Analysing retinal images using (extended) persistent homology

Yuchen Jiang

November 2022

A thesis submitted for the degree of Master of Mathematical Sciences
(Advanced)
of the Australian National University

Australian National University

For family

Declaration

The work in this thesis is my own except where otherwise stated.

Yuchen Jiang

Acknowledgements

First and foremost, I would like to thank my supervisors, Associate Professor Vanessa Robins and Dr. Katharine Turner, for their ongoing support during this project. As someone who started this partially applied project with limited experience in image analysis, I struggled with processing the data and learning new programming languages. Thank you for brainstorming for solutions and checking the codes with me when I got stuck. I am also grateful that you put in the time and effort to create a lovely applied topology community within ANU. I have learned a lot from presentations in our weekly group meetings, and the thought of being part of a circle of people working in the same field is extremely comforting.

I would not have left my beloved home in Nanjing and study in a foreign country without the support from my parents. They give me an opportunity to follow my dream and see the world that they themselves had not gotten a chance to discover. While it has been over 3 years now since I last set foot back on my long-missed hometown, they always make me feel loved and cared for through each video or voice call, or even a simple message. Thank you for supporting every decision I make and being the most understanding parents I could ever hope for. I cannot wait to finally meet you again in person at the end of this year.

Thank you also to my grandparents, who were my first teachers growing up. You taught me to be curious about the world, to be kind to people around me and to work hard but smart to get what I want. It is going to be your 50th anniversary at the beginning of next year, and I look forward to celebrating with you.

To all my friends, I am grateful to have you in my life, and to know that I can count on you whenever I need a quick break from the grind. Even though a catchup in person is not possible sometimes, a short check-in message is enough to make my day.

Last but not least, to my partner, Fabian Circelli, you have made life in Canberra so much brighter and you feel like home to me. Thank you for encouraging me and cheering me up when I feel down, for calming me down when I am stressed or anxious, for your amazing homely food that always comforts me, and for those fun maths debates that keeps me on my toes. The list goes on, and I am grateful to have you in my life.

Abstract

Biometrics are data used for automatic verification and identification of individuals, two important routines commonly performed to enhance the level of security within a system. Therefore, improvements to the analysis of biometrics are crucial. Common examples of biometrics include fingerprints and facial features. In this thesis, we consider retinal fundus images, which are scans of a person's retina blood vessels at the back of the eyeballs. They have become a popular choice for these tasks due to their uniqueness and stability over time.

Traditional methods mainly utilise specific biological features observed in the scans. These methods generally rely on highly accurate automated extractions of these traits, which are challenging to produce especially when abnormalities appear in diseased individuals. In this paper, we instead propose a novel approach, which is more tolerant of the errors from the feature extraction process, to analyse retina biometrics. In particular, we compute the *(extended) persistent homology* of the blood vessel structure (viewed as a manifold with boundary embedded in $\mathbb{R}^2$) in a retinal image with respect to some filtration and produce a summary statistic called a *persistence diagram*. This then allows us to perform further statistical tests.

We test our method on a publicly available database using different choices of filtrations to capture the shapes of the vessels. Some of these choices achieve a high level of accuracy compared with tests done on the same database. Our method also take significantly less time compared to other proposed methods. In the future, we can explore more filtrations and/or use combinations of results obtained from different filtrations to see if we can further increase the level of accuracy.

Keywords: persistent homology, extended persistent homology, personal identification and verification, retinal fundus images.

Contents

Notation and terminology

Notation

$\mathbb{N}$	The set of non-negative integers.
$\mathbb{R}^*$	The set of extended real numbers, i.e. $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, +\infty\}$.
$A \sqcup