $r < r_{\max}$. Then there exists an invariant measure $d\mu_{k,d}(V)$ on $\mathrm{Gr}(k, T_cM) = \mathrm{Gr}(k,d)$, such that for every $f \in C^\infty(M^{k+1}; \mathbb{R})$*

$$\int_{M^{k+1}} f(\mathbf{y})\mathbf{1}_r(\mathbf{y})\mathbf{1}\{c(\mathbf{y}) \in M'\} \left| \mathrm{dvol}_g(\mathbf{y}) \right| = \int_{M'} \left| \mathrm{dvol}_g(c) \right| \int_0^r du u^{dk-1} \int_{\mathrm{Gr}(k,T_cM)} d\mu_{k,d}(V) \times \left(\int_{(S_1)^{k+1}} \Upsilon_1^{d-k}(w) f(\exp_c(uw)) \prod_{i=1}^{k+1} \sqrt{\det(g_{\exp_c(uw_i)})} \left| \mathrm{dvol}_{S_1(V)}(w_i) \right| \right).$$

Proof. First note that if M' has positive codimension then both expressions are zero, so assume M' has zero codimension. If $\mathbf{y} \in M^{k+1}$ is such that $E_{r_{\max}}(\mathbf{y}) \neq \emptyset$, then the induced centre $c(\mathbf{y})$ is uniquely defined (Lemma 4.3.14); hence:

$$\{\mathbf{y} \in M^{k+1} \mid c(\mathbf{y}) \in M' \text{ and } E_{r_{\max}}(\mathbf{y}) \neq \emptyset\} = \bigcup_{c \in M'} \mathcal{Y}(c),$$

where $\mathcal{Y}(c) := \{\mathbf{y} \in M^{k+1} \mid E_{r_{\max}}(\mathbf{y}) \neq \emptyset \text{ and } c(\mathbf{y}) = c\}$, and this union is disjoint (by uniqueness of the centre).

Thus

$$\int_{M^{k+1}} f(\mathbf{y})\mathbf{1}_r(\mathbf{y})\mathbf{1}\{c(\mathbf{y}) \in M'\}\Big|\operatorname{dvol}_g(\mathbf{y})\Big| = \int_{M'}\Big|\operatorname{dvol}_g(c)\Big|\int_{\mathcal{Y}(c)} f(\mathbf{y})\mathbf{1}\{\rho(\mathcal{Y}) \le r\}\Big|\operatorname{dvol}_g(\mathbf{y})\Big|.$$

Now fix $c \in M'$ with local normal coordinates $(x^1, \ldots, x^d)$; for $\mathbf{y} \in \mathcal{Y}(c)$ with $\rho(\mathbf{y}) \le r < r_{\max}$ (this last condition ensures that y_i can be written as $\exp_c(v_i)$, $v_i \in T_cM$ and $|v_i| = u \le r$, for all $i \in [k+1]$), we find:

$$\begin{aligned}
\Big|\operatorname{dvol}_g(\mathbf{y})\Big| &= \left|\bigwedge_{i=1}^{k+1}\operatorname{dvol}_g(y_i)\right| \\
&= \left|\bigwedge_{i=1}^{k+1}\sqrt{\left|\det(g_{y_i})\right|}dx^1(y_i)\wedge\cdots\wedge dx^d(y_i)\right| \\
&= \prod_{i=1}^{k+1}\sqrt{\left|\det(g_{y_i})\right|}\left|\bigwedge_{i=1}^{k+1} dx^1(y_i)\wedge\cdots\wedge dx^d(y_i)\right| \\
&= \prod_{i=1}^{k+1}\sqrt{\left|\det(g_{y_i})\right|}\left|\bigwedge_{i=1}^{k+1}\operatorname{dvol}_{g_{E_d}}(v_i)\right| \\
&= \prod_{i=1}^{k+1}\sqrt{\left|\det(g_{y_i})\right|}\Big|\operatorname{dvol}_{g_{E_d}}(\mathbf{v})\Big|,
\end{aligned}$$

Note that we have the polar decomposition:

$$\Big|\operatorname{dvol}_{g_{E_d}}(\mathbf{v})\Big| = du\Big|\operatorname{dvol}_{\mathbb{S}_u(E_d)(\mathbf{v})}\Big|,$$

and so by the Blaschke-Petkantschin formula, since $\{v_i : i \in [k+1]\}$ lies in a k-dimensional subspace $V \subset T_cM$, we have

$$\Big|\operatorname{dvol}_{\mathbb{S}_u(E_d)}(\mathbf{v})\Big| = \mathrm{Y}_u(\mathbf{v})^{d-k}d\mu_{k,d}(V)\Big|\operatorname{dvol}_{\mathbb{S}_u(V)}(\mathbf{v})\Big|,$$

hence, we deduce that

$$\Big|\operatorname{dvol}_g(\mathbf{y})\Big| = \prod_{i=1}^{k+1}\sqrt{\left|\det(g_{y_i})\right|}du\mathrm{Y}_u(\mathbf{v})^{d-k}d\mu_{k,d}(V)\Big|\operatorname{dvol}_{\mathbb{S}_u(V)}(\mathbf{v})\Big|.$$

This shows that for $c \in M'$

$$\begin{aligned}
&\int_{\mathcal{Y}(c)} f(\mathbf{y})\mathbf{1}\{\rho(\mathcal{Y}) \le r\}\Big|\operatorname{dvol}_g(\mathbf{y})\Big| = \\
&\qquad \int_0^r du u^{dk-1}\int_{\mathrm{Gr}(k,T_cM)} d\mu_{k,d}(V) \\
&\qquad \times\left(\int_{(\mathbb{S}_1)^{k+1}}\mathrm{Y}_1^{d-k}(\mathbf{w})f(\exp_c(u\mathbf{w}))\prod_{i=1}^{k+1}\sqrt{\det(g_{\exp_c(uw_i)})}\Big|\operatorname{dvol}_{\mathbb{S}_1(V)}(w_i)\Big|\right),
\end{aligned}$$

and thus the result follows. □

4.5.3 *Blaschke-Petkantschin formula for Compact Riemannian Manifolds with Non-Empty Boundary*

Using the change of variable formula established in Lemma 4.5.2 for compact closed Riemannian manifolds we attain a formula for Riemannian manifolds with non-empty boundary. We note that our formula is altered near the boundary since we must restrict our integral in the tangent space to those points whose image under the exponential map remain in the manifold.

Lemma 4.5.3. *Suppose that M is a compact Riemannian manifold with non-empty boundary, let DM be its double manifold and let $M' \subset M$ be a submanifold. Then we have:*

$$\begin{aligned}\int_{M^{k+1}} & f(\mathbf{y})\mathbf{1}_r(\mathbf{y})\mathbf{1}\{c(\mathbf{y}) \in M'\}|\operatorname{dvol}_g(\mathbf{y})| \\ &= \int_{c\in M'} |\operatorname{dvol}_g(c)| \int_0^r du \int_{\mathrm{Gr}(k,T_cDM)} d\mu_{k,d}(V) \\ &\times \int_{(\mathbb{S}_u(V)\cap \exp_c^{-1}(M))^{k+1}} \Upsilon_u^{d-k}(\mathbf{v}) f(\exp_c(\mathbf{v})) \prod_{i=1}^{k+1} \sqrt{|\det(g_{\exp_c(v_i)})|}\, |\operatorname{dvol}_{\mathbb{S}_u(V)^{k+1}}(v_i)|.\end{aligned}$$

Proof. We have

$$\begin{aligned}\int_{M^{k+1}} & f(\mathbf{y})\mathbf{1}_r(\mathbf{y})\mathbf{1}\{c(\mathbf{y}) \in M'\}|\operatorname{dvol}_g(\mathbf{y})| \\ &= \int_{DM^{k+1}} f(\mathbf{y})\mathbf{1}\{y \in M^{k+1}\}\mathbf{1}_r(\mathbf{y})\mathbf{1}\{c(\mathbf{y}) \in M'\}|\operatorname{dvol}_g(\mathbf{y})|;\end{aligned}$$

the double manifold DM is closed, hence applying the change of variables formula in Lemma 4.5.2 to the function

$$\mathbf{y} \longmapsto f(\mathbf{y})\mathbf{1}\{y \in M^{k+1}\},$$

we find

$$\begin{aligned}&\int_{DM^{k+1}} f(\mathbf{y})\mathbf{1}\{y \in M^{k+1}\}\mathbf{1}_r(\mathbf{y})\mathbf{1}\{c(\mathbf{y}) \in M'\}|\operatorname{dvol}_g(\mathbf{y})| = \\ &\int_{c\in M'} |\operatorname{dvol}_g(c)| \int_0^r du \int_{\mathrm{Gr}(k,T_cDM)} d\mu_{k,d}(V) \\ &\times \int_{\mathbb{S}_u(V)^{k+1}} \Upsilon_u^{d-k}(\mathbf{v}) f(\exp_c(\mathbf{v}))\mathbf{1}\{\exp_c(\mathbf{v}) \in M^{k+1}\} \prod_{i=1}^{k+1} \sqrt{|\det(g_{\exp_c(v_i)})|}\, |\operatorname{dvol}_{\mathbb{S}_u(V)^{k+1}}(v_i)|,\end{aligned}$$

which gives the desired result. □

4.6 *Upper Threshold*

In this section we produce an upper bound on the expected number of critical points induced by a Poisson process on our manifold. Similar to the argument presented in (79) we utilise an auxiliary radius

r_0 and count critical points with critical value in the range $[r, r_0)$. The auxiliary radius is chosen to be sufficiently large so that the r_0 neighbourhood of the Poisson process covers the manifold w.h.p., but is simultaneously sufficiently small so that the number of critical points with critical values in the range $[r, r_0)$ is asymptotically zero.

Let $\Lambda = \omega_d n r^d$ denote the expected number of points of a Poisson process with intensity n inside a d-dimensional ball of radius r. If we append a subscript to Λ this corresponds to changing the radius value in this expectation: $\Lambda_{r_0} = \omega_d n r_0^d$. Let $\beta_k(r)$ denote the k^{th} Betti number of the Čech complex associated to the submanifold $B_r(\mathcal{P}_n)$. We shall count critical points of the associated Morse min-type function treating points close to the boundary and far from the boundary separately.

Proposition 4.6.1 (Betti Number Upper Bound)**.** *Let M be a compact unit volume d-dimensional manifold with boundary. If $n \to \infty$ and $r, r_0 \to 0$ such that $\Lambda \to \infty$, $\Lambda_{r_0} r \to 0$, $\Lambda_{r_0} r_0^2 \to 0$ and $r_0 \geq r(\frac{\omega_d}{\kappa}(1 + |\log r|))^{1/d}$ (where $\kappa = \omega_d 2^{-(d+1)}(1 - s_{max} r_0^2)$), then for all $1 \leq k \leq d-1$:*

$$\mathbb{E}[\beta_k(r)] \leq \beta_k(M) + O(n\Lambda^k e^{-\Lambda} + n^{1-\frac{1}{d}} \Lambda^k e^{-\frac{1}{2}\Lambda})$$

To prove the above let us follow (79) in attempting to bound the number of k-critical points of the distance function associated to a Poisson process on M with critical value in the range $[r, r_0)$. Let us denote this set of critical points by $C_k^{\rho_M}(r, r_0)$.

Lemma 4.6.2. *If $n \to \infty$ and $r, r_0 \to 0$ such that $\Lambda \to \infty$, $\Lambda_{r_0} r \to 0$, $\Lambda_{r_0} r_0^2 \to 0$ and $r = o(r_0)$ then for all $1 \leq k \leq d$:*

$$\mathbb{E}[C_k^{\rho_M}(r, r_0)] = O(n^{1-\frac{1}{d}} \Lambda^{k-1} e^{-\frac{1}{2}\Lambda} + n\Lambda^{k-1} e^{-\Lambda})$$

The bounding term $n\Lambda^{k-1}e^{-\Lambda}$ arises from counting critical points at distance more than r_0 from the boundary of the manifold and the bounding term $n^{1-\frac{1}{d}}\Lambda^{k-1}e^{-\frac{1}{2}\Lambda}$ arises from counting critical points at distance less than r_0 from the boundary of the manifold.

In what follows let $f = \mathbf{1}\{x \in M\}$ and $\mathcal{P}_n$ denote a Poisson process of intensity nf on DM. Let us use the shorthand notation ∂M_{r_0} for an r_0 neighbourhood of the boundary, and $M_{r_0} = M \setminus \partial M_{r_0}$.

Let us define the following indicator functions which we use to count critical points:

- $h(\mathcal{Y}) = \mathbf{1}\{0 \in \text{Grad}(\mathcal{Y})\}$
- $h_{r,r_0}(\mathcal{Y}) = h(\mathcal{Y})\mathbf{1}\{r \leq \rho(\mathcal{Y}) < r_0\}$

- $g_{r,r_0}(\mathcal{Y}, \mathcal{P}_n) = h_{r,r_0}(\mathcal{Y})\mathbf{1}\{B(\mathcal{Y}) \cap \mathcal{P}_n = \emptyset\}$; where $B(\mathcal{Y})$ is the open ball $B_{\rho(\mathcal{Y})}(c(\mathcal{Y}))$
- $g^{M_{r_0}}_{r,r_0}(\mathcal{Y}, \mathcal{P}_n) = g_{r,r_0}(\mathcal{Y}, \mathcal{P}_n)\mathbf{1}\{c(\mathcal{Y}) \in M_{r_0}\}$
- $g^{\partial M_{r_0}}_{r,r_0}(\mathcal{Y}, \mathcal{P}_n) = g_{r,r_0}(\mathcal{Y}, \mathcal{P}_n)\mathbf{1}\{c(\mathcal{Y}) \in \partial M_{r_0}\}$

We can express the number of critical points as the following sum over generic $\mathcal{Y} \subset \mathcal{P}_n$:

$$|C_k^{\rho_M}(r, r_0)| = \sum_{\substack{\mathcal{Y}\subset\mathcal{P}_n \\ |\mathcal{Y}|=k+1}} |g^{M_{r_0}}_{r,r_0}(\mathcal{Y}, \mathcal{P}_n)| + |g^{\partial M_{r_0}}_{r,r_0}(\mathcal{Y}, \mathcal{P}_n)|$$

We can count the critical points in M_{r_0} and ∂M_{r_0} separately. Let us denote the critical points in ∂M_{r_0} with critical value in the interval $[r, r_0)$ as $C_{\partial M_{r_0}}$, and the critical points in M_{r_0} with critical value in the interval $[r, r_0)$ as $C_{M_{r_0}}$. The counting of the critical points $C_{M_{r_0}}$ is not affected by the presence of a non-trivial boundary and so we can use the analysis of (79) to bound the term $\mathbb{E}[|C_{M_{r_0}}|]$ by an order of $O(n\Lambda^{k-1}e^{-\Lambda})$.

We require new analysis to calculate an upper bound for the expected number of critical points $\mathbb{E}[|C_{\partial M_{r_0}}|]$. Applying Palm Theory (Theorem 4.2.4) to the uniform Poisson Process on M with intensity n we yield:

$$\begin{aligned}\mathbb{E}[|C_{\partial M_{r_0}}|] &= \mathbb{E}_{\mathcal{P}_n}[\sum_{\substack{\mathcal{Y}\subset\mathcal{P}_n \\ |\mathcal{Y}|=k+1}} g^{\partial M_{r_0}}_{r,r_0}(\mathcal{Y}, \mathcal{P}_n)] \\ &= \frac{n^{k+1}}{(k+1)!}\mathbb{E}_{\mathcal{P}_n,\mathcal{Y}'}[g^{\partial M_{r_0}}_{r,r_0}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)]\end{aligned}$$

Let us now condition on a given sample $\mathcal{Y}'$ of $k+1$ points of M with a view to then integrating over the whole double manifold. Recall $B(\mathcal{Y}')$ denotes the ball centred at $c(\mathcal{Y}')$ of radius $\rho(\mathcal{Y}')$ in DM, and that the number of points in a given subset of a uniform Poisson process has Poisson distribution with parameter proportional to the volume of the given subset.

$$\mathbb{E}[\mathbf{1}\{B(\mathcal{Y}') \cap \mathcal{P}_n = \emptyset\}|\mathcal{Y}'] = \mathbb{P}(\mathcal{P}_n(B(\mathcal{Y}')) = 0|\mathcal{Y}') = e^{-n\mathrm{Vol}(B(\mathcal{Y}')\cap M)}$$

$$\mathbb{E}[|C_{\partial M_{r_0}}|] = \frac{n^{k+1}}{(k+1)!}\int_{DM^{k+1}} \mathbf{1}\{c(\mathbf{y}) \in \partial M_{r_0}\}h_{r,r_0}(\mathbf{y})e^{-n\mathrm{Vol}(B(\mathbf{y})\cap M)}|\mathrm{dvol}_g(\mathbf{y})|$$

For sufficiently small r_0 we have the following lower bound for the volume of a ball with centre at distance $\delta(\mathbf{y})$ from the boundary via Proposition 4.2.9:

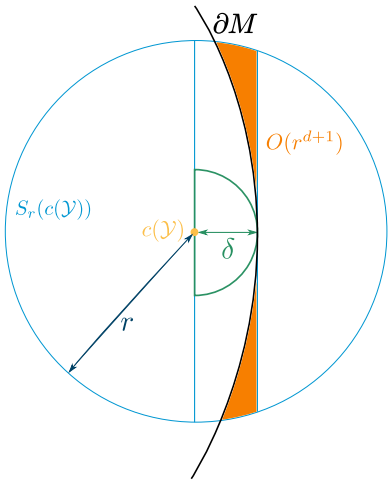


Figure 4.2: The contributions to our lower bound for $\mathrm{Vol}(B(\mathbf{y}) \cap M)$, are illustrated in this figure. Proposition 4.2.9 establishes that the shaded region has volume $O(r^{d+1})$. We note that for $\delta(\mathbf{y}) < \tau_M$ (τ_M the reach of the manifold) then the half-ball of radius $\delta(\mathbf{y})$ contributes $\frac{1}{2}\omega_d\delta(\mathbf{y})^d$ to the volume.

$$\mathrm{Vol}(B(\mathbf{y}) \cap M) \geq (\frac{1}{2} - O(r))\mathrm{Vol}(B(\mathbf{y})) + \frac{1}{2}\omega_d \delta(\mathbf{y})^d.$$

Note that $e^{nO(r)\mathrm{Vol}(B(\mathbf{y}))} = O(1)$ since $\Lambda_{r_0} r \to 0$, and so we have an upper bound:

$$\mathbb{E}[|C_{\partial M_{r_0}}|] \leq C\frac{n^{k+1}}{(k+1)!}\int_{DM^{k+1}} \mathbf{1}\{c(\mathbf{y}) \in \partial M_{r_0}\} h_{r,r_0}(\mathbf{y}) e^{-n\frac{1}{2}\left(\mathrm{Vol}(B(\mathbf{y}))+\omega_d\delta(\mathbf{y})^d\right)} |\mathrm{dvol}_g(\mathbf{y})|$$

Using the Blaschke-Petkantschin formula from Section 4.5 with $f(\mathbf{y}) = h_{r,r_0}(\mathbf{y}) e^{-n\frac{1}{2}\left(\mathrm{Vol}(B(\mathbf{y}))+\omega_d\delta(\mathbf{y})^d\right)}$, $c = c(\mathbf{y})$, $u = \rho(\mathbf{y})$ and $\mathbf{y} = \exp_c(u\mathbf{w})$ we attain the bound:

$$\begin{aligned}\mathbb{E}[|C_{\partial M_{r_0}}|] \leq \frac{n^{k+1}}{(k+1)!}\int_{\partial M_{r_0}} |\mathrm{dvol}_g(c)| \int_r^{r_0} du\ u^{dk-1} \int_{\mathrm{Gr}(k,T_cM)} d\mu_{k,d}(V) \\ \times \prod_{i=1}^{k}\left(\int_{\mathrm{S}_1(V)} \sqrt{|\det(g_{\exp_c(uw_i)})|}|\mathrm{dvol}_{\mathrm{S}_1(V)}(w_i)|\right) \mathbf{Y}_1^{d-k}(\mathbf{w}) f(\exp_c(u\mathbf{w}))\end{aligned}$$

Using compactness of the Grassmanian we can bound the final integral over $(\mathrm{S}_1(V))^k$ by some maximising subspace. The parallelogram volume $\mathbf{Y}_1^{d-k}(\mathbf{w})$ is taken over unit vectors and so bounded. Since our manifold is compact and g is smooth we may bound $\sqrt{|\det(g_{\exp_c(uw_i)})|}$ by some constant. (For details of the bounding constants see (79)). Thus we yield for some constant C:

$$\mathbb{E}[|C_{\partial M_{r_0}}|] \leq Cn^{k+1}\int_{\partial M_{r_0}} |\mathrm{dvol}_g(c)| \int_r^{r_0} du\ u^{dk-1} e^{-n\frac{1}{2}\left(\mathrm{Vol}(B_u(c))+\omega_d\delta(\mathbf{y})^d\right)}$$

Since $\Lambda_{r_0} r_0^2 \to 0$ we may bound the term $e^{-n\frac{1}{2}\left(\mathrm{Vol}(B_u(c))+\omega_d\delta(\mathbf{y})^d\right)}$ using the second order Taylor expansion of exp and Lemma 4.2.5 :

$$e^{-n\frac{1}{2}\left(\mathrm{Vol}(B_u(c))+\omega_d\delta(\mathbf{y})^d\right)} \leq e^{-n\frac{1}{2}\omega_d\left(u^d+\delta(\mathbf{y})^d\right)}\left(1 + s_{\max} n\omega_d r_0^{d+2}\right)$$

Let us apply a change of variables $s = \frac{u}{r}$ and additionally separate our integral to integrate over the distance of the critical point, $\delta = \delta(\mathbf{y})$, to the boundary. We make a change of variables using the diffeomorphism $\partial M_{r_0} \cong \partial M \times [0, r_0)$ given to us from the Collar Neighbourhood Theorem. Since our manifold is compact the Jacobian term introduced will be bounded by a constant which we absorb into the constant term C.

$$\begin{aligned}\mathbb{E}[|C_{\partial M_{r_0}}|] &\leq Cn^{k+1}(1 + s_{\max} n\omega_d r_0^{d+2}) \int_{\partial M_{r_0}} e^{-n\frac{1}{2}\omega_d\delta^d} |\mathrm{dvol}_g(c)| \int_1^{\frac{r_0}{r}} ds\ r^{dk} s^{dk-1} e^{-n\frac{1}{2}\omega_d r^d s^d} \\ &\leq Cn\Lambda^k(1 + s_{\max}\Lambda_{r_0} r_0^2) \int_0^{r_0} e^{-n\frac{1}{2}\omega_d\delta^d} d\delta \int_1^{\frac{r_0}{r}} ds\ s^{dk-1} e^{-\frac{1}{2}\Lambda s^d}\end{aligned}$$

We may bound the first integral by a term $O(n^{-\frac{1}{d}})$ by changing variables ($t = \frac{1}{2}\Lambda_\delta$) and observing that the integral takes the form of an upper incomplete gamma function:

$$\int_0^{r_0} e^{-n\frac{1}{2}\omega_d\delta^d} d\delta = Cn^{-\frac{1}{d}} \int_0^{\frac{1}{2}\Lambda_{r_0}} t^{\frac{1}{d}-1} e^{-t} dt \leq C\Gamma(\frac{1}{d}) n^{-\frac{1}{d}}$$

The last integral also has the form of an upper incomplete gamma function.

$$\Gamma(k,x) = \int_x^\infty t^{k-1}e^{-t}dt = (k-1)!e^{-x}\sum_{i=0}^{k-1}\frac{x^i}{i!}$$

Let $t = \frac{1}{2}\Lambda s^d$

$$\int_1^{\frac{r_0}{r}} \left(\frac{1}{2}\Lambda\right)^k s^{dk-1}e^{-\frac{1}{2}\Lambda s^d} ds = \frac{1}{d}\int_{\frac{1}{2}\Lambda}^{\frac{1}{2}\Lambda_{r_0}} t^{k-1}e^{-t}dt = \Gamma(k,\frac{1}{2}\Lambda) - \Gamma(k,\frac{1}{2}\Lambda_{r_0})$$

Absorbing surplus constants into the term C we attain:

$$\mathbb{E}[|C_{\partial M_{r_0}}|] \leq Cn^{1-\frac{1}{d}}(1 + s_{\max}\Lambda_{r_0}r_0^2)\left(e^{-\frac{1}{2}\Lambda}\sum_{j=0}^{k-1}\frac{(\frac{1}{2}\Lambda)^j}{j!} - e^{-\frac{1}{2}\Lambda_{r_0}}\sum_{j=0}^{k-1}\frac{(\frac{1}{2}\Lambda_{r_0})^j}{j!}\right)$$

Using the assumptions that $\Lambda_{r_0}r_0^2 \to 0$, $\Lambda \to \infty$ and $r = o(r_0)$ yields that $\mathbb{E}[|C_{\partial M_{r_0}}|] = O(n^{1-\frac{1}{d}}\Lambda^{k-1}e^{-\frac{1}{2}\Lambda})$. Thus we prove the Lemma 4.6.2.

Proof. (Proposition 4.6.1)

Having established Lemma 4.6.2 this proof proceeds like the proof of Proposition 6.1 in (79) mutatis mutandis. Details are provided in Section 4.9.

□

4.7 *Lower Threshold*

In this section we shall find a lower bound for the expected Betti numbers of the Čech complex $\mathcal{C}(n,r)$. We utilise the concept of a special type of critical point called a Θ-cycle defined in (78). Such a critical point is guaranteed to induce a non-trivial cycle in the homology of the resulting Čech complex. The lower bound shows that if the convergence of $\Lambda \to \infty$ is sufficiently slow then, w.h.p the Čech complex will have Betti numbers larger than that of the manifold. Thus we will attain a lower threshold for Λ.

The paper (79) counts Θ-cycles, a special class of critical points $c(\mathcal{Y})$ on a general compact closed Riemannian manifold with critical values in the range $(r_1, r]$. The radius r_1 is chosen to be sufficiently

small that there are a large number of critical points with critical value in the range $(r_1, r]$ but simultaneously sufficiently large that a Θ-cycle with critical value r_1 persists and remains a Θ-cycle in the Čech complex at radius r.

The conditions that determine a critical point to be a Θ-cycle are all local conditions. As such we can replicate the analysis in (79) to count the Θ-cycles in a manifold with boundary which are sufficiently distant from the boundary. It transpires that this lower bound can be improved by counting a collection of cycles which occur close to the boundary, which we shall call Θ-like-cycles.

Let us proceed to define the conditions given in (79) that determine a critical point to be a Θ-cycle. Let $\mathcal{Y} \subset \mathcal{P}_n \subset M$ be a generic subset of a Poisson point process on M inducing centre $c(\mathcal{Y})$ and $\varepsilon \in (0,1)$. Let us denote the closed annulus in M by:

$$A_\varepsilon(c(\mathcal{Y})) = \overline{B_{\rho(\mathcal{Y})}(c(\mathcal{Y}))} \setminus B_{\varepsilon\rho(\mathcal{Y})}(c(\mathcal{Y}))$$

Intuitively a Θ-cycle is formed at critical point c when an annulus surrounding c is covered and the critical point c introduces a k simplex which cuts across this annulus and introduces an erroneous homological cycle. If our centre is close to the boundary, our annulus will be cut by the boundary and so we must modify the argument in (79) to apply to this case.

Lemma 4.7.1 is a slight modification of that presented in (78) and having taken into consideration that the critical point lies far from the boundary the proof follows identically.

Lemma 4.7.1 ((78))**.** *Let $\mathcal{Y} \subset \mathcal{P}_n \subset M$ with $\mathcal{Y}$ inducing a critical point $c(\mathcal{Y})$ of index k. Let us define:*

$$\phi(\mathcal{Y}) = \frac{1}{2\rho(\mathcal{Y})} \min_{v \in \partial\Delta(\mathcal{Y})} |v|$$

where $\Delta(\mathcal{Y}) = \mathrm{Grad}(\mathcal{Y})$. Suppose $\rho(\mathcal{Y}) < r_{max}$, $dist(c(\mathcal{Y}), \partial M) > \rho(\mathcal{Y})$ and $A_\phi(\mathcal{Y}) \subset B_{\rho(\mathcal{Y})}(\mathcal{P}_n)$, then the critical point $c(\mathcal{Y})$ generates a new non-trivial cycle in $H_k(B_{\rho(\mathcal{Y})}(\mathcal{P}_n))$ which we call a Θ-cycle.

The following technical Lemma is established by showing that a Θ-cycle with critical value in the range $(r_1, r]$ will remain a Θ-cycle at radius r. We introduce an auxiliary radius $r_2 > r$ and require that $B_{r_2} \cap \mathcal{P}_n = \mathcal{Y}$. This condition ensures that the homology class introduce by a Θ-cycle is not annihilated by the introduction of another critical point with critical value in the range $(r_1, r]$.

Lemma 4.7.2 ((79))**.** *Suppose $r_2 > r > 0$ and that $r_1 > r\sqrt{1 - \frac{1}{c_8^2}(\frac{r_2}{r} - 1)^2}$. Let $\varepsilon \in (0,1)$ and $\Theta_k^\varepsilon(r_1, r)$ denote the number of Θ-cycles induced by subsets of $(k+1)$ points $\mathcal{Y}$ with the properties:*

$$\rho(\mathcal{Y}) \in (r_1, r], \quad B_{r_2}(c(\mathcal{Y})) \cap \mathcal{P}_n = \mathcal{Y}, \quad \phi(\mathcal{Y}) \geq \varepsilon, \quad dist(c(\mathcal{Y}), \partial M) > r_2$$

For any $\varepsilon \in (0,1)$ we have that $\beta_k(r) \geq \Theta_k^\varepsilon(r_1, r)$.

Let us use the following notation for indicator functions which track when a subset of $(k+1)$ elements $\mathcal{Y} \subset \mathcal{P}_n$ induces an element of $\Theta_k^\varepsilon(r_1, r)$. Recall the definition of $h_{r_1,r}(\mathcal{Y})$ which indicates when the collection of points induces a critical point with critical value in the range $(r_1, r]$.

- $h_r^\varepsilon(\mathcal{Y}) = h_{r_1,r}(\mathcal{Y})\mathbf{1}\{\phi(\mathcal{Y}) \geq \varepsilon\}\mathbf{1}\{\text{dist}(c(\mathcal{Y}), \partial M) > r_2\}$
- $g_r^\varepsilon(\mathcal{Y}, \mathcal{P}_n) = h_r^\varepsilon(\mathcal{Y})\mathbf{1}\{B_{r_2}(c(\mathcal{Y})) \cap (\mathcal{P}_n \setminus \mathcal{Y}) = \emptyset\}\mathbf{1}\{A_\varepsilon \subset B_{\rho(\mathcal{Y})}(\mathcal{P}_n)\}$

Hence we may write $\Theta_k^\varepsilon(r_1, r)$ as the sum:

$$\Theta_k^\varepsilon(r_1, r) = \sum_{|\mathcal{Y}|=k+1} g_r^\varepsilon(\mathcal{Y}, \mathcal{P}_n)$$

Lemma 4.7.3 ((79))**.** *Suppose $\varepsilon > 0$ is sufficiently small (independent of r and n), and $r > 0$ such that $\Lambda \to \infty, \Lambda r^2 \to 0$. Then for suitably chosen r_1, r_2 with $r_2 > r > r_1 > 0$ we have that:*

$$\mathbb{E}[\Theta_k^\varepsilon(r_1, r)] = \Omega(n\Lambda^{k-2}e^{-\Lambda})$$

The proof of the above lemma follows almost identically to the proof supplied in (79).

Proposition 4.7.4 ((79))**.** *Suppose $\varepsilon > 0$ is sufficiently small, $\omega(n) \to \infty$, $\gamma \in (0,1)$ is fixed and $\Lambda = \log n + (k-2)\log\log n - \omega(n)$, then for all $1 \leq k \leq d-1$:*

$$\lim_{n\to\infty} \mathbb{P}(\Theta_k^\varepsilon(r_1, r) > \gamma\mathbb{E}[\Theta_k^\varepsilon(r_1, r)]) = 1$$

This Proposition is proven using a second moment argument based on Chebyshev's inequality and the details may be found in (79).

The results from this section yield a lower threshold for Λ:

Proposition 4.7.5. *Let M be a unit volume, compact, Riemannian manifold with boundary. Let $\Lambda = \log n + (k-2)\log\log n - \omega(n)$ then for all $1 \leq k \leq d-1$:*

$$\lim_{n\to\infty} \mathbb{P}(H_k(\mathcal{C}(n, r)) \cong H_k(M)) = 0$$

This threshold is identical to that found in (79). We adapted the proof from (79) in order that we only counted Θ-cycles sufficiently far from the boundary that the local considerations did not detect the manifold had a boundary.

4.7.1 *Lower Threshold Refined*

We shall now adapt the argument counting Θ-cycles to establish that a large number of Θ-like-cycles are present close to the boundary. From this we can deduce a greater lower threshold. Let $\Theta_k^{\varepsilon,\partial M}(r_1, r)$ denote the number Θ-like-cycles of index k with critical value in the range $(r_1, r]$. This type of critical point induces an erroneous homological cycle near the boundary and will be defined thoroughly below, (Definition 4.7.9).

Similarly to the previous section we count Θ-like-cycles with critical values in the range $(r_1, r]$ with auxilliary radius r_2 used as in the previous section to ensure a Θ-like-cycle with critical value r_1 remains a Θ-like-cycle at scale r.

Lemma 4.7.6. *Let M be a unit volume compact Riemannian manifold with boundary. Suppose $n \to \infty$ and $r \to 0$ such that $\Lambda \to \infty$, $\Lambda r^2 \to 0$ and suppose $\alpha = \frac{1}{2} + O((\log n)^{-1})$. Then for all $1 \leq k \leq d-1$ we have that:*

$$\mathbb{E}[|\Theta_k^{\varepsilon,\partial M}(r_1, r)|] = \Omega(n\Lambda^{k-2}e^{-\alpha\Lambda}r(\log n)^{-(k+1)})$$

Heuristically a Θ-cycle is a critical point lying in the centre of a surrounding annulus, and thus induces a new cycle. For a critical point close to the boundary, the whole annulus is not contained in the manifold and so we define a Θ-like-cycle for this corresponding situation. The annuli which are cut by the boundary form a cup which for certain critical points give rise to non-trivial homological cycles. We shall identify which critical points introduce new homological cycles.

The argument to establish Lemma 4.7.6 in this section has the following structure. We first define a special class of critical points, Θ-like-cycles, and establish that these critical points introduce new spurious homological cycles near the boundary (Lemma 4.7.8). We then wish to bound from below the expected number of Θ-like-cycles. In order for a critical point to be a Θ-like-cycle we require that both a partial annulus surrounding the critical point is covered and that the simplex induced by the critical point is approximately tangential to the boundary. We define the partial annulus $A_\varepsilon^{(\varphi)}(c)$ which is constructed in order that if the points inducing the centre all lie in this annulus then the simplex introduced by the critical point is approximately tangential to the boundary. Then for suitable φ the partial annulus is covered w.h.p. and so both conditions are met and the critical point c thus induces a Θ-like-cycle.

Definition 4.7.7 (The Partial Annulus)**.** *Let c be a critical point with radius ρ such that $dist(\partial M, c) \in (0, \rho)$. Let p be the closest point to c on*

∂M. Let $\mathbf{n} := \exp_c^{-1} p$, which can be thought of as the vector pointing in the direction of the normal to the boundary at p. We define the partial annulus:

$$A_\varepsilon^{(\varphi)}(c) = \{x \in A_\varepsilon(c) \ : \ exp_c^{-1} x \text{ makes angle greater than } \pi/2 - \varphi \text{ with } \mathbf{n}\}$$

See Figure 4.3.

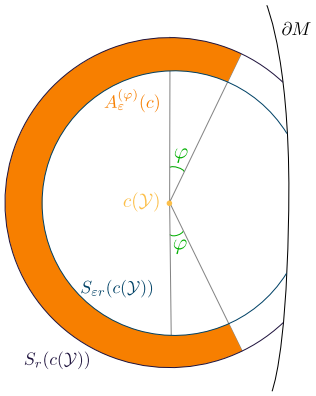


Figure 4.3: A sketch of the partial annulus $A_\varepsilon^{(\varphi)}(c)$. Our calculations in Section 4.7.4 show that if the centre $c(\mathcal{Y})$ is at distance $\delta \sim \frac{r}{\log n}$ from the boundary and $\varphi \sim (\log n)^{-1}$ then $A_\varepsilon^{(\varphi)}(c)$ is covered w.h.p.

4.7.2 Θ-like-cycles Induce Homology

Let us define a function $\psi(\mathcal{Y}, \varphi)$ for Θ-like-cycles that plays the same role as ϕ did for Θ-cycles. Recall $\phi(\mathcal{Y})$ was defined to be half of the distance from the centre $c(\mathcal{Y})$ to the closest point on $\partial\Delta(\mathcal{Y}) \subset T_{c(\mathcal{Y})}M$ the boundary of the k-simplex $\text{Conv}\{\nabla\rho_{y_i}(c)\}_{i=0}^k \subset T_{c(\mathcal{Y})}M$ divided by $\rho(\mathcal{Y})$. Let $\iota_k : C_k^\partial(\mathcal{C}(\mathcal{P}_n, r)) \to C_k^{\text{sing}}(B_r(\mathcal{P}_n))$ denote the natural map between the simplicial k-chains of the Čech complex and the singular k-chains of the topological space $B_r(\mathcal{P}_n)$. The value $\phi(\mathcal{Y})$ is such that $\iota_{k-1}(\partial\mathcal{Y}) \in C_{k-1}^{\text{sing}}(A_\phi(c(\mathcal{Y})))$. If also $A_\phi(c(\mathcal{Y})) \subset B_{\rho(\mathcal{Y})}(\mathcal{P}_n)$ then $\iota_{k-1}(\partial\mathcal{Y})$ is a boundary of a k-chain γ contained in $C_k^{\text{sing}}(B_{\rho(\mathcal{Y})}(\mathcal{P}_n))$. As a result, the k-simplex $\mathcal{Y}$ introduces a new k-cycle $(\iota_k(\mathcal{Y}) - \gamma)$ at radius $\rho(\mathcal{Y})$.

In order to prove a k-simplex $\mathcal{Y}$ associated to a critical point near the boundary introduces a new degree k homology class we seek to show that $\partial\mathcal{Y}$ is the boundary of a k-chain γ contained in $B_{\rho(\mathcal{Y})}(\mathcal{P}_n)$ which does not include the simplex $\mathcal{Y}$. In the case a critical point $c(\mathcal{Y})$ is near the boundary, the full annulus $A_\varepsilon(c(\mathcal{Y}))$ is cut by the boundary. This has a couple of important implications. The location of the $\mathcal{Y}$ about their centre relative to the boundary is now important since the annulus is no longer rotationally symmetric and so the probability of fulfilling the condition that the boundary of the simplex $\mathcal{Y}$ is covered by $B_{\rho(\mathcal{Y})}(\mathcal{P}_n)$ is dependent on the position of these points relative to the centre. In short, the location of the points $\mathcal{Y}$ about their centre now impacts whether the simplex introduces a new homology class.

We define ψ to ensure that the condition $\iota_{k-1}(\partial\mathcal{Y}) \subset A_\psi^{(\beta)}(c(\mathcal{Y}))$ is fulfilled and thus $\partial\mathcal{Y}$ is a boundary of a k-chain contained in $C_k^{\text{sing}}(A_\psi^{(\beta)}(c(\mathcal{Y}))) \subset C_k^{\text{sing}}(B_r(\mathcal{P}_n))$ with high probability:

$$\psi(\mathcal{Y}, \varphi) = \frac{1}{2} \sup\left(\{\varepsilon \geq 0 \mid \partial\Delta(\mathcal{Y}) \subset \exp_c^{-1} A_\varepsilon^{(\varphi)}(c(\mathcal{Y}))\} \cup \{0\}\right)$$

Note that like ϕ, the term $\psi(\mathcal{Y}, \varphi)$, is not sensitive to the scale $\rho(\mathcal{Y})$ in the sense that $\psi(\mathcal{Y}, \varphi)$ is dependent on the distribution of the points $\mathcal{Y}$ on the sphere centred at $c(\mathcal{Y}))$ and so has a lower bound

which does not scale with the critical value $\rho(\mathcal{Y})$. We will sometimes abbreviate $\psi(\mathcal{Y}, \varphi)$ to just ψ.

Lemma 4.7.8 (Θ-like-cycles induce non-trivial homological cycles)**.** *Let $\mathcal{Y} \subset \mathcal{P} \subset M$ be a set of points inducing an index k critical point with critical value ρ, dist$(c(\mathcal{Y}), \partial M)$, and $1 \leq k \leq d-1$. If $\psi = \psi(\mathcal{Y}, \varphi) > 0$ and $A_\psi^{(\varphi)}(c(\mathcal{Y})) \subset B_\rho(\mathcal{P})$, then $c(\mathcal{Y})$ induces a non-trivial cycle in $H_k(\mathcal{C}(\mathcal{P}, \rho))$.*

The proof of Lemma 4.7.8 follows the same structure as a proof in (78) that shows Θ-cycles induce non-trivial homology. For completeness we include the proof of Lemma 4.7.8 in Section 4.9.

Definition 4.7.9 (Θ-like-cycle)**.** *A critical point $c(\mathcal{Y})$ which satisfies the conditions of the above lemma is a Θ-like-cycle at scale $\psi(\mathcal{Y}, \varphi)$.*

Note that if $\partial\Delta(\mathcal{Y})$ intersects $A_0 \setminus A_0^{(\varphi)}$ then for all $\varepsilon > 0$ then $\psi(\mathcal{Y}, \varphi) = 0$. We do not want to count the critical points induced by such $\mathcal{Y}$ since we cannot prove that they introduce non-trivial homological cycle. We sketch an example of a critical point inducing a Θ-like-cycle in Figure 4.4, contrast this with an example of a critical point not inducing a Θ-like-cycle.

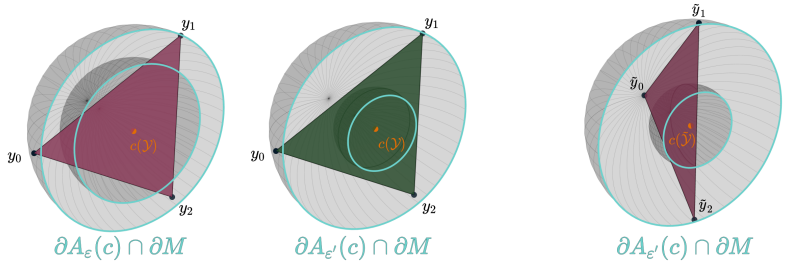


Figure 4.4: A sketch of an index 2 critical point near the boundary which induces a Θ-like-cycle (left), and an index 2 critical point near the boundary which does not inducing a Θ-like-cycle (right). Suppose $\varphi > 0$ is such that $A_\varepsilon^{(\varphi)}(c) \cap \partial M = \emptyset$ for all $\varepsilon > 0$. In the leftmost sketch we depict the critical point for the set $\mathcal{Y} = \{y_0, y_1, y_2\}$. We highlight the intersection of the annulus $A_\varepsilon(c)$ with the boundary of the manifold. The boundary of the simplex $\mathcal{Y}$ is not contained in the annulus $A_\varepsilon(c)$. Since the points $\mathcal{Y}$ lie in a hyperplane V tangential to the boundary the value $\psi(\mathcal{Y}, \varphi) > 0$ is dependent on the distribution of the points $\mathcal{Y}$ on the sphere $\mathbb{S}_{\rho(\mathcal{Y})}(V)$, but independent of how close $c(\mathcal{Y})$ is to the boundary of the manifold. The annulus $A_{\varepsilon'}^{(\varphi)}(c)$ witnesses that this critical point induces a Θ-like-cycle. In contrast, in the rightmost sketch we depict the critical point for the set $\tilde{\mathcal{Y}} = \{\tilde{y}_0, \tilde{y}_1, \tilde{y}_2\}$. Since the points $\tilde{\mathcal{Y}}$ lie in a hyperplane orthogonal to the boundary of the simplex $\tilde{\mathcal{Y}}$ is not contained in $A_{\varepsilon'}^{(\varphi)}(c(\tilde{\mathcal{Y}}))$ for all $\varepsilon' > 0$ and so this critical point does not induce a Θ-like-cycle.

4.7.3 Θ-like-cycle Lower Bound Computation

Let us modify our indicator functions from the previous section to count Θ-like-cycles with critical values in the range $[r_1, r)$. We count cycles whose critical point lies at a distance to the boundary in the range $[\delta, 2\delta]$, for suitably chosen δ.

- $h_r^{\varepsilon,\delta}(\mathcal{Y}) = h_{r_1,r}(\mathcal{Y})\mathbf{1}\{\psi(\mathcal{Y}, \varphi) \geq \varepsilon\}\mathbf{1}\{\delta \leq \text{dist}(c(\mathcal{Y}), \partial M) \leq 2\delta\}$
- $g_r^{\varepsilon,\delta}(\mathcal{Y}, \mathcal{P}) = h_r^{\varepsilon,\delta}(\mathcal{Y})\mathbf{1}\{B_{r_2}(c(\mathcal{Y})) \cap \mathcal{P} = \emptyset\}\mathbf{1}\{A_\varepsilon^{(\varphi)}(c(\mathcal{Y})) \subset B_{\rho(\mathcal{Y})}(\mathcal{P})\}$

Recall that the parameter r_2 is used to ensure that the other points of the Poisson process are sufficiently distant that from the Θ-like-cycles we count with critical value in the range $[r_1, r)$ that they not

annihilated by the introduction of another simplex with critical value in the range $[r_1, r)$. In particular r_2 must satisfy the hypotheses of Lemma 4.7.2

We shall choose δ to maximise our count of Θ-like-cycles. The dominant term in the lower bound of the count of Θ-like-cycles is the exponential term: $\exp(-n\alpha \mathrm{Vol}(B_{r_2}(c)))$ where $\alpha = \frac{1}{2} + O(r + \frac{\delta}{r})$. In order to match the dominant exponential term of the upper bound of critical points (Section 4.6) which dictates the leading term of the upper threshold, we require $\exp(-n\alpha \mathrm{Vol}(B_{r_2}(c))) \sim \exp(-\frac{1}{2}\Lambda)$. It follows that the optimal choice for δ to maximise the lower threshold is $\delta \sim \frac{r}{\log n}$.

The parameter φ is chosen to be as large as possible whilst satisfying the condition that the partial annulus $A_\varepsilon^{(\varphi)}(c)$ is covered w.h.p. for a critical point c with $\mathrm{dist}(c, \partial M) \in [\delta, 2\delta]$.

Lemma 4.7.8 shows that the indicator function $g_r^{\varepsilon,\delta}(\mathcal{Y}, \mathcal{P})$ is supported on subsets $\mathcal{Y} \subset \mathcal{P}$ which introduce erroneous cycles. If $\mathcal{P}_n$ is our Poisson process on M then we count the Θ-like-cycles of index k with the following sum of indicator functions:

$$\Theta_k^{\varepsilon,\partial M}(r_1, r) = \sum_{|\mathcal{Y}|=k+1} g_r^{\varepsilon,\delta}(\mathcal{Y}, \mathcal{P}_n)$$

4.7.4 *Partial ε-Annulus Coverage*

Let us denote the probability of covering the partial ε-annulus of a critical point $c(\mathbf{y})$ conditioned on the critical point having no points but $\mathbf{y}$ in an r_2 neighbourhood as:

$$p_{\varepsilon,\varphi}(\mathbf{y}) = \mathbb{P}(A_\varepsilon^{(\varphi)}(\mathcal{Y}') \subset B_{\rho(\mathcal{Y}')}(\mathcal{P}_n)) \mid \mathcal{Y}' = \mathbf{y},\ \mathcal{P}_n \cap B_{r_2}(c(\mathcal{Y}')) = \mathcal{Y}')$$

As in previous sections we use Palm Theory (Theorem 4.2.4) applied to the Poisson process on M to attain an integral expression:

$$\begin{aligned}\mathbb{E}[|\Theta_k^{\varepsilon,\partial M}|] &= \frac{n^{k+1}}{(k+1)!}\mathbb{E}[g_r^{\varepsilon,\delta}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)] \\ &= \frac{n^{k+1}}{(k+1)!}\int_{DM^{k+1}} h_r^{\varepsilon,\delta}(\mathbf{y}) p_{\varepsilon,\varphi}(\mathbf{y}) e^{-n\mathrm{Vol}(B_{r_2}(\mathbf{y})\cap M)} |\mathrm{dvol}_g(\mathbf{y})|\end{aligned}$$

The expectation depends on φ in the support of the functions $\mathbf{1}\{\psi(\mathcal{Y}, \varphi) \geq \varepsilon\}, \mathbf{1}\{A_\varepsilon^{(\varphi)} \subset B_{\rho(\mathcal{Y})}(\mathcal{P})\}$. Increasing φ, augments the support of $\mathbf{1}\{\psi(\mathcal{Y}, \varphi) \geq \varepsilon\}$ and diminishes the support of $\mathbf{1}\{A_\varepsilon^{(\varphi)} \subset B_{\rho(\mathcal{Y})}(\mathcal{P})\}$. The term $\mathbf{1}\{\psi(\mathcal{Y}, \varphi) \geq \varepsilon\}$ enforces that the points $\mathcal{Y}$ must lie in a hyperplane approximately tangent to the boundary, and the influence of this term on the lower bound is computed in the

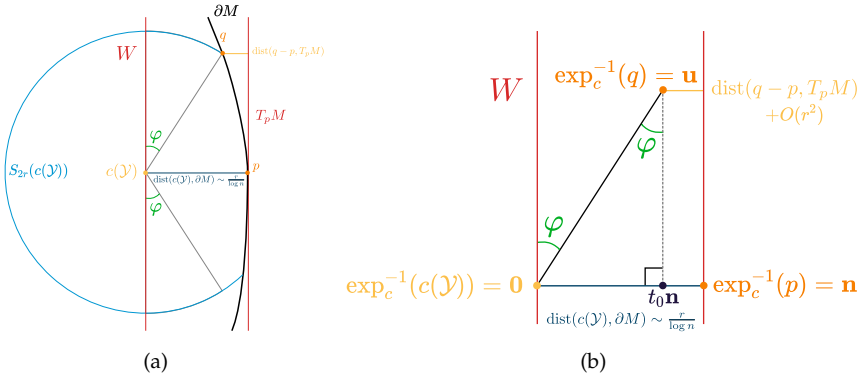
following section. Meanwhile the term $\mathbf{1}\{A_\varepsilon^{(\varphi)} \subset B_{\rho(\mathcal{Y})}(\mathcal{P})\}$ insists that the partial annulus is covered by the point process and is tracked by $p_{\varepsilon,\varphi}(\mathbf{y})$.

Recall our asymptotic notation $f(n) \sim g(n)$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$. Our next Lemma shows that if $\varphi \sim (\log n)^{-1}$ and $\delta(\mathbf{y}) \sim \frac{r}{\log n}$ then the term $p_{\varepsilon,\varphi}(\mathbf{y}) \to 1$ uniformly as $\Lambda \to \infty$ for fixed ε.

Lemma 4.7.10. *Let M be a smooth compact manifold with boundary and $r \sim \left(\frac{\log n}{n}\right)^{\frac{1}{d}}$. There exists a constant C such that if $\varphi = \frac{C}{\log n}$ and $\varepsilon \in (0,1)$ then for all $c(\mathbf{y}) \in M$ with $dist(c(\mathbf{y}), \partial M) \in [\frac{r}{\log n}, \frac{2r}{\log n}]$ the probability $p_{\varepsilon,\varphi}(\mathbf{y}) \to 1$ uniformly as $\Lambda \to \infty$.*

Proof. First observe that for any angle φ and any centre c we have that $\mathrm{Vol}(A_\varepsilon^{(\varphi)}(c)) \sim \rho^d(1-\varepsilon^d)$ and so there is a constant C' dependent only on the metric g such that there is an $\frac{\varepsilon\rho}{10}$-net of $A_\varepsilon^{(\varphi)}(c)$, $\mathcal{S}$, with $|\mathcal{S}| \leq C'\frac{1-\varepsilon^d}{\varepsilon^d}$. If for all $s \in \mathcal{S}$ we have $\mathcal{P}_n \cap B_{\rho(1-\varepsilon/10)}(s) \neq \emptyset$ then $A_\varepsilon^{(\varphi)}(c) \subset B_\rho(\mathcal{P}_n)$. Thus conditioning on the event that $\{\mathcal{P}_n \cap B_{r_2} = \emptyset\}$ we may bound $p_{\varepsilon,\varphi}$ from below:

$$p_{\varepsilon,\varphi} \geq 1 - C' \max_{s\in\mathcal{S}} e^{-n\mathrm{Vol}(B_{\rho(1-\varepsilon/10)}(s)\setminus B_{r_2}(c))}$$



(a) (b)

Figure 4.5: We sketch the geometric constructions present in the argument of Section 4.7.4. In Figure 4.5(a) we sketch $p \in \partial M$ the closest point to $c(\mathcal{Y})$ lying on the boundary and the point $q \in \partial M \cap S_{2r}(c(\mathcal{Y}))$ which is such that the angle φ between $\mathbf{u} = \exp_c^{-1}(q)$ and the *tangent* hyperplane $W \subset T_cM$ is minimal over all points in the intersection $\partial M \cap S_{2r}(c(\mathcal{Y}))$. In Figure 4.5(b) we sketch the preimages of p and q in the tangent space T_cM and use the fact that the metric on the tangent space deviates from the Riemannian metric on M by a term of order $O(r^2)$ in order to compute that $\varphi \sim \sin\varphi \sim \frac{\|t_0\mathbf{n}\|}{\|\mathbf{u}\|} = \frac{\mathrm{dist}(c(\mathbf{y}),\partial M) \pm d(q-p,T_pM)+O(r^2)}{2r} \sim \frac{1}{\log n}$.

In order to show that $p_{\varepsilon,\varphi}(\mathbf{y}) \to 1$ we will show that there is a constant C independent of $\mathbf{y}$ such that if $\varphi = \frac{C}{\log n}$ then the volume $\mathrm{Vol}(B_{\rho(1-\varepsilon/10)}(x) \setminus B_{r_2}(c(\mathbf{y})))$ is bounded below by $\frac{\omega_d}{20^d}\varepsilon^d r^d$ for all $x \in A_\varepsilon^{(\varphi)}(c(\mathbf{y}))$.

For the next part of the proof we will show there is a constant C independent of $\mathbf{y}$ such that for all $c(\mathbf{y}) \in M$ with $\mathrm{dist}(c(\mathbf{y}), \partial M) \in [\frac{r}{\log n}, \frac{2r}{\log n}]$ and for all $x \in A_\varepsilon^{(\frac{C}{\log n})}(c(\mathbf{y}))$ the geodesic emanating from $c(\mathbf{y})$ passing through x of length $2r$ is completely contained in M and does not intersect ∂M.

Fix some critical point $c = c(\mathbf{y})$ with $\mathrm{dist}(c(\mathbf{y}), \partial M) \in [\frac{r}{\log n}, \frac{2r}{\log n}]$ and let $p \in \partial M$ be the unique point on the boundary closest to c,

and let $\mathbf{n} \in T_cM$ be the *normal* vector at such that $\exp_c(\mathbf{n}) = p$. This induces the *tangent* hyperplane W at c; $W \subset T_cM$ such that $\mathbf{n} \cdot W = 0$ (see Figure 4.5(a)).

Let us define $U_c = \{\mathbf{u} \in T_cM : \exp_c(\mathbf{u}) \in \partial M \text{ and } \|u\| = 2r\} = \exp_c^{-1}(S_{2r}(c) \cap \partial M)$ and define $\varphi_{\mathbf{u}}$ to be the acute angle formed between $\mathbf{u}$ and W, and let $\varphi = \min_{\mathbf{u} \in U_c} \varphi_{\mathbf{u}}$. Suppose the minimum is attained in direction $\mathbf{u}$ so that $\varphi = \varphi_{\mathbf{u}}$ and the point $q = \exp_c(\mathbf{u})$ lies on the boundary and is distance $2r$ from $c(\mathbf{y})$.

We form a right angled triangle in T_cM by adding a construction line perpendicular to $\mathbf{n}$ passing through $\mathbf{u}$ and intersecting the line $\{t\mathbf{n} : t \in \mathbb{R}\}$ at the point $t_0\mathbf{n}$. This triangle has vertices $\mathbf{0}, \mathbf{u}, t_0\mathbf{n}$ and the angle at vertex $\mathbf{u}$ is φ (see Figure 4.5(b)). Thus $\sin\varphi = \frac{\|t_0\mathbf{n}\|}{\|u\|} = \frac{\|t_0\mathbf{n}\|}{2r}$.

We can compute $\|t_0\mathbf{n}\|$ using the Euclidean metric on the tangent space T_cM as an approximation for g the metric of the manifold. In the local $2r$-neighbourhood of c the Euclidean metric differs from g by a term of order $O(r^2)$ (Equation (4.8)).

We compute that $\|t_0\mathbf{n}\| = \operatorname{dist}(c(\mathbf{y}), \partial M) \pm d(q-p, T_pM) + O(r^2)$ (see Figure 4.5(b)). Using Theorem 4.2.8 we bound $\operatorname{dist}(q-p, T_pM) \leq \frac{\|q-p\|^2}{2\tau_M} \leq \frac{4r^2}{2\tau_M}$. Note that $r^2 = o(\frac{r}{\log n})$ and hence $\|t_0\mathbf{n}\| = \operatorname{dist}(c(\mathbf{y}), \partial M) + O(r^2) \sim \frac{r}{\log n}$. It follows that $\sin\varphi \sim \varphi \sim \frac{1}{\log n}$. Hence there exists a constant C such that for every centre $c(\mathbf{y})$ with $\operatorname{dist}(c(\mathbf{y}), \partial M) \in [\frac{r}{\log n}, \frac{2r}{\log n}]$ for any $x \in A_\varepsilon^{(\frac{C}{\log n})}(c(\mathbf{y}))$ the radial geodesic emanating from c passing through x of length $2r$ does not intersect ∂M.

Next let us show that for all $x \in A_\varepsilon^{(\frac{C}{\log n})}(c)$ the volume $\operatorname{Vol}(B_{r(1-\varepsilon/10)}(x) \setminus B_{r_2}(c)) \geq \frac{\omega_d}{20^d}\varepsilon^d r^d$. We have depicted the constructions in this argument in Figure 4.6 for clarity. Let $\mathbf{v}$ be a unit vector making acute angle $\frac{C}{\log n}$ with the *tangent* hyperplane W normal to $\mathbf{n}$.

In order to show that $\operatorname{Vol}(B_{r(1-\varepsilon/10)}(x) \setminus B_{r_2}(c)) \geq \frac{\omega_d}{20^d}\varepsilon^d r^d$ for all $x \in A_\varepsilon^{(\frac{C}{\log n})}(c)$, it suffices to show this property for $x = \exp_c(\varepsilon r\mathbf{v})$ since the volume $\operatorname{Vol}(B_{r(1-\varepsilon/10)}(x) \setminus B_{r_2}(c))$ increases as x moves away from c in the radial direction.

Let $t = r_2 + \frac{1}{2}((1+\frac{9\varepsilon}{10})r - r_2)$ and consider the volume of $B_{(t-r_2)}(v)$ the radius $t - r_2$ ball centred at $v = \exp_c(t\mathbf{v})$. At least half of this ball does not intersect the boundary and since $r_2 \to r$ as $n \to \infty$ we may assume $t - r_2 \geq \frac{\varepsilon r}{20}$ thus the volume of this ball is at least $\frac{\omega_d}{20^d}\varepsilon^d r^d$. By construction $B_{(t-r_2)}(v)$ is completely contained in $B_{r(1-\varepsilon/10)}(x) \setminus B_{r_2}(c)$ as $\rho(c,v) \geq r_2 + (t - r_2)$, and $\rho(x,v) + (t - r_2) = 2t - \varepsilon r - r_2 = r(1 - \frac{\varepsilon}{10})$. Thus $\operatorname{Vol}(B_{r(1-\varepsilon/10)}(x) \setminus B_{r_2}(c)) > \frac{\omega_d}{20^d}\varepsilon^d r^d$.

Hence given we have shown for any $c(\mathbf{y})$ that $\operatorname{Vol}(B_{r(1-\varepsilon/10)}(x) \setminus$

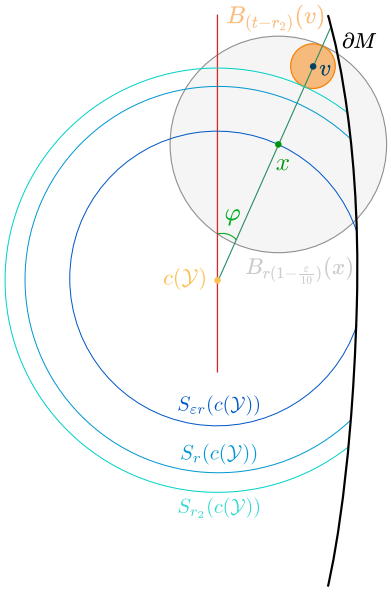


Figure 4.6: Let $\mathbf{v} \in T_cM$ be a unit vector making acute angle $\frac{C}{\log n}$ with the *tangent* hyperplane W. Let $x = \exp_c(\varepsilon r\mathbf{v})$, $t = r_2 + \frac{1}{2}((1+\frac{9\varepsilon}{10})r - r_2)$ and $v = \exp_c(t\mathbf{v})$. We sketch the ball $B_{(t-r_2)}(v)$ which witnesses that $\operatorname{Vol}(B_{r(1-\varepsilon/10)}(x) \setminus B_{r_2}(c)) > \frac{\omega_d}{20^d}\varepsilon^d r^d$.

$B_{r_2}(c)) \geq \frac{\omega_d}{20^d}\varepsilon^d r^d$ for all $x \in A_\varepsilon^{(\frac{C}{\log n})}(c)$ then it follows that:

$$\begin{aligned} p_{\varepsilon,\varphi} &\geq 1 - C' \max_{s\in\mathcal{S}} e^{-n\mathrm{Vol}(B_{\rho(1-\varepsilon/10)}(s)\setminus B_{r_2}(c))} \\ &\geq 1 - C' e^{-n\varepsilon^d r^d/1000} = 1 - C' e^{-\varepsilon^d \Lambda/1000\omega_d} \to 1 \end{aligned}$$

uniformly as $\Lambda \to \infty$ for fixed ε. □

Since $p_{\varepsilon,\varphi}(\mathbf{y}) \to 1$ uniformly as $\Lambda \to \infty$ for fixed ε, we may remove the term $p_{\varepsilon,\varphi}$ from the integral and replace it by a constant.

$$\mathbb{E}[|\Theta_k^{\varepsilon,\partial M}|] \geq C\frac{n^{k+1}}{(k+1)!}\int_{DM^{k+1}} h_r^{\varepsilon,\delta}(\mathbf{y}) e^{-n\mathrm{Vol}(B_{r_2}(c(\mathbf{y}))\cap M)} |\mathrm{dvol}_g(\mathbf{y})|$$

The next term we shall attack in the lower bound is $e^{-n\mathrm{Vol}(B_{r_2}(c(\mathbf{y}))\cap M)}$. Using the result of Proposition 4.2.9 and the Blaschke Petkantschin formula we attain the lower bound:

$$\begin{aligned} \mathbb{E}[|\Theta_k^{\varepsilon,\partial M}|] &\geq C\frac{n^{k+1}}{(k+1)!}\int_{DM^{k+1}} h_r^{\varepsilon,\partial M_\delta}(\mathbf{y}) e^{-n\alpha\mathrm{Vol}(B_{r_2}(c(\mathbf{y})))} |\mathrm{dvol}_g(\mathbf{y})| \\ &= C\frac{n^{k+1}}{(k+1)!}\int_{\partial M_{[\delta,2\delta]}} |\mathrm{dvol}_g(c)| \int_{r_1}^{r} du\ u^{dk-1} \int_{Gr(k,T_cM)} d\mu_{k,d}(V) \\ &\times \prod_{i=1}^{k}\left(\int_{\mathrm{S}_1(V)} \sqrt{|\det(g_{\exp_c(uw_i)})|}|\mathrm{dvol}_{\mathrm{S}_1(V)}(w_i)|\right) \mathbf{Y}_1^{d-k}(\mathbf{w}) f(\exp_c(u\mathbf{w})) \end{aligned}$$

Where $f(\mathbf{y}) = h_r^{\varepsilon,\partial M_\delta}(\mathbf{y}) e^{-n\alpha\mathrm{Vol}(B_{r_2}(c(\mathbf{y})))}$ and $\alpha = \frac{1}{2} + O(r, \frac{\delta}{r}) = \frac{1}{2} + O((\log n)^{-1})$. We can bound from below the components of this integral using the fact that the manifold is compact. In particular using Lemma 4.2.5:

$$e^{-n\alpha\mathrm{Vol}(B_{r_2}(c(\mathbf{y})))} \geq e^{-\alpha\Lambda_{r_2}}(1 + s_{\min}\Lambda_{r_2} r^2)$$

$$\begin{aligned} \mathbb{E}[|\Theta_k^{\varepsilon,\partial M}|] &\geq C\frac{n^{k+1}}{(k+1)!} e^{-\alpha\Lambda_{r_2}}(1 + s_{\min}\Lambda_{r_2} r^2) \int_{\partial M_{[\delta,2\delta]}} |\mathrm{dvol}_g(c)| \int_{r_1}^{r} du\ u^{dk-1} \int_{Gr(k,T_cM)} d\mu_{k,d}(V) \\ &\times \prod_{i=1}^{k}\left(\int_{\mathrm{S}_1(V)} \sqrt{|\det(g_{\exp_c(uw_i)})|}|\mathrm{dvol}_{\mathrm{S}_1(V)}(w_i)|\right) \mathbf{Y}_1^{d-k}(\mathbf{w}) h_r^{\varepsilon,\delta}(\exp_c(u\mathbf{w})) \end{aligned}$$

4.7.5 *Grassmannian Volume*

We require subtle analysis to bound from below the contribution of the integrals over the Grassmannian. Let us define:

$$D_{k,\varphi}^{\varepsilon} = \int_{Gr(k,T_cM)} d\mu_{k,d}(V) \times \prod_{i=1}^{k}\left(\int_{\mathrm{S}_1(V)} \sqrt{|\det(g_{\exp_c(uw_i)})|}|\mathrm{dvol}_{\mathrm{S}_1(V)}(w_i)|\right) \mathbf{Y}_1^{d-k}(\mathbf{w}) h_r^{\varepsilon,\delta}(\exp_c(u\mathbf{w}))$$

Taking sufficiently small r we may assume the determinant term is arbitrarily close to 1. Moreover given $\psi(\mathbf{y}, \varphi) \geq \varepsilon$ it is clear the volume $\Upsilon(\mathbf{w})$ is bounded below by a term of order $\text{dist}(A_\varepsilon^{(\varphi)}, c(\mathcal{Y}))^k = \Omega(\varepsilon^k)$.

Let us calculate a lower bound for the volume of the subspace of the Grassmannian $\text{Gr}(k, d)$ for which the points inducing our centre must lie in order that $c(\mathcal{Y})$ induces a Θ-like cycle of order k, see Figure 4.7. Since we are using the invariant measure $\mu_{k,d}$ on $\text{Gr}(k, d)$ as in the Blaschke-Petkantschin Formula we can appeal to Theorem 13.1.5 from *Stochastic and Integral Geometry* (106), and compute a bound for this volume using the action of $SO(d)$ on $\text{Gr}(k, d)$.

For sufficiently small r there is a unique closest point to c on the boundary $p \in \partial M$. Let $\mathbf{n} \in T_c M$ be such that $\exp_c(\mathbf{n}) = p$ and let $e_d = \mathbf{n}/\|\mathbf{n}\|$ be the unit vector *normal* to the boundary at c.

Define V to be a subset of the sphere in the tangent space at c as sketched in Figure 4.7:

$$V := \{\mathbf{v} \in T_c M : \|v\| = 1, |\langle \mathbf{v}, \mathbf{n} \rangle| \leq \sin(\varphi)\} \subset T_{c(\mathcal{Y})} M$$

This induces a subset of the Grassmannian:

$$\mathcal{V} := \{W \in \text{Gr}(k, d) : W = \langle v_1, ..., v_k \rangle, v_i \in V\} \subset \text{Gr}(k, d)$$

For our suitably chosen φ, a critical point induced by points lying in V will introduce a simplex which lies approximately tangential to the boundary and thus will induce a Θ-like-cycle. In order to calculate the volume of the associated subset of the Grassmannian $\mathcal{V}$ we shall use Theorem 4.7.11.

Theorem 4.7.11 ((106, Theorem 13.1.5)). *Let G be a compact group operating continuously and transitively on a Hausdorff space E, and suppose G and E have countable bases. Let ν be a Haar measure on G with $\nu(G) = 1$. Then there exists a unique G-invariant Borel measure ρ on E with $\rho(E) = 1$ defined by:*

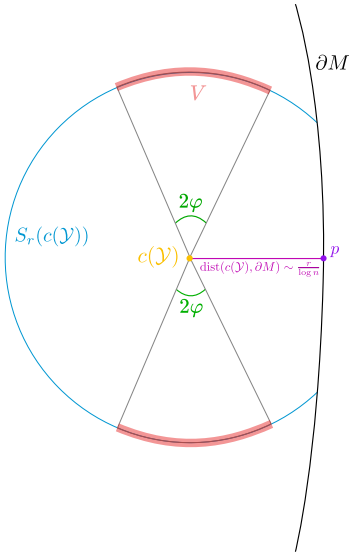
$$\rho(B) = \nu(\{g \in G \ : \ g \cdot x_0 \in B\}), \quad B \in \mathcal{B}(E)$$

for arbitrary $x_0 \in E$.

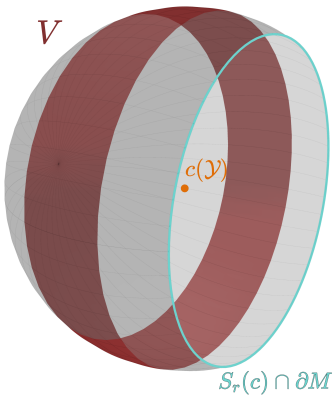
We observe that the compact group $SO(d)$ acts continuously and transitively on $\text{Gr}(k, d)$. We can parametrise $SO(d)$ with $\binom{d}{2}$ angles $\{\phi_{i,j}\}_{1 \leq i \leq j \leq d-1}$ (107; 108).

Let us briefly recall this parametrization. Denote the augmentation of vectors and matrices via a superscript a:

$$R^a = \begin{bmatrix} R & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (v_1, ..., v_l)^a = (v_1, ..., v_l, 0)$$



(a) A cartoon of a critical point near the boundary.



(b) A critical point near the boundary in a 3 manifold with boundary.

Figure 4.7: In Figure 4.7(a) we sketch a critical point $c(\mathcal{Y})$ close to the boundary to highlight the important quantities in computing a lower bound for the number of Θ-like-cycles. We highlight a subset $V \subset S_r(c(\mathcal{Y}))$ such that if every point of $\mathcal{Y}$ lies in V the simplex introduced by the critical point c is a k-dimensional lid on the sliced-annulus centred around c and so induces a Θ-like cycle. The corresponding subset $\mathcal{V} \subset \mathrm{Gr}(k, T_c M)$ consists of those k-planes whose intersection with the sphere lies in $V \subset S_r(c(\mathcal{Y}))$. In Figure 4.7(b) we give a concrete example for M a 3-manifold with boundary. We highlight where the sphere intersects the boundary. If the points $\mathcal{Y}$ inducing an index 2 critical point at c lie in the highlighted subset they induce a Θ-like-cycle as in Figure 4.4.

Recall the hyperspherical coordinate system of the unit d-sphere S^d given by d angles $\theta_1, ..., \theta_d$, $\Sigma_d : \mathbb{R}^d \to S^d$. Let π_i denote the projection onto the i^{th} coordinate.

$$\pi_i(\Sigma_d(\theta_1, ..., \theta_d)) = \begin{cases} \sin\theta_1 ... \sin\theta_d & \text{if } i = 1 \\ \cos\theta_{i-1} \sin\theta_i ... \sin\theta_d & \text{if } 2 \leq i \leq d+1 \end{cases}$$

We have an orthonormal basis $\{\mathbf{u}_{d,1}, ..., \mathbf{u}_{d,d}, \Sigma_d(\theta_1, ..., \theta_d)\}$ where $\mathbf{u}_{d,i}$ is the unit vector in the direction $\frac{\partial \Sigma_d(\theta_1,...,\theta_d)}{\partial \theta_i}$. Let the rotation $R_{d+1}(\theta_1, ..., \theta_d) \in SO(d+1)$ be the matrix:

$$R_{d+1}(\theta_1, ..., \theta_d) = \begin{bmatrix} \mathbf{u}_{d,1} & ... & \mathbf{u}_{d,d} & \Sigma_d(\theta_1, ..., \theta_d) \end{bmatrix}$$

Observe that R_{d+1} maps the standard basis vector $\mathbf{e}_{d+1}$ to the point $\Sigma_d(\theta_1, ..., \theta_d) \in S^d$.

We inductively construct $\Phi_d \in SO(d)$ from $\binom{d}{2}$ angles $\{\phi_{i,j}\}_{1 \leq i \leq j \leq d-1}$ as follows:

$$\Phi_2(\phi_{1,1}) = \begin{bmatrix} \cos\phi_{1,1} & \sin\phi_{1,1} \\ -\sin\phi_{1,1} & \cos\phi_{1,1} \end{bmatrix}$$

$$\Phi_d(\{\phi_{i,j}\}_{1 \leq i \leq j \leq d-1}) = R_d(\phi_{1,d-1}, ..., \phi_{d-1,d-1}) \begin{pmatrix} \Phi_{d-1}(\{\phi_{i,j}\}_{1 \leq i \leq j \leq d-2}) \end{pmatrix}^a$$

$$\phi_{1,j} \in [0, 2\pi],\ \phi_{i,j} \in [0, \pi] \text{ for } 2 \leq i \leq d-1, 1 \leq j \leq d-1$$

The Haar measure on $SO(d)$ is then given up to normalising con-

stant by:

$$d\nu = \left(\prod_{1 \leq i \leq j \leq d-1} \sin^{i-1} \phi_{i,j} \right) d\phi_{1,1} ... d\phi_{d-1,d-1}$$

Let P_k denote the projection onto the k-plane x_0. Observe that if $\|P_k \Phi_d e_d\| \leq \sin \varphi$ then $\Phi_d x_0 \in \mathcal{V}$. Recall $x_0 = \langle e_1, ..., e_k \rangle$ where e_i are standard basis vectors. Using the inductive construction we observe that $\|P_k \Phi_d e_d\| = \|P_k \Sigma_{d-1}(\phi_{1,d-1}, ..., \phi_{d-1,d-1})\| \leq \sin \phi_{k,d-1}$ since the first k coordinates of $\Sigma_{d-1}(\phi_{1,d-1}, ..., \phi_{d-1,d-1})$ contain a factor of $\sin \phi_{k,d-1}$.

Thus we attain the following estimate for $\mu_{k,d}(\mathcal{V})$:

$$\begin{aligned} \mu_{k,d}(\mathcal{V}) &= \int_{\Phi \in SO(d)} \mathbf{1}\{\Phi \cdot x_0 \in \mathcal{V}\} d\nu \sim \int_{\phi_{i,j}} \mathbf{1}\{\Phi \cdot x_0 \in \mathcal{V}\} \left(\prod_{1 \leq i \leq j \leq d-1} \sin^{i-1} \phi_{i,j} \right) d\phi_{1,1} ... d\phi_{d-1,d-1} \\ &\geq C \int_0^{\varphi/2} \sin^{k-1} \phi_{k,d-1} d\phi_{k,d-1} = \Omega(\varphi^k) \end{aligned}$$

Thus we have a lower bound estimate for the Grassmannian volume in terms of the angle φ.

Given a critical point c induced by $\mathcal{Y} \subset A_\varepsilon^{(\varphi)}(c)$ and associated critical simplex Δ, for sufficiently small ε we have $\partial\Delta \subset A_\varepsilon^{(\varphi)}(c)$ and thus c induces a Θ-like cycle.

We can compute a lower bound for the support of the term $\mathbf{1}\{\psi(\mathbf{y}, \varphi) \geq \varepsilon\}$ by observing that:

$$\mathbf{1}\{\psi(\mathbf{y}, \varphi) \geq \varepsilon\} \geq \mathbf{1}\{\phi(\mathbf{y}) \geq \varepsilon\} \mathbf{1}\{\mathbf{v} \in V \subset T_{c(\mathbf{y})} M\}$$

The support of the term $\mathbf{1}\{\mathbf{v} \in V \subset T_{c(\mathbf{y})} M\}$ may be bounded below using the Grassmannian volume calculation above and as observed in (79) the support of the term $\mathbf{1}\{\phi(\mathbf{y}) \geq \varepsilon\}$ is bounded below by a constant for fixed ε. Hence we conclude that for sufficiently small ε we have $D_{k,\varphi}^\varepsilon = \Omega((\log n)^{-k})$.

We can calculate a lower bound using this estimate and with $r_1 = r(1 - \frac{\xi^2}{2c_g^2})$, $r_2 = r(1 + \xi)$:

$$\begin{aligned} \mathbb{E}[|\Theta_k^{\varepsilon, \partial M}|] &\geq C(1 - c_R r^2)(1 + s_{\min} \Lambda_{r_2} r^2) D_{k,\varphi}^\varepsilon n^{k+1} e^{-\alpha \Lambda_{r_2}} \int_{\partial M_{[\delta, 2\delta]}} |\mathrm{dvol}_g(c)| \int_{r_1}^r du \; u^{dk-1} \\ &\geq C D_{k,\varphi}^\varepsilon n^{k+1} e^{-\alpha \Lambda_{r_2}} \delta r^{dk} \int_{r_1/r}^1 s^{dk} \, ds \\ &\geq C D_{k,\varphi}^\varepsilon n \Lambda^k e^{-\alpha \Lambda_{r_2}} \delta \xi^2 \end{aligned}$$

For $\xi = \Lambda^{-1}$ and $\Lambda \sim \log n$ we observe that asymptotically $e^{-\alpha \Lambda_{r_2}} = \Omega(e^{-\alpha \Lambda})$ and so:

$$\begin{aligned}\mathbb{E}[|\Theta_k^{\varepsilon,\partial M}|] &\geq CD_{k,\varphi}^{\varepsilon} n\Lambda^{k-2}e^{-\alpha\Lambda_{r_2}}\delta \\ &\geq Cn\Lambda^{k-2}e^{-\alpha\Lambda}r(\log n)^{-(k+1)}\end{aligned}$$

Thus we see that $\mathbb{E}[|\Theta_k^{\varepsilon,\partial M}|] = \Omega(n\Lambda^{k-2}e^{-\alpha\Lambda}r(\log n)^{-(k+1)})$

In order to complete the argument for the lower threshold we need to verify that the second moment arguments in (79) carry over to our modified notion of a Θ-like-cycle, and so w.h.p. the number of Θ-like-cycles is bounded below by the same regime as the expected number of Θ-like-cycles. We compute the second moments in Section 4.10 and verify that indeed w.h.p. the number of Θ-like-cycles is bounded below by the same regime as the expected number of Θ-like-cycles.

4.8 Main Result

Collecting the results contained in the previous sections we attain thresholds in terms of Λ for the k^{th} homological connectivity of a compact manifold with boundary M. For Λ greater than the upper threshold we recover the k^{th} homology of M with high probability, and for Λ less than the lower threshold we do not recover the k^{th} homology of M with high probability.

A recent coverage result from (104) yields a sharp coverage threshold for Riemannian manifolds with boundary. For $\Lambda = (2-\frac{2}{d})\log n + 2(d-2)\log\log n + w(n)$ the manifold is covered with high probability. Our result shows that the homological connectivity threshold for lower homology groups occurs before coverage and that the first order term for homological connectivity is $(2-\frac{2}{d})\log n$.

Theorem 4.8.1 (Homological Connectivity of a Riemannian Manifold with Boundary)**.** *Let M be a unit volume compact Riemannian manifold with smooth non-empty boundary. Let $d \geq 2$ be the dimension of M, $\Lambda = \omega_d n r^d$ and $\mathcal{P}_n$ a Poisson process of intensity n on M. Suppose $w(n)$ is any function with $w(n) \to \infty$ as $n \to \infty$. Then for $1 \leq k \leq d-1$*

$$\lim_{n\to\infty} \mathbb{P}(H_k(\mathcal{C}(n,r)) \cong H_k(M)) = \begin{cases} 1 & \Lambda = (2-\frac{2}{d})\log n + 2k\log\log n + w(n), \\ 0 & \Lambda = (2-\frac{2}{d})\log n + 2(k-2-(k+1-\frac{1}{d}))\log\log n - w(n), \end{cases}$$

Proof. (Upper Threshold)

Let $r_0 = r(\frac{\omega_d}{\kappa}(1+|\log r|))^{1/d}$ so that the conditions of Lemma 4.6.2 are met. Asymptotically $\Lambda_{\frac{r_0}{2}} \geq 2\log n$. Using our Asymptotic Coverage result Corollary 4.4.2 we observe that $M \subset B_{r_0}(\mathcal{P}_n)$ w.h.p. and

so for sufficiently small r_0, by the Nerve Lemma $(H_k(\mathcal{C}(n,r_0)) \cong H_k(M))$ w.h.p.

Moreover for Λ as described in the upper threshold we observe that $n\Lambda^k e^{-\Lambda}, n^{1-\frac{1}{d}}\Lambda^k e^{-\frac{1}{2}\Lambda} \to 0$ so by Lemma 4.6.2 the expected number of k-critical and $(k+1)$-critical points with critical value in the range $(r, r_0]$ goes to zero:

$$\Lambda = (2-\frac{2}{d})\log n + 2k\log\log n + w(n), \implies n\Lambda^k e^{-\Lambda} = O\left(\frac{n(\log n)^k}{n^{(2-2/d)}(\log n)^{2k}e^{w(n)}}\right) \to 0$$

$$\Lambda = (2-\frac{2}{d})\log n + 2k\log\log n + w(n), \implies n^{1-\frac{1}{d}}\Lambda^k e^{-\frac{1}{2}\Lambda} = O\left(\frac{n^{1-\frac{1}{d}}(\log n)^k}{n^{1-1/d}(\log n)^k e^{w(n)}}\right) \to 0$$

Using Morse Theory for manifolds with boundary we yield that $H_k(\mathcal{C}(n,r)) \cong H_k(\mathcal{C}(n,r_0))$ w.h.p. since the probability of there being a critical point in the range $(r, r_0]$ tends to zero, and thus $H_k(\mathcal{C}(n,r)) \cong H_k(M)$ w.h.p. for $\Lambda = (2-\frac{2}{d})\log n + 2k\log\log n + w(n)$. □

Proof. (Lower Threshold)

Recall that by Lemma 4.7.6 we have $\mathbb{E}[\Theta_k^{\varepsilon,\partial M}] = \Omega(e^{-\alpha\Lambda}n\Lambda^{k-2}r(\log n)^{-(k+1)})$ for $\alpha = \frac{1}{2} + O((\log n)^{-1})$. Second Moment Calculations in Section 4.10 verify that w.h.p. $\Theta_k^{\varepsilon,\partial M} > \frac{1}{2}\mathbb{E}[\Theta_k^{\varepsilon,\partial M}]$ and thus w.h.p. $|\beta_k(r)| = \Omega(e^{-\alpha\Lambda}n\Lambda^{k-2}r(\log n)^{-(k+1)})$.

Then we note that for $\Lambda = (2-\frac{2}{d})\log n + 2(k-2-(k+1-\frac{1}{d}))\log\log n - w(n)$:

$$e^{-\alpha\Lambda}n\Lambda^{k-2}r(\log n)^{-(k+1)} \sim \frac{n^{1-\frac{1}{d}}(\log n)^{k-2+\frac{1}{d}-(k+1)}}{e^{\frac{\Lambda}{2}}} \sim \frac{n^{1-\frac{1}{d}}(\log n)^{k-2+\frac{1}{d}-(k+1)}}{n^{1-\frac{1}{d}}(\log n)^{k-2+\frac{1}{d}-(k+1)}e^{-\frac{w(n)}{2}}} \to \infty$$

Thus for Λ in this regime the Betti numbers tend to infinity w.h.p. and so we do not recover the homology of M. □

4.9 Conclusion

Having adapted the techniques from (79) in order to count critical points near to the boundary we have attained thresholds similar to those of the Homological Connectivity Theorem from (79), applicable to compact, closed Riemannian Manifolds.

Theorem 4.9.1 (Homological Connectivity Thresholds for Compact Manifolds without Boundary (79))**.** *Let M be a unit volume compact, Riemannian manifold without boundary. Suppose that as $n \to \infty$, $w(n) \to$*

∞. *Then for* $1 \leq k \leq d-1$

$$\lim_{n\to\infty} \mathbb{P}(H_k(\mathcal{C}(n,r)) \cong H_k(M)) = \begin{cases} 1 & \Lambda = \log n + k\log\log n + w(n), \\ 0 & \Lambda = \log n + (k-2)\log\log n - w(n), \end{cases}$$

The first point of similarity to note between Theorem 4.8.1 and Theorem 4.9.1 is that whilst neither Theorem identifies a sharp threshold, both identify the first order term for the transition to homological connectivity: $(2-\frac{2}{d})\log n$ and $\log n$ respectively. It is worth noting how the geometric differences between building a Čech complex on a manifold with boundary, rather than a closed manifold, inform the differences in the coefficients of the terms in the homological connectivity thresholds.

Our analysis shows that the distance function of a Poisson point process yields a large number of critical points near to the boundary, and this results in the disparity between the constant factor of the first order terms. For a collection of points $\mathcal{Y}$ of the Poisson process to induce a critical point at the centre of these points $c(\mathcal{Y})$, one requires that no other point of the Poisson point process lies in the ball of radius $\rho(\mathcal{Y})$ centred at $c(\mathcal{Y})$. The existence of critical points near to the boundary is made more likely by the fact that the ball of radius $\rho(\mathcal{Y})$ centred at $c(\mathcal{Y})$ is cut by the boundary, and so it is more likely that no other point of the Poisson point process lies in this cut ball. In the most extreme case the ball's volume is cut in half by the boundary and this introduces the factor of 2 in the leading term of the threshold. Since this phenomenon occurs only for critical points near to the boundary we have to scale our count by the volume of an r-collar neighbourhood of the boundary. This volume behaves like $\mathrm{Vol}(\partial M_r) \sim r \sim (\frac{\log n}{n})^{\frac{1}{d}}$ since $\Lambda = \omega_d n r^d \sim \log n$, and in particular the factor $n^{-\frac{1}{d}}$ introduces the term $-\frac{1}{d}$ to the leading coefficient in Theorem 4.8.1.

The second order disparity is identified to lie in the range $[(k-2)\log\log n, k\log\log n]$ by Theorem 4.9.1 and in the range $[2(\frac{1}{d}-3)\log\log n, 2k\log\log n]$ by Theorem 4.8.1. The factor of 2 in the second order terms is again induced by the effect of the boundary cutting volumes in half. The second order term of the lower threshold is affected by counting special erroneous cycles which occur near the boundary which we call Θ-like-cycles. Since we only count points near the boundary, again a $\frac{1}{d}$ term is introduced to the coefficient of the second order term. The coefficient of the second order term is further impacted by a lower bound for the volume of a subset of a Grassmannian, and may be able to be improved if this bound is sharpened.

Relative homology. Here we studied the homological connectivity of a compact Riemannian manifold with boundary in terms of the absolute homology. Equally we could have chosen to study the homology of the manifold relative to its boundary. We could perform the same analysis to attain connectivity thresholds for the relative homology by counting critical points of a Morse function induced by the Point process. Recall that the Morse complex of the Morse function recovers absolute homology if our Morse function attains a maximum on the boundary, and relative homology if our Morse function attains a minimum on the boundary (Theorem 4.3.4). Thus the negative of the distance function from the point process would induce a Morse complex which calculates the relative homology.

Taking the negative of the distance function converts index k critical points into index $d-k$ critical points. With this setup we lose the geometric interpretation that the Morse complex at scale r computes the homology of the union of radius r balls built around the point process. This correspondence is a result analogous to Lefschetz Duality combined with the Universal Coefficient Theorem for Cohomology for calculating the relative homology, which makes it clear that given $H_k(M)$ and $H_{k-1}(M)$ we can compute $H_{d-k}(M, \partial M)$.

There is a geometric interpretation for the dual to our Θ-like-cycles when we consider the Morse complex of the negative of the distance function. For a Θ-like-cycle of index k we require a k-simplex to be introduced approximately tangential to the boundary. The corresponding orthogonal dual $(d-k)$-simplex crosses the part of the partial annulus cut by the boundary and introduces a new homological cycle when we take homology relative to the boundary. See Figure 4.4, for an illustration of such a dual simplex which introduces a relative cycle.

Appendix

Upper Threshold Proofs

Proof. (Proposition 4.6.1)

Let us define $\hat{\beta}_k(r) = \beta_k(r) - \beta_k(M)$ then use Lemma 4.6.2 to bound $\mathbb{E}[\hat{\beta}_k(r)]$. Let's condition on the event $E = \{M \subset B_{r_0}(\mathcal{P}_n)\}$ that we cover our manifold with the radius r_0 neighbourhood of our Poisson Process. Conditioning on E we get:

$$\mathbb{E}[\hat{\beta}_k(r)] = \mathbb{E}[\hat{\beta}_k(r)|E]\mathbb{P}(E) + \mathbb{E}[\hat{\beta}_k(r)|E^c]\mathbb{P}(E^c)$$

We can bound the term $\mathbb{E}[\hat{\beta}_k(r)|E]\mathbb{P}(E)$ by observing that any non-trivial k-cycle in $\mathcal{C}_r$ that is trivial in M must be annihilated by a $(k+1)$-critical point in $C^{\rho_M}_{k+1}(r, r_0)$. Thus assuming $\Lambda \to \infty$, $r = o(r_0)$ and $\Lambda_{r_0} r_0^2 \to 0$, using Lemma 4.6.2 we may bound this term by $\mathbb{E}[|C^{\rho_M}_{k+1}(r, r_0)|] = O(n^{1-\frac{1}{d}}\Lambda^k e^{-\frac{1}{2}\Lambda} + n\Lambda^k e^{-\Lambda})$

The second term, $\mathbb{E}[\hat{\beta}_k(r)|E^c]\mathbb{P}(E^c)$ can be bounded using an $\frac{r_0}{2}$-net to bound the non-coverage probability. Let us begin by bounding $\mathbb{E}[\beta_k(r)|E^c]\mathbb{P}(E^c)$. A coarse upper bound for $\beta_k(r)$ is the number of k-dimensional faces of $\mathcal{C}_r$:

$$\begin{aligned}
\mathbb{E}[\beta_k(r)|E^c]\mathbb{P}(E^c) &\leq \mathbb{E}\left[\binom{|\mathcal{P}_n|}{k+1}|E^c\right]\mathbb{P}(E^c) \\
&= \sum_{k+1}^{\infty}\binom{m}{k+1}\mathbb{P}(|\mathcal{P}_n| = m \mid E^c)\mathbb{P}(E^c) \\
&= \sum_{k+1}^{\infty}\binom{m}{k+1}\mathbb{P}(E^c \mid |\mathcal{P}_n| = m)\mathbb{P}(|\mathcal{P}_n| = m)
\end{aligned}$$

Since $\mathcal{P}_n$ is a Poisson process of intensity n then $\mathbb{P}(|\mathcal{P}_n| = m) = \frac{e^{-n}n^m}{m!}$ and conditioned on $\{|\mathcal{P}_n| = m\}$ we may write $\mathcal{P}_n$ as a set of m independent uniformly distributed random variables $\chi_m = \{X_1, ..., X_m\}$

$$\mathbb{P}(E^c \mid |\mathcal{P}_n| = m) = \mathbb{P}(B_{r_0}(\chi_m) \subsetneq M)$$

Let $\mathcal{N}$ be a $\frac{r_0}{2}$-net for M chosen such that $|\mathcal{N}| \leq c_d r_0^{-d}$ for some constant c_d dependent only on the dimension of our manifold and the metric, and such that at least $\frac{1}{2}$ of the volume of radius $\frac{r_0}{2}$ ball lies in M. We then use this net to bound the non-coverage probability:

$$\begin{aligned}
\mathbb{P}(B_{r_0}(\chi_m) \subsetneq M) &\leq \sum_{x\in\mathcal{N}}\mathbb{P}\left(\rho_{\chi_m}(x) > \frac{r_0}{2}\right) \\
&\leq c_d r_0^{-d}\max_{x\in\mathcal{N}}\{(1 - \mathrm{Vol}(B_{\frac{r_0}{2}}(x)))^m\} \\
&\leq c_d r_0^{-d}(1 - \omega_d 2^{-(d+1)}(1 - s_{\max}r_0^2)r_0^d)^m \\
&\leq c_d r_0^{-d}(1 - \kappa r_0^d)^m
\end{aligned}$$

where $s_{\max}$ is a constant intrinsic to the manifold as defined in Lemma 4.2.5. Thus we yield:

$$\begin{aligned}
\mathbb{E}[\beta_k(r)|E^c]\mathbb{P}(E^c) &\leq \sum_{m=k+1}^{\infty} \binom{m}{k+1} \mathbb{P}(E^c \mid |\mathcal{P}_n| = m)\mathbb{P}(|\mathcal{P}_n| = m) \\
&\leq \sum_{m=k+1}^{\infty} \binom{m}{k+1} c_d r_0^{-d}(1-\kappa r_0^d)^m \frac{e^{-n}n^m}{m!} \\
&\leq r_0^{-d} n^{k+1}(1-\kappa r_0^d)^{k+1} e^{-\kappa n r_0^d} \sum_{j=0}^{\infty} \frac{c_d}{(k+1)!} \frac{e^{-n(1-\kappa r_0^d)} n^j (1-\kappa r_0^d)^j}{j!} \\
&\leq C r_0^{-d} n^{k+1}(1-\kappa r_0^d)^{k+1} e^{-\kappa n r_0^d} \leq C r_0^{-d} n^{k+1} e^{-\kappa n r_0^d}.
\end{aligned}$$

Calculating the asymptotic behaviour of $r_0^{-d} n^{k+1} e^{-\kappa n r_0^d} / n\Lambda^k e^{-\Lambda}$ given the condition that $r_0 \geq r(\frac{\omega_d}{\kappa}(1+|\log r|))^{1/d}$ we see:

$$\begin{aligned}
\frac{r_0^{-d} n^{k+1} e^{-\kappa n r_0^d}}{n\Lambda^k e^{-\Lambda}} &= \frac{n^k e^{\Lambda - \kappa n r_0^d}}{r_0^d \Lambda^k} \leq \frac{n^k e^{-|\log r|\Lambda}}{r^d \Lambda^k (\frac{\omega_d}{\kappa}(1+|\log r|))} \\
&= \frac{n^k r^{\Lambda - d}}{\Lambda^k (\frac{\omega_d}{\kappa}(1+|\log r|))} = \frac{n^k r^{\Lambda - d}}{n^k r^{dk} \omega_d^k (\frac{\omega_d}{\kappa}(1+|\log r|))} = \frac{r^{\Lambda - d - dk}}{\omega_d^k (\frac{\omega_d}{\kappa}(1+|\log r|))} \to 0.
\end{aligned}$$

Hence $\mathbb{E}[\beta_k(r)|E^c]\mathbb{P}(E^c) = o(n\Lambda^k e^{-\Lambda})$. Then we note that $\mathbb{P}(E^c) = o(n\Lambda^k e^{-\Lambda})$ by an almost identical argument and so $\mathbb{E}[\hat{\beta}_k(r)|E^c]\mathbb{P}(E^c) = \mathbb{E}[\beta_k(r)|E^c]\mathbb{P}(E^c) - \beta_k(M)\mathbb{P}(E^c) = o(n\Lambda^k e^{-\Lambda})$. □

Lower Threshold Proofs

Proof. (Lemma 4.7.8)

We may assume $\mathcal{P}$ to be generic in the sense that the critical values of each simplex are distinct, so there is some $\rho_- < \rho$ such that $\mathcal{C}(\mathcal{P}, \rho) = \mathcal{C}(\mathcal{P}, \rho_-) \cup \Delta$ where Δ is the k-simplex $\mathcal{Y}$. Moreover the boundary $\partial\Delta \in \mathcal{C}(\mathcal{P}, \rho_-)$, and Δ is not the face of any higher simplex in $\mathcal{C}(\mathcal{P}, \rho)$ by the construction of the Čech complex.

Suppose that $\partial\Delta$ is a boundary in $\mathcal{C}(\mathcal{P}, \rho_-)$ so that $\partial\Delta = \partial\gamma$ for some k chain $\gamma \in \mathcal{C}(\mathcal{P}, \rho_-)$. Then clearly $\Delta - \gamma$ is a k-cycle in $\mathcal{C}(\mathcal{P}, \rho)$. However $\Delta - \gamma$ is not a boundary or homologous to another cycle since Δ is not the face of any higher simplex in $\mathcal{C}(\mathcal{P}, \rho)$, and so we introduce a new non-trivial k-cycle.

Thus it suffices to show that $\partial\Delta$ is indeed a boundary. Consider the natural map from the simplicial chains to the singular chains $\iota_k : C_k^{\Delta}(\mathcal{C}(\mathcal{P}, \rho_-)) \to C_k^{\text{sing}}(B_{\rho_-}(\mathcal{P}))$. For sufficiently small ρ_-, by the Nerve Lemma the induced map on homology is an isomorphism $h_k : H_k(\mathcal{C}(\mathcal{P}, \rho_-)) \cong H_k(B_{\rho_-}(\mathcal{P}))$.

If ρ_- is sufficiently close to ρ then given that $A_{\psi}^{(\varphi)}(\mathcal{Y}) \subset B_{\rho}(\mathcal{P})$ we also have $A_{2\psi}^{(\varphi)}(\mathcal{Y}) \subset B_{\rho_-}(\mathcal{P})$. Our ψ is constructed in order

that $\iota_{k-1}(\partial\Delta) \in C_{k-1}^{\text{sing}}(A_{2\psi}^{(\varphi)}) \subset C_{k-1}^{\text{sing}}(B_{\rho_-}(\mathcal{P}))$. $A_{2\psi}^{(\varphi)}$ is homotopic to a d-dimensional annulus sliced by a $d-1$-hyperplane and so $H_{k-1}(A_{2\psi}^{(\varphi)}) = 0$, and thus $h_{k-1}(\partial\Delta) = 0$ □

4.10 *Second Moment Calculations*

We have attained a lower bound for the expected number of Θ-like-cycles near the boundary of the manifold. Let us now seek to show that w.h.p. the number of Θ-like-cycles is bounded below by this regime. Some of the notation used in this section is defined in Section 4.7. Let us make the abbreviation $\Theta_k^{\varepsilon,\partial M} = \Theta_k^{\varepsilon,\partial M}(r_1, r)$. Using Chebyshev's inequality we observe that:

$$\mathbb{P}(\Theta_k^{\varepsilon,\partial M} \leq \gamma\mathbb{E}[\Theta_k^{\varepsilon,\partial M}]) \leq \frac{\text{Var}(\Theta_k^{\varepsilon,\partial M})}{(1-\gamma)^2\mathbb{E}[\Theta_k^{\varepsilon,\partial M}]^2}$$

It suffices to show that the right-hand-side goes to zero. In bounding the variance we will be interested in the following term:

$$\left(\Theta_k^{\varepsilon,\partial M}(r_1, r)\right)^2 = \sum_{\substack{\mathcal{Y}_1,\mathcal{Y}_2\subset\mathcal{P}_n \\ |\mathcal{Y}_i|=k+1}} g_r^{\varepsilon,\delta}(\mathcal{Y}_1, \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2, \mathcal{P}_n)$$

Let us separate the distinct situations in which the centres induced by $\mathcal{Y}_i$ are either close together T_2, or far apart T_1. Let us define:

$$\Phi_r(\mathcal{Y}_1, \mathcal{Y}_2) = \mathbf{1}\{B_r(c(\mathcal{Y}_1) \cap B_r(c(\mathcal{Y}_2) = \emptyset\}$$

$$T_1 = \Theta_k^{\varepsilon,\partial M}(r_1, r)^2\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2) = \sum_{\substack{\mathcal{Y}_1,\mathcal{Y}_2\subset\mathcal{P}_n \\ |\mathcal{Y}_i|=k+1}} g_r^{\varepsilon,\delta}(\mathcal{Y}_1, \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2, \mathcal{P}_n)\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)$$

$$T_2 = \Theta_k^{\varepsilon,\partial M}(r_1, r)^2(1-\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) = \sum_{\substack{\mathcal{Y}_1,\mathcal{Y}_2\subset\mathcal{P}_n \\ |\mathcal{Y}_i|=k+1}} g_r^{\varepsilon,\delta}(\mathcal{Y}_1, \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2, \mathcal{P}_n)(1-\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2))$$

Thus we may express the variance as:

$$\text{Var}(\Theta_k^{\varepsilon,\partial M}) = \mathbb{E}[(\Theta_k^{\varepsilon,\partial M})^2] - \mathbb{E}[\Theta_k^{\varepsilon,\partial M}]^2 = (\mathbb{E}[T_1] - \mathbb{E}[\Theta_k^{\varepsilon,\partial M}]^2) + \mathbb{E}[T_2]$$

Using Palm Theory Theorem 4.2.4 to simplify the sums taken over $(k+1)$-sized subsets we obtain:

$$\mathbb{E}[\Theta_k^{\varepsilon,\partial M}]^2 = \frac{n^{2k+2}}{((k+1)!)^2}\mathbb{E}[g_r^{\varepsilon,\delta}(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n')]$$

$$\mathbb{E}[T_1] = \frac{n^{2k+2}}{((k+1)!)^2}\mathbb{E}[g_r^{\varepsilon,\delta}(\mathcal{Y}_1', \mathcal{Y}' \cup \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2', \mathcal{Y}' \cup \mathcal{P}_n)]$$

Where $\mathcal{P}_n, \mathcal{P}_n'$ are i.i.d Poisson processes, $\mathcal{Y}_1', \mathcal{Y}_2' \subset M$ are i.i.d, and $\mathcal{Y}' = \mathcal{Y}_1' \cup \mathcal{Y}_2'$. Notice that when $\Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2') \neq 0$ then $g_r^{\varepsilon,\delta}(\mathcal{Y}_i', \mathcal{Y}' \cup$

$\mathcal{P}_n) = g_r^{\varepsilon,\delta}(\mathcal{Y}_i', \mathcal{Y}_i' \cup \mathcal{P}_n)$ since $g_r^{\varepsilon,\delta}$ is is only dependent on the points in the second argument which are sufficiently close to the points in the first argument. Thus we can make the following simple bound by omitting the negative contribution of the third expectation term in the first expression:

$$\begin{aligned}
\mathbb{E}[T_1] - \mathbb{E}[\Theta_k^{\varepsilon,\partial M}]^2 = \frac{n^{2k+2}}{((k+1)!)^2}\Big(&\mathbb{E}[g_r^{\varepsilon,\delta}(\mathcal{Y}_1', \mathcal{Y}' \cup \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2', \mathcal{Y}' \cup \mathcal{P}_n) B_r(\mathcal{Y}_1')] \\
&- \mathbb{E}[g_r^{\varepsilon,\delta}(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n') \Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2')] \\
&- \mathbb{E}[g_r^{\varepsilon,\delta}(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n') (1 - \Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2'))]\Big) \\
\leq \frac{n^{2k+2}}{((k+1)!)^2}\Big(&\mathbb{E}[g_r^{\varepsilon,\delta}(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n) \Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2')] \\
&- \mathbb{E}[g_r^{\varepsilon,\delta}(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n') \Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2')]\Big) \\
=: \frac{n^{2k+2}}{((k+1)!)^2}&\mathbb{E}[\Delta g_r^{\varepsilon,\delta}]
\end{aligned}$$

Let us condition on the subsets $\mathcal{Y}_1', \mathcal{Y}_2'$ and note that whenever $\Delta g_r^{\varepsilon,\delta} \neq 0$ these two subsets are sufficiently separated so as not to interact. In particular note that:

$$\begin{aligned}
&\Delta g_r^{\varepsilon,\delta} = h_r^{\varepsilon,\delta}(\mathcal{Y}_1') h_r^{\varepsilon,\delta}(\mathcal{Y}_2') \mathbf{1}\{B_{2r}(\mathcal{Y}_1') \cap B_{2r}(\mathcal{Y}_2') = \emptyset\} \\
&\times \Big(\mathbf{1}\{B_{r_2}(\mathcal{Y}_1') \cap \mathcal{P}_n = \emptyset\} \mathbf{1}\{A_\varepsilon^{(\varphi)}(c(\mathcal{Y}_1')) \subset B_{\rho(\mathcal{Y}_1')}(\mathcal{P}_n)\} \mathbf{1}\{B_{r_2}(\mathcal{Y}_2') \cap \mathcal{P}_n = \emptyset\} \mathbf{1}\{A_\varepsilon^{(\varphi)}(c(\mathcal{Y}_2')) \subset B_{\rho(\mathcal{Y}_2')}(\mathcal{P}_n)\} \\
&- \mathbf{1}\{B_{r_2}(\mathcal{Y}_1') \cap \mathcal{P}_n = \emptyset\} \mathbf{1}\{A_\varepsilon^{(\varphi)}(c(\mathcal{Y}_1')) \subset B_{\rho(\mathcal{Y}_1')}(\mathcal{P}_n)\} \mathbf{1}\{B_{r_2}(\mathcal{Y}_2') \cap \mathcal{P}_n' = \emptyset\} \mathbf{1}\{A_\varepsilon^{(\varphi)}(c(\mathcal{Y}_2')) \subset B_{\rho(\mathcal{Y}_2')}(\mathcal{P}_n')\}\Big)
\end{aligned}$$

and so if $\Delta g_r^{\varepsilon,\delta} \neq 0$ then $B_{2r}(\mathcal{Y}_1') \cap B_{2r}(\mathcal{Y}_2') = \emptyset$. Hence conditioned on $\mathcal{Y}_1', \mathcal{Y}_2'$ we see that $B_{r_2}(\mathcal{Y}_2') \cap \mathcal{P}_n$ and $A_\varepsilon^{(\varphi)}(c(\mathcal{Y}_2')) \subset B_{\rho(\mathcal{Y}_2')}(\mathcal{P}_n)$ are independent of $B_{r_2}(\mathcal{Y}_1') \cap \mathcal{P}_n$ and $A_\varepsilon^{(\varphi)}(c(\mathcal{Y}_1')) \subset B_{\rho(\mathcal{Y}_1')}(\mathcal{P}_n)$ by the spatial independence property of Poisson point processes. Moreover $B_{r_2}(\mathcal{Y}_2') \cap \mathcal{P}_n$ has the same distribution as $B_{r_2}(\mathcal{Y}_1') \cap \mathcal{P}_n'$ and $A_\varepsilon^{(\varphi)}(c(\mathcal{Y}_1')) \subset B_{\rho(\mathcal{Y}_1')}(\mathcal{P}_n)$ has the same distribution as $A_\varepsilon^{(\varphi)}(c(\mathcal{Y}_1')) \subset B_{\rho(\mathcal{Y}_1')}(\mathcal{P}_n')$.

Thus the terms cancel upon taking expectation and

$$\mathbb{E}_{\mathcal{P}_n, \mathcal{P}_n', \mathcal{Y}_1', \mathcal{Y}_2'}[\Delta g_r^{\varepsilon,\delta}] = \mathbb{E}_{\mathcal{P}_n, \mathcal{P}_n', \mathcal{Y}_1', \mathcal{Y}_2'}[\mathbb{E}_{\mathcal{P}_n, \mathcal{P}_n'}[\Delta g_r^{\varepsilon,\delta} | \mathcal{Y}_1', \mathcal{Y}_2']] = 0.$$

It remains to bound the second term of the variance $\mathbb{E}[T_2]$. Let us split the term T_2 into the separate cases in which $\mathcal{Y}_1, \mathcal{Y}_2$ share j points:

$$T_2 = \sum_{j=0}^{k+1} \sum_{|\mathcal{Y}_1 \cap \mathcal{Y}_2| = j} g_r^{\varepsilon,\delta}(\mathcal{Y}_1, \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2, \mathcal{P}_n)(1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) = \sum_{j=0}^{k+1} I_j$$

We next require a Corollary of Theorem 4.2.4.

Corollary 4.10.1. *Let (X, ρ) be a metric space, $f : X \to \mathbb{R}$ a probability density and $\mathcal{P}_n$ a Poisson process on X with intensity $\lambda_n = nf$. If $h(\mathcal{Y}, \mathcal{X})$ is a measurable function for all finite subsets $\mathcal{Y}, \mathcal{X} \subset X$, with $\mathcal{Y} \subset \mathcal{X}$ and $|\mathcal{Y}| = k+1$ then:*

$$\mathbb{E}\left[\sum_{\substack{|\mathcal{Y}_1| = |\mathcal{Y}_2| = k+1, \\ |\mathcal{Y}_1 \cap \mathcal{Y}_2| = j}} h(\mathcal{Y}_1, \mathcal{P}_n) h(\mathcal{Y}_2, \mathcal{P}_n) \right] = \frac{n^{2k-j}}{j!(k-j)!} \mathbb{E}[h(\mathcal{Y}_1', \mathcal{Y}' \cup \mathcal{P}_n) h(\mathcal{Y}_2', \mathcal{Y}' \cup \mathcal{P}_n)]$$

where $\mathcal{Y}' = \mathcal{Y}_1' \cup \mathcal{Y}_2'$ is a set of $2k - j$ i.i.d points in X with density f, with $|\mathcal{Y}_1' \cap \mathcal{Y}_2'| = j$ and $\mathcal{Y}'$ independent of $\mathcal{P}_n$.

Using the result given in Corollary 4.10.1 yields:

$$\begin{aligned}
\mathbb{E}[I_j] &= \sum_{|\mathcal{Y}_1 \cap \mathcal{Y}_2| = j} g_r^{\varepsilon,\delta}(\mathcal{Y}_1, \mathcal{P}_n) g_r^{\varepsilon,\delta}(\mathcal{Y}_2, \mathcal{P}_n)(1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) \\
&\leq \frac{n^{2k+2-j}}{j!((k+1-j)!)^2} \int_{DM^{2k+2-j}} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-n\mathrm{Vol}(\mathbf{y}_1, \mathbf{y}_2)} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) |\, \mathrm{dvol}_g(\mathbf{y})|
\end{aligned}$$

We bound the volume term using the results of Section 4.2.4:

$$\mathrm{Vol}(\mathbf{y}_1, \mathbf{y}_2) = \mathrm{Vol}(B_r(c(\mathbf{y}_1) \cup B_r(c(\mathbf{y}_2)) \geq (\frac{1}{2} + \alpha)(1 - (dv'r + \nu r^2))(1 + \frac{\omega_{d-1} d(c_1, c_2)}{\omega_d r} + O(\frac{d(c_1, c_2)^2}{r}))\omega_d r^d$$

where $\alpha = O(\frac{\delta}{r}) + O(r)$.

Next we require a multivariable Blaschke-Petkantschin change of variable formula in order to bound the expectations $\mathbb{E}[I_j]$. We show how to bound a change of variable formula when integrating over two variables in M^{k+1} where M has non-empty boundary. This formula is already used (without proof) in (79) in the case where M is closed.

Lemma 4.10.2. *Let M be a compact Riemannian manifold with non-empty boundary and let $\mathbf{y}_1, \mathbf{y}_2 \in M^{k+1}$. Denote the respective centres by c_1, c_2. Define*

$$\Omega := \left\{ (\mathbf{y}_1, \mathbf{y}_2) \in \left(M^{k+1}\right)^2 \mid a \leq \rho_M(c_1, c_2) \leq b \right\}.$$

Then there is some constant C_M dependent solely on M for which the following bound holds:

$$\int_\Omega f_1(\mathbf{y}_1) f_2(\mathbf{y}_2) \mathbf{1}_r(\mathbf{y}_1, \mathbf{y}_2) \left| \operatorname{dvol}_g(\mathbf{y}_1, \mathbf{y}_2) \right| \leq$$
$$C_M \int_M \left| \operatorname{dvol}_g(c_1) \right| \int_a^b ds \int_{\mathbb{S}_1(T_{c_1}M)} s^{d-1} \left| \operatorname{dvol}_{\mathbb{S}_1(T_{c_1}M)}(w) \right|$$
$$\times \prod_{i=1}^{2} \int_0^r du_i u_i^{kd-1} \int_{\mathrm{Gr}(k,d)} d\mu_{k,d}(V) \int_{\mathbb{S}_1(V)^{k+1}} \left| \operatorname{dvol}_{\mathbb{S}_1(V)^{k+1}}(\mathbf{w}_i) \right| f_i(\exp_{c_i}(u_i \mathbf{w}_i)).$$

Proof. We use the Blaschke-Petkantschin formula for integrals over one variable in M^{k+1} to attain the result. Given $\mathbf{y}_1 \in M^{k+1}$ and c_1 the induced center, let

$$\Omega(\mathbf{y}_1) := \left\{ \mathbf{y}_2 \in M^{k+1} \mid a \leq \rho_M(c_1, c_2) \leq b \right\}.$$

Denote the above integral on the LHS by I, we have

$$I = \int_{M^{k+1}} f_1(\mathbf{y}_1) \mathbf{1}_r(\mathbf{y}_1) \left(\int_{\Omega(\mathbf{y}_1)} f_2(\mathbf{y}_2) \mathbf{1}_r(\mathbf{y}_2) \left| \operatorname{dvol}_g(\mathbf{y}_2) \right| \right) \left| \operatorname{dvol}_g(\mathbf{y}_1) \right|.$$

We first compute the inner integral, for fixed $c_1 \in M$.

Note that

$$\int_{\Omega(\mathbf{y}_1)} f_2(\mathbf{y}_2) \mathbf{1}_r(\mathbf{y}_2) \left| \operatorname{dvol}_g(\mathbf{y}_2) \right| = \int_M f_2(\mathbf{y}_2) \mathbf{1}_r(\mathbf{y}_2) \mathbf{1}\{c_2 \in A_a^b(c_1)\} \left| \operatorname{dvol}_g(\mathbf{y}_2) \right|,$$

where $A_a^b(c_1) := B_b(c_1) \setminus B_a(c_1)^o$.

By the Blaschke Petkantschin formula for manifolds with non-empty boundary, we then find

$$\int_{\Omega(\mathbf{y}_1)} f_2(\mathbf{y}_2) \mathbf{1}_r(\mathbf{y}_2) \left| \operatorname{dvol}_g(\mathbf{y}_2) \right|$$
$$= \int_{A_a^b(c_1)} \left| \operatorname{dvol}_g(c_2) \right| \int_0^r du_2 \int_{\mathrm{Gr}(k, T_c DM)} d\mu_{k,d}(V)$$
$$\times \left(\int_{(\mathbb{S}_u(V) \cap \exp_{c_2}^{-1}(M))^{k+1}} \left| \operatorname{dvol}_{\mathbb{S}_u(V)^{k+1}}(\mathbf{v}_2) \right| f_2(\exp_{c_2}(\mathbf{v}_2)) \Upsilon_u^{d-k}(\mathbf{v}_2) \prod_{j=1}^{k+1} \sqrt{\det(g_{\exp_{c_2}(v_j)})} \right)$$
$$\leq C \int_{A_a^b(c_1)} \left| \operatorname{dvol}_g(c_2) \right| \int_0^r du_2 u_2^{dk-1} \int_{\mathrm{Gr}(k,d)} d\mu_{k,d}(V)$$
$$\times \left(\int_{\mathbb{S}_1(V)^{k+1}} \left| \operatorname{dvol}_{\mathbb{S}_1(V)^{k+1}}(\mathbf{w}_2) \right| f_2(\exp_{c_2}(u_2 \mathbf{w}_2)) \Upsilon_1^{d-k}(\mathbf{w}_2) \prod_{j=1}^{k+1} \sqrt{\det(g_{\exp_{c_2}(u_2 w_j)})} \right).$$

Using the compactness of M, the above is

$$\leq C \int_{A_a^b(c_1)} \left| \operatorname{dvol}_g(c_2) \right| \int_0^r du_2 u_2^{dk-1} \int_{\mathrm{Gr}(k,d)} d\mu_{k,d}(V) \int_{\mathbb{S}_1(V)^{k+1}} \left| \operatorname{dvol}_{\mathbb{S}_1(V)^{k+1}}(\mathbf{w}_2) \right| f_2(\exp_{c_2}(u_2 \mathbf{w}_2)).$$

Furthermore, using the Riemannian approximation results and polar decomposition, we have

$$\begin{aligned}\int_{A_a^b(c_1)} \left| \operatorname{dvol}_g(c_2) \right| &\leq C \int_a^b ds \int_{\mathrm{S}_s(T_{c_1}DM)} \left| \operatorname{dvol}_{\mathrm{S}_s(T_{c_1}DM)}(w) \right| \\ &= \int_a^b ds \int_{\mathrm{S}_1(T_{c_1}DM)} s^{d-1} \left| \operatorname{dvol}_{\mathrm{S}_1(T_{c_1}DM)}(w) \right|.\end{aligned}$$

The outer integral in the expression of I is estimated again with the Blaschke-Petkantschin formula for manifolds with non-empty boundary. Doing so and combining the result with the above expression for the inner integral of I yields the claimed formula. □

We now make two separate change of variables separating the cases where $j = 0$ and $j \neq 0$. We take the centre $c(\mathbf{y}_2)$ to be in polar coordinates around $c(\mathbf{y}_1)$. For details see Lemma 4.10.2 and Section 4.5 on Blaschke Petkantshin Formulae. Let us further partition our integral into two regions Ω_a, Ω_b:

$$\Omega_a = \{(\mathbf{y}_1, \mathbf{y}_2) : 0 \leq \frac{\rho(c(\mathbf{y}_1), c(\mathbf{y}_2))}{r} \leq \epsilon\}$$
$$\Omega_b = \{(\mathbf{y}_1, \mathbf{y}_2) : \epsilon \leq \frac{\rho(c(\mathbf{y}_1), c(\mathbf{y}_2))}{r} \leq 4\}$$

Under the assumption that $\Lambda r \to 0$ and $r_1 = (1 - \frac{1}{2c_g^2\Lambda^2})r$ we attain:

$$\begin{aligned}\mathbb{E}[I_0^{(a)}] &\leq \frac{n^{2k+2}}{((k+1)!)^2} \int_{\Omega_a} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-n\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) | \operatorname{dvol}_g(\mathbf{y})| \\ &\leq Cn^{2k+2} \int_{\Omega_a} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-\frac{\Lambda}{2}(1-(d\eta' r+\eta r^2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) | \operatorname{dvol}_g(\mathbf{y})| \\ &\leq Cn^{2k+2} e^{-\frac{\Lambda}{2}} \int_{\partial M_{[\delta,2\delta]}} | \operatorname{dvol}_g(c_1)| \int_0^{\epsilon r} ds \int_{\mathrm{S}_1(T_{c_1}M)} s^{d-1} \operatorname{dvol}_{\mathrm{S}_1(T_{c_1}M)}(w) \\ &\times \prod_{i=1}^{2} \int_{r_1}^{r} du_i \int_{\mathrm{Gr}(k,d)} u_i^{k(d-k)} d\mu_{k,d}(V) \int_{(\mathrm{S}_1(V))^{k+1}} u_i^{(k-1)(k+1)} | \operatorname{dvol}_{(\mathrm{S}_1(V))^{k+1}}(\mathbf{w}_i) | h_r^{\varepsilon,\delta}(\exp_{c_i}(u_1 \mathbf{w}_i)) \\ &\leq Cn^{2k+2} e^{-\frac{\Lambda}{2}} \delta \epsilon^d r^{d(2k+1)} \Lambda^{-4} \leq Cne^{-\frac{\Lambda}{2}} \Lambda^{2k-3} \epsilon^d r\end{aligned}$$

$$\begin{aligned}\mathbb{E}[I_0^{(b)}] &\leq \frac{n^{2k+2}}{((k+1)!)^2} \int_{\Omega_b} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-n\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) | \operatorname{dvol}_g(\mathbf{y})| \\ &\leq Cn^{2k+2} \int_{\Omega_b} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-\frac{\Lambda}{2}(1+\frac{\epsilon\omega_d}{\omega_{d-1}})(1-(d\eta' r+\eta r^2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) | \operatorname{dvol}_g(\mathbf{y})| \\ &\leq Cn^{2k+2} e^{-\frac{\Lambda}{2}(1+\frac{\epsilon\omega_d}{\omega_{d-1}})} \int_{\partial M_{[\delta,2\delta]}} | \operatorname{dvol}_g(c_1)| \int_{\epsilon r}^{4r} ds \int_{\mathrm{S}_1(T_{c_1}M)} s^{d-1} \operatorname{dvol}_{\mathrm{S}_1(T_{c_1}M)}(w) \\ &\times \prod_{i=1}^{2} \int_{r_1}^{r} du_i \int_{\mathrm{Gr}(k,d)} u_i^{k(d-k)} d\mu_{k,d}(V) \int_{(\mathrm{S}_1(V))^{k+1}} u_i^{(k-1)(k+1)} | \operatorname{dvol}_{(\mathrm{S}_1(V))^{k+1}}(\mathbf{w}_i) | h_r^{\varepsilon,\delta}(\exp_{c_i}(u_1 \mathbf{w}_i)) \\ &\leq Cn^{2k+2} e^{-\frac{\Lambda}{2}(1+\frac{\epsilon\omega_d}{\omega_{d-1}})} \delta r^{d(2k+1)} \Lambda^{-4} \leq Cne^{-\frac{\Lambda}{2}} e^{-\frac{\Lambda\epsilon\omega_d}{2\omega_{d-1}}} \Lambda^{2k-3} r\end{aligned}$$

where $c_2 = \exp_{c_1}(sw)$, $\mathbf{y}_i = \exp_{c_i}(\mathbf{w}_i)$ and $\mathrm{S}_1(V)$ is the unit sphere.

Similarly for $j \neq 0$ we attain bounds:

$$\begin{aligned}
\mathbb{E}[I_j^{(a)}] &\leq \frac{n^{2k+2-j}}{j!((k+1-j)!)^2} \int_{\Omega_a} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-n\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2))| \operatorname{dvol}_g(\mathbf{y})| \\
&\leq n^{2k+2-j} \int_{\Omega_a} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-\frac{\Lambda}{2}(1-(d\eta' r + \eta r^2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2))| \operatorname{dvol}_g(\mathbf{y})| \\
&\leq C n^{2k+2-j} e^{-\frac{\Lambda}{2}} \int_{\partial M_{[\delta,2\delta]}} |\operatorname{dvol}_g(c_1)| \\
&\times \int_{r_1}^{r} du_1 \int_{\mathrm{Gr}(k,d)} u_1^{k(d-k)} d\mu_{k,d}(V_1) \int_{(\mathrm{S}_1(V_1))^{k+1}} u_i^{(k-1)(k+1)} \operatorname{dvol}_{(\mathrm{S}_1(V_1))^{k+1}} \\
&\times \int_0^{\epsilon r} ds \int_{\mathrm{S}_1(T_{c_1}E)} s^{d-j} \operatorname{dvol}_{\mathrm{S}^{d-1}}(w) \int_{r_1}^{r} du_2 \int_{\mathrm{Gr}(k-j,d)} u_i^{(k-j)(d-(k-j))} d\mu_{k-j,d}(W) \\
&\times \int_{(\mathrm{S}_1(V_2))^{k+1-j}} u_2^{(k-1)(k+1-j)} \operatorname{dvol}_{(\mathrm{S}_1(V_2))^{k+1-j}} h_r^{\varepsilon,\delta}(\exp_{c_1}(u_1\mathbf{w}_1)) h_r^{\varepsilon,\delta}(\exp_{c_2}(u_2\mathbf{w}_2)) \\
&\leq C n^{2k+2-j} e^{-\frac{\Lambda}{2}} \delta \epsilon^{d-j+1} r^{d(2k+1-j)+j(k-j)+1} \Lambda^{-4} \leq C n e^{-\frac{\Lambda}{2}} \Lambda^{2k-3} \epsilon^{d-j+1} r^{j(k-j)+1}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[I_j^{(b)}] &\leq \frac{n^{2k+2-j}}{j!((k+1-j)!)^2} \int_{\Omega_b} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-n\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2))| \operatorname{dvol}_g(\mathbf{y})| \\
&\leq n^{2k+2-j} \int_{\Omega_b} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-\frac{\Lambda}{2}(1+\frac{\epsilon\omega_d}{\omega_{d-1}})(1-(d\eta' r + \eta r^2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2))| \operatorname{dvol}_g(\mathbf{y})| \\
&\leq C n^{2k+2-j} e^{-\frac{\Lambda}{2}(1+\frac{\epsilon\omega_d}{\omega_{d-1}})} \int_{\partial M_{[\delta,2\delta]}} |\operatorname{dvol}_g(c_1)| \\
&\times \int_{r_1}^{r} du_1 \int_{\mathrm{Gr}(k,d)} u_1^{k(d-k)} d\mu_{k,d}(V_1) \int_{(\mathrm{S}_1(V_1))^{k+1}} u_i^{(k-1)(k+1)} \operatorname{dvol}_{(\mathrm{S}_1(V_1))^{k+1}} \\
&\times \int_{\epsilon r}^{4r} ds \int_{\mathrm{S}_1(T_{c_1}E)} s^{d-j} \operatorname{dvol}_{\mathrm{S}^{d-1}}(w) \int_{r_1}^{r} du_2 \int_{\mathrm{Gr}(k-j,d)} u_i^{(k-j)(d-(k-j))} d\mu_{k-j,d}(W) \\
&\times \int_{(\mathrm{S}_1(V_2))^{k+1-j}} u_2^{(k-1)(k+1-j)} \operatorname{dvol}_{(\mathrm{S}_1(V_2))^{k+1-j}} h_r^{\varepsilon,\delta}(\exp_{c_1}(u_1\mathbf{w}_1)) h_r^{\varepsilon,\delta}(\exp_{c_2}(u_2\mathbf{w}_2)) \\
&\leq C n^{2k+2-j} e^{-\frac{\Lambda}{2}(1+\frac{\epsilon\omega_d}{\omega_{d-1}})} \delta r^{d(2k+1-j)+j(k-j)+1} \Lambda^{-4} \leq C n e^{-\frac{\Lambda}{2}} e^{-\frac{\Lambda\epsilon\omega_d}{2\omega_{d-1}}} \Lambda^{2k-3} r^{j(k-j)+1}
\end{aligned}$$

We note that our largest bounding function is for the expectation $\mathbb{E}[I_0]$ and so we have the following bound for all j and for arbitrary $\epsilon \in (0,1)$:

$$\mathbb{E}[I_j] \leq C n e^{-\frac{\Lambda}{2}} \Lambda^{2k-3} r (e^{-\frac{\Lambda\epsilon\omega_d}{2\omega_{d-1}}} + \epsilon^d)$$

We may therefore use this same regime to bound $\mathbb{E}[T_2]$. Let us choose $\epsilon = \frac{2(2k-1)\omega_{d-1}}{\omega_d} \frac{\log\log n}{\log n}$. Then for $\Lambda = (2 - \frac{2}{d}) \log n + 2(k - 2 - (k + 1 - \frac{1}{d})) \log\log n - \omega(n)$, (recalling that $\mathbb{E}[\Theta_k^{\varepsilon,\partial M}] = \Omega(e^{-\alpha\Lambda} n \Lambda^{k-2} r (\log n)^{-(k+1)})$ and that $\alpha = \frac{1}{2} + O((\log n)^{-1})$), we

calculate:

$$\frac{\mathrm{Var}(\Theta_k^{\varepsilon,\partial M})}{\mathbb{E}[\Theta_k^{\varepsilon,\partial M}]^2} = \frac{\mathbb{E}[T_2]}{\mathbb{E}[\Theta_k^{\varepsilon,\partial M}]^2} \leq C\frac{ne^{-\frac{\Lambda}{2}}\Lambda^{2k-3}r(e^{-\frac{\Lambda\epsilon\omega_d}{2\omega_{d-1}}}+\epsilon^d)}{n^2\Lambda^{2k-4}e^{-2\alpha\Lambda}r^2(\log n)^{-2(k+1)}} \sim \frac{\Lambda e^{\frac{\Lambda}{2}}(\log n)^{2(k+1)}(e^{-\frac{\Lambda\epsilon\omega_d}{2\omega_{d-1}}}+\epsilon^d)}{nr}$$
$$\sim e^{-\frac{\omega(n)}{2}}(\log n)^{2k}(e^{-\frac{\Lambda\epsilon\omega_d}{2\omega_{d-1}}}+\epsilon^d) \sim e^{-\frac{\omega(n)}{2}}\left(\frac{1}{\log n}+\frac{(\log\log n)^d}{(\log n)^d}\right) \to 0$$

Thus we conclude that w.h.p. $\Theta_k^{\varepsilon,\partial M} \geq \frac{1}{2}\mathbb{E}[\Theta_k^{\varepsilon,\partial M}]$.

List of Symbols

M	Compact Riemannian Manifold with boundary
g	Smooth Riemannian metric
$\rho(\cdot,\cdot)$	Distance induced by Riemannian Metric
τ_M	The reach of the manifold M
∂M_r	An r neighbourhood of the boundary
P	A finite sample of points of a manifold
$\mathcal{P}, \mathcal{Q}$	A Poisson point process
$B_r(P)$	The union of radius r balls centred at each $p \in P$
$\mathcal{C}(n,r)$	Čech complex at radius r on a point process of intensity n
$c(\mathcal{Y}), c(\mathbf{y})$	The centre of a finite collection of points
$\rho(\mathcal{Y}), \rho(\mathbf{y})$	The critical value of a finite collection of points
$B(\mathcal{Y}), B(\mathbf{y})$	The ball centred at $c(\mathcal{Y})$ with radius $\rho(\mathcal{Y})$
$\beta_k(r)$	The k-th Betti number associated to a Čech complex at radius r
ω_d	Volume of a unit radius d-dimensional ball
Λ	Expected number of points lying in an r-ball $\Lambda = \omega_d n r^d$
C	Constant factor, a product of constant terms used to simplify expressions in inequalities
$C_k^{\rho_M}(r, r_0)$	Set of index k critical points with critical values in the range $[r, r_0)$
$\Theta_k^{\varepsilon}(r_1, r)$	Set of index k critical points with critical values in the range $[r_1, r)$ inducing Θ-cycles
$\Theta_k^{\varepsilon,\partial M}(r_1, r)$	Set of index k critical points with critical values in the range $[r_1, r)$ inducing Θ-like-cycles
$A_\varepsilon(c)$	ε-annulus about centre $c(\mathbf{y})$ of radius $\rho(\mathbf{y})$
$A_\varepsilon^{(\varphi)}(c)$	Partial ε-annulus about centre $c(\mathbf{y})$ of radius $\rho(\mathbf{y})$
$\mathbb{S}_u(V)$	Sphere of radius u in the plane V

Bibliography

[1] Agnese Barbensi, et al., "A Topological Selection of Folding Pathways from Native States of Knotted Proteins", *Symmetry*, 13(9), 2021, URL https://www.mdpi.com/2073-8994/13/9/1670.

[2] O Vipond, et al., "Multiparameter persistent homology landscapes identify immune cell spatial patterns in tumors", *Proceedings of the National Academy of Sciences of USA*, 2021, publisher: National Academy of Sciences.

[3] Marcio Gameiro, et al., "A topological measurement of protein compressibility", *Japan Journal of Industrial and Applied Mathematics*, 32(1), 2015, pp. 1–17, URL https://link.springer.com/article/10.1007/s13160-014-0153-5.

[4] Xia Kelin and Wei Guo-Wei, "Multidimensional persistence in biomolecular data", *Journal of Computational Chemistry*, 36(20), 2015, pp. 1502–1520, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.23953.

[5] Monica Nicolau, et al., "Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival", *Proceedings of the National Academy of Sciences of the United States of America*, 108(17), 2011, pp. 7265–7270, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3084136/.

[6] Herbert Edelsbrunner and John Harer, *Computational Topology - an Introduction.*, American Mathematical Society, 2010.

[7] Steve Y. Oudot, *Persistence Theory: From Quiver Representations to Data Analysis*, no. 209 in Mathematical Surveys and Monographs, American Mathematical Society, 2015, URL https://hal.inria.fr/hal-01247501.

[8] Peter Bubenik, et al., "Persistent homology detects curvature", *Inverse Problems*, 2019, URL http://iopscience.iop.org/10.1088/1361-6420/ab4ac0.

[9] Paul Bendich, et al., "Persistent homology analysis of brain artery trees", *Annals of Applied Statistics*, 10(1), 2016, pp. 198–218, URL https://uncch.pure.elsevier.com/en/publications/persistent-homology-analysis-of-brain-artery-trees.

[10] Peter Bubenik, "Statistical topological data analysis using persistence landscapes", *Journal of Machine Learning Research*, 16(1), 2015, pp. 77–102, URL http://jmlr.org/papers/v16/bubenik15a.html.

[11] Peter Bubenik, "The Persistence Landscape and Some of Its Properties", in *Topological Data Analysis - The Abel Symposium 2018*, Springer International Publishing, 2018, section: 4.

[12] Henry Adams, et al., "Persistence Images: A Stable Vector Representation of Persistent Homology", *Journal of Machine Learning Research*, 18(8), 2017, pp. 1–35, URL http://jmlr.org/papers/v18/16-337.html.

[13] Bryn Keller, et al., "PHoS: Persistent homology for virtual screening", *ChemRxiv*, 2018, URL https://chemrxiv.org/articles/PHoS_Persistent_Homology_for_Virtual_Screening/6969260.

[14] Gunnar Carlsson and Afra Zomorodian, "The theory of multidimensional persistence", *Discrete and Computational Geometry*, 42(1), 2009, pp. 71–93.

[15] Patrizio Frosini and Claudia Landi, "Persistent Betti Numbers for a Noise Tolerant Shape-Based Approach to Image Retrieval", in *Computer Analysis of Images and Patterns*, Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, 2011, pp. 294–301, URL https://link.springer.com/chapter/10.1007/978-3-642-23672-3_36.

[16] Andrea Cerri, et al., "Betti numbers in multidimensional persistent homology are stable functions", *Mathematical Methods in the Applied Sciences*, 36(12), 2013, pp. 1543–1557, URL http://doi.wiley.com/10.1002/mma.2704.

[17] Jacek Skryzalin and Gunnar Carlsson, "Numeric invariants from multidimensional persistence", *Journal of Applied and Computational Topology*, 1(1), 2017, pp. 89–119, URL http://link.springer.com/10.1007/s41468-017-0003-z.

[18] Aaron Adcock, et al., "The ring of algebraic functions on persistence bar codes", *Homology, Homotopy and Applications*, 18(1),

2016, pp. 381–402, URL https://doi.org/10.4310/hha.2016.v18.n1.a21.

[19] Sara Kališnik, "Tropical Coordinates on the Space of Persistence Barcodes", *Foundations of Computational Mathematics*, 19(1), 2019, pp. 101–129, URL https://doi.org/10.1007/s10208-018-9379-y.

[20] Heather A. Harrington, et al., "Stratifying Multiparameter Persistent Homology", *SIAM Journal on Applied Algebra and Geometry*, 3(3), 2019, pp. 439–471, URL https://doi.org/10.1137/18M1224350.

[21] Ezra Miller, "Data structures for real multiparameter persistence modules", *arXiv e-prints*, 2017, p. arXiv:1709.08155.

[22] Leo Betthauser, et al., "Graded Persistence Diagrams and Persistence Landscapes", *Discrete & Computational Geometry*, 2021, URL https://doi.org/10.1007/s00454-021-00316-1, publisher: Springer Science and Business Media LLC.

[23] Michael Lesnick and Matthew Wright, "Interactive Visualization of 2-D Persistence Modules", *arXiv:151200180 [cs, math]*, 2015, URL http://arxiv.org/abs/1512.00180, arXiv: 1512.00180.

[24] Peter Bubenik, et al., "Metrics for Generalized Persistence Modules", *Foundations of Computational Mathematics*, 15(6), 2015, pp. 1501–1531, URL https://doi.org/10.1007/s10208-014-9229-5.

[25] Gunnar Carlsson and Afra Zomorodian, "The Theory of Multidimensional Persistence", *Discrete & Computational Geometry*, 42(1), 2009, pp. 71–93, URL https://doi.org/10.1007/s00454-009-9176-0.

[26] Michael Lesnick, "The Theory of the Interleaving Distance on Multidimensional Persistence Modules", *Foundations of Computational Mathematics*, 15(3), 2015, pp. 613–650, URL https://doi.org/10.1007/s10208-015-9255-y.

[27] William Crawley-Boevey, "Decomposition of pointwise finite-dimensional persistence modules", *Journal of Algebra and its Applications*, 14(05), 2015, p. 1550066, URL https://www.worldscientific.com/doi/abs/10.1142/S0219498815500668.

[28] Peter Bubenik, et al., "Metrics for generalized persistence modules", *Foundations of Computational Mathematics*, 15(6), 2015, pp. 1501–1531.

[29] Håvard Bakke Bjerkevik and Magnus Bakke Botnan, "Computational Complexity of the Interleaving Distance", in *34th International Symposium on Computational Geometry (SoCG 2018)*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018, Leibniz International Proceedings in Informatics (LIPIcs), pp. 1–15.

[30] Håvard Bakke Bjerkevik, et al., "Computing the interleaving distance is NP-hard", *Foundations of Computational Mathematics*, 2019, URL https://doi.org/10.1007/s10208-019-09442-y.

[31] Ulrich Bauer and Michael Lesnick, "Induced Matchings of Barcodes and the Algebraic Stability of Persistence", in *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, ACM, New York, NY, USA, 2014, SOCG'14, pp. 355:355–355:364, URL http://doi.acm.org/10.1145/2582112.2582168.

[32] Håvard Bakke Bjerkevik, "On the Stability of Interval Decomposable Persistence Modules", *Discrete & Computational Geometry*, 66(1), 2021, pp. 92–121, URL https://doi.org/10.1007/s00454-021-00298-0, publisher: Springer Science and Business Media LLC.

[33] Michel Ledoux and Michel Talagrand, *Probability in Banach Spaces, Isoperimetry and Processes*, Springer-Verlag Berlin Heidelberg, 2011.

[34] Frédéric Chazal, et al., "Stochastic Convergence of Persistence Landscapes and Silhouettes", *Journal of Computational Geometry*, 6, 2014, pp. 140–161.

[35] Frédéric Chazal, et al., "On the bootstrap for persistence diagrams and landscapes", *Modeling and Analysis of Information Systems*, 20(6), 2015, pp. 111–120, URL https://www.mais-journal.ru/jour/article/view/162.

[36] The RIVET Developers, "RIVET", , 2020, URL https://github.com/rivetTDA/rivet/.

[37] OliverVipond, "Multiparameter Persistence Landscapes", , 2020, URL https://github.com/OliverVipond/Multiparameter_Persistence_Landscapes, original-date: 2020-07-02T15:15:02Z.

[38] Peter Bubenik and Paweł Dłotko, "A persistence landscapes toolbox for topological statistics", *Journal of Symbolic Computation*, 78, 2017, pp. 91–114, URL http://www.sciencedirect.com/science/article/pii/S0747717116300104.

[39] David Cohen-Steiner, et al., "Vines and Vineyards by Updating Persistence in Linear Time", in *Proceedings of the Twenty-second Annual Symposium on Computational Geometry (SoCG 2016)*, ACM, 2006, pp. 119–126, URL http://doi.acm.org/10.1145/1137856.1137877.

[40] Irving J. Good and Ray A. Gaskins, "Density Estimation and Bump-Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data", *Journal of the American Statistical Association*, 75(369), 1980, pp. 42–56, URL http://www.jstor.org/stable/2287377.

[41] B. W. Silverman, "Using Kernel Density Estimates to Investigate Multimodality", *Journal of the Royal Statistical Society: Series B (Methodological)*, 43(1), 1981, pp. 97–99, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1981.tb01155.x.

[42] J. A. Hartigan and P. M. Hartigan, "The Dip Test of Unimodality", *The Annals of Statistics*, 13(1), 1985, pp. 70 – 84, URL https://doi.org/10.1214/aos/1176346577.

[43] N. I. Fisher and J. S. Marron, "Mode Testing via the Excess Mass Estimate", *Biometrika*, 88(2), 2001, pp. 499–517, URL http://www.jstor.org/stable/2673496.

[44] Chul Moon, et al., "Persistence terrace for topological inference of point cloud data.", *Journal of Computational and Graphical Statistics*, 27(3), 2018, pp. 576–586, URL https://doi.org/10.1080/10618600.2017.1422432.

[45] Robert F. Geitz, "Pettis Integration", *Proceedings of the American Mathematical Society*, 82(1), 1981, p. 81, URL http://www.jstor.org/stable/2044321?origin=crossref.

[46] Kazimierz Musiał, *Topics in the theory of Pettis integration*, Universita' degli Studi di Trieste. Dipartimento di Scienze Matematiche, 1991.

[47] Jørgen Hoffmann-Jørgensen and Gilles Pisier, "The Law of Large Numbers and the Central Limit Theorem in Banach Spaces", *The Annals of Probability*, 4(4), 1976, pp. 587–599.

[48] Oliver Vipond, "Multiparameter Persistence Landscapes", *Journal of Machine Learning Research*, 21(61), 2020, pp. 1–38, URL http://jmlr.org/papers/v21/19-054.html.

[49] Michael Kerber, et al., "Exact Computation of the Matching Distance on 2-Parameter Persistence Modules", in *35th International Symposium on Computational Geometry (SoCG 2019)*, edited by Gill Barequet and Yusu Wang, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2019, vol. 129 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 46:1–46:15, URL http://drops.dagstuhl.de/opus/volltexte/2019/10450, iSSN: 1868-8969.

[50] Tamal K Dey and Cheng Xin, "Computing Bottleneck Distance for 2-D Interval Decomposable Modules", *j34th International Symposium on Computational Geometry (SoCG 2018)*, 99, 2018, URL https://par.nsf.gov/biblio/10064706.

[51] Michael Lesnick and Matthew Wright, "Computing Minimal Presentations and Bigraded Betti Numbers of 2-Parameter Persistent Homology", *arXiv:190205708 [cs, math]*, 2019, URL http://arxiv.org/abs/1902.05708, arXiv: 1902.05708.

[52] Håvard Bakke Bjerkevik, et al., "Computing the Interleaving Distance is NP-Hard", *Foundations of Computational Mathematics*, 2019, URL https://doi.org/10.1007/s10208-019-09442-y.

[53] Alexander McCleary and Amit Patel, "Edit Distance and Persistence Diagrams Over Lattices", *arXiv:201007337 [cs, math]*, 2021, URL http://arxiv.org/abs/2010.07337, arXiv: 2010.07337.

[54] René Corbet, et al., "A kernel for multi-parameter persistent homology", *Computers & Graphics: X*, 2, 2019, p. 100005, URL http://www.sciencedirect.com/science/article/pii/S2590148619300056.

[55] Jérémy Cochoy and Steve Oudot, "Decomposition of Exact pfd Persistence Bimodules", *Discrete & Computational Geometry*, 63(2), 2020, pp. 255–293, URL https://doi.org/10.1007/s00454-019-00165-z.

[56] Magnus Bakke Botnan and Michael Lesnick, "Algebraic stability of zigzag persistence modules", *Algebraic & Geometric Topology*, 18(6), 2018, pp. 3133–3204, URL https://projecteuclid.org/euclid.agt/1540605637.

[57] Magnus Bakke Botnan, et al., "On Rectangle-Decomposable 2-Parameter Persistence Modules", in *36th International Symposium on Computational Geometry, SoCG 2020, June 23-26, 2020, Zürich, Switzerland*, edited by Sergio Cabello and Danny Z. Chen, Schloss Dagstuhl - Leibniz-Zentrum für

Informatik, 2020, vol. 164 of *LIPIcs*, pp. 22:1–22:16, URL https://doi.org/10.4230/LIPIcs.SoCG.2020.22.

[58] Michael Lesnick, "The Theory of the Interleaving Distance on Multidimensional Persistence Modules", *Foundations of Computational Mathematics*, 15(3), 2015, pp. 613–650, URL http://link.springer.com/10.1007/s10208-015-9255-y.

[59] Claudia Landi, *Research in computational topology*, Springer Science+Business Media, New York, NY, 2018.

[60] Martina Scolamiero, et al., "Multidimensional Persistence and Noise", *Foundations of Computational Mathematics*, 17(6), 2017, pp. 1367–1406, URL https://link.springer.com/article/10.1007/s10208-016-9323-y.

[61] Heather A. Harrington, et al., "Stratifying Multiparameter Persistent Homology", *SIAM Journal on Applied Algebra and Geometry*, 3(3), 2019, pp. 439–471, URL https://doi.org/10.1137/18M1224350, _eprint: https://doi.org/10.1137/18M1224350.

[62] The RIVET Developers, "RIVET", , 2018, URL http://rivet.online.

[63] Mathieu Carrière and Steve Oudot, "Local Equivalence and Intrinsic Metrics between Reeb Graphs", in *Symposium on Computational Geometry*, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017, vol. 77 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 25:1–25:15, URL http://drops.dagstuhl.de/opus/volltexte/2017/7179.

[64] Peter Bubenik, et al., "Wasserstein distance for generalized persistence modules and abelian categories", *arXiv:180909654 [math]*, 2018, URL http://arxiv.org/abs/1809.09654, arXiv: 1809.09654.

[65] Bryn Keller, et al., "PHoS: Persistent Homology for Virtual Screening", *figshare*, 2018, URL https://chemrxiv.org/articles/PHoS_Persistent_Homology_for_Virtual_Screening/6969260.

[66] Michael Kerber and Arnur Nigmetov, "Efficient Approximation of the Matching Distance for 2-Parameter Persistence", in *36th International Symposium on Computational Geometry (SoCG 2020)*, edited by Sergio Cabello and Danny Z. Chen, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2020, vol. 164 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 53:1–53:16, URL

https://drops.dagstuhl.de/opus/volltexte/2020/12211, iSSN: 1868-8969.

[67] Irena Peeva, *Graded syzygies*, Springer, London New York, 2011.

[68] Mathieu Carrière and Steve Oudot, "Structure and Stability of the One-Dimensional Mapper", *Foundations of Computational Mathematics*, 18(6), 2018, pp. 1333–1396, URL https://doi.org/10.1007/s10208-017-9370-z.

[69] Nicolas Berkouk, "Stable resolutions of multi-parameter persistence modules", *arXiv:190109824 [math]*, 2019, URL http://arxiv.org/abs/1901.09824, arXiv: 1901.09824.

[70] Frédéric Chazal, et al., *The Structure and Stability of Persistence Modules*, SpringerBriefs in Mathematics, Springer Verlag, 2016, URL https://hal.inria.fr/hal-01330678.

[71] Primoz Skraba and Mikael Vejdemo-Johansson, "Persistence modules: Algebra and algorithms", *arXiv:13022015 [cs, math]*, 2013, URL http://arxiv.org/abs/1302.2015, arXiv: 1302.2015 version: 2.

[72] Mikhael Gromov, et al., *Metric structures for Riemannian and non-Riemannian spaces : with Appendices by M. Katz, P. Pansu and S. Semmes*, Progress in mathematics ; v. 152, Birkhauser, Boston, 1998.

[73] Gunnar Carlsson, et al., "Zigzag Persistent Homology and Real-Valued Functions", in *Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry*, Association for Computing Machinery, New York, NY, USA, 2009, SCG '09, pp. 247–256, URL https://doi.org/10.1145/1542362.1542408, event-place: Aarhus, Denmark.

[74] Paul Bendich, et al., "The Robustness of Level Sets", in *Algorithms – ESA 2010*, edited by David Hutchison, et al., Springer Berlin Heidelberg, Berlin, Heidelberg, vol. 6346, 2010, pp. 1–10, URL http://link.springer.com/10.1007/978-3-642-15775-2_1.

[75] Vin de Silva, et al., "Categorified Reeb Graphs", *Discrete & Computational Geometry*, 55(4), 2016, pp. 854–906, URL http://link.springer.com/10.1007/s00454-016-9763-9.

[76] Ulrich Bauer, et al., "Strong Equivalence of the Interleaving and Functional Distortion Metrics for Reeb Graphs", in

Symposium on Computational Geometry, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015, vol. 34 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 461–475, URL http://drops.dagstuhl.de/opus/volltexte/2015/5146.

[77] Ulrich Bauer, et al., "Measuring Distance between Reeb Graphs", in *Proceedings of the thirtieth annual symposium on Computational geometry*, Association for Computing Machinery, Kyoto, Japan, 2014, SOCG'14, pp. 464–473, URL https://doi.org/10.1145/2582112.2582169.

[78] Omer Bobrowski and Shmuel Weinberger, "On the vanishing of homology in random Čech complexes", *Random Structures & Algorithms*, 51(1), 2017, pp. 14–51, URL https://ezproxy-prd.bodleian.ox.ac.uk:4563/10.1002/rsa.20697.

[79] Omer Bobrowski and Goncalo Oliveira, "Random Čech complexes on Riemannian manifolds", *Random Structures & Algorithms*, 54(3), 2019, pp. 373–412, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20800.

[80] Omer Bobrowski and Matthew Kahle, "Topology of random geometric complexes: a survey", *Journal of Applied and Computational Topology*, 1(3), 2018, pp. 331–364, URL https://doi.org/10.1007/s41468-017-0010-0.

[81] Vanessa Robins, "Betti number signatures of homogeneous Poisson point processes", *Phys Rev E*, 74, 2006, p. 061107, URL https://link.aps.org/doi/10.1103/PhysRevE.74.061107.

[82] Nathan Linial and Roy Meshulam, "Homological Connectivity Of Random 2-Complexes", *Combinatorica*, 26(4), 2006, pp. 475–487, URL https://doi.org/10.1007/s00493-006-0027-9.

[83] Rahul Paul and Stephan K. Chalup, "A study on validating non-linear dimensionality reduction using persistent homology", *Pattern Recognition Letters*, 100, 2017, pp. 160–166.

[84] Joshua B. Tenenbaum, et al., "A Global Geometric Framework for Nonlinear Dimensionality Reduction", *Science*, 290(5500), 2000, pp. 2319–2323, URL https://science.sciencemag.org/content/290/5500/2319.

[85] Afra Zomorodian and Gunnar Carlsson, "Computing Persistent Homology", in *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, ACM, New York, NY, USA, 2004, SCG '04, pp. 347–356, URL http://doi.acm.org/10.1145/997817.997870.

[86] Herbert Edelsbrunner and John L. Harer, *Computational topology - an introduction*, American Mathematical Society, Providence, RI, 2010.

[87] Nina Otter, et al., "A roadmap for the computation of persistent homology", *EPJ Data Science*, 6(1), 2017, p. 17, URL https://doi.org/10.1140/epjds/s13688-017-0109-5.

[88] Omer Bobrowski, et al., "Maximally persistent cycles in random geometric complexes", *Ann Appl Probab*, 27(4), 2017, pp. 2032–2060, URL https://ezproxy-prd.bodleian.ox.ac.uk:4563/10.1214/16-AAP1232.

[89] Matthew Kahle, "Sharp vanishing thresholds for cohomology of random flag complexes", *Annals of Mathematics*, 179(3), 2014, pp. 1085–1107, URL http://www.jstor.org/stable/24522785.

[90] Partha Niyogi, et al., "Finding the Homology of Submanifolds with High Confidence from Random Samples", *Discrete & Computational Geometry*, 39(1-3), 2008, pp. 419–441, URL https://link.springer.com/article/10.1007/s00454-008-9053-2.

[91] Frédéric Chazal, et al., "A Sampling Theory for Compact Sets in Euclidean Space", *Discrete & Computational Geometry*, 41(3), 2009, pp. 461–479, URL https://doi.org/10.1007/s00454-009-9144-8.

[92] Yuan Wang and Bei Wang, "Topological inference of manifolds with boundary", *Computational Geometry*, 88, 2020, p. 101606, URL https://doi.org/10.1016/j.comgeo.2019.101606.

[93] V. Gershkovich and H. Rubinstein, "Morse theory for Min-type functions", *Asian J Math*, 1(4), 1997, pp. 696–715, URL https://ezproxy-prd.bodleian.ox.ac.uk:4563/10.4310/AJM.1997.v1.n4.a3.

[94] Allen Hatcher, *Algebraic Topology*, Cambridge University Press, Cambridge, 2002.

[95] Adrian Baddeley, "Spatial point processes and their applications", in *Stochastic geometry*, Springer, Berlin, vol. 1892 of *Lecture Notes in Math.*, 2007, pp. 1–75, URL https://ezproxy-prd.bodleian.ox.ac.uk:4563/10.1007/978-3-540-38175-4_1.

[96] Mathew Penrose, *Random Geometric Graphs (Oxford Studies in Probability)*, Oxford University Press, 2003, URL https://www.amazon.com/Random-Geometric-Graphs-Studies-Probability/dp/0198506260?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=

chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0198506260.

[97] Günter Last, *Lectures on the Poisson process*, Cambridge University Press, Cambridge, 2018.

[98] John M. Lee, *Introduction to smooth manifolds*, vol. 218 of *Graduate Texts in Mathematics*, Springer, New York, 2nd edn., 2013.

[99] John Nash, "The imbedding problem for Riemannian Manifolds", *Annals of Mathematics*, 63, 1956, pp. 20 – 63.

[100] Herbert Federer, "Curvature measures", *Transactions of the American Mathematical Society*, 93(3), 1959, pp. 418–491, URL http://www.ams.org/home/page/.

[101] John Milnor, "Lectures on the H-Cobordism Theorem", *Princeton University Press*, 1965.

[102] M. F. Smiley and John Milnor, "Morse Theory.", *The American Mathematical Monthly*, 71(8), 1964, p. 936, URL http://www.jstor.org/stable/2312441?origin=crossref.

[103] Leopold Flatto and Donald J. Newman, "Random coverings", *Acta Math*, 138(3–4), 1977, pp. 241–264, URL https://ezproxy-prd.bodleian.ox.ac.uk:4563/10.1007/BF02392317.

[104] Wei Chai, *Random Topological Structures*, Ph.D. thesis, University of Chicago, 2018, URL http://knowledge.uchicago.edu/record/425.

[105] R. E. Miles, "Isotropic random simplices", *Advances in Appl Probability*, 3, 1971, pp. 353–382, URL https://ezproxy-prd.bodleian.ox.ac.uk:4563/10.2307/1426176.

[106] Rolf Schneider and Wolfgang Weil, *Stochastic and Integral Geometry*, Probability and Its Applications, Springer-Verlag, Berlin Heidelberg, 2008, URL //www.springer.com/gb/book/9783540788584.

[107] A. Hurwitz, "über die Erzeugung der Invarianten durch Integration", *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1897, 1897, pp. 71–2, URL https://eudml.org/doc/58378.

[108] Persi Diaconis and Peter J. Forrester, "Hurwitz and the origins of random matrix theory in mathematics", *Random Matrices: Theory and Applications*, 06(01), 2017, p. 1730001, URL https://doi.org/10.1142/S2010326317300017, _eprint: https://doi.org/10.1142/S2010326317300017.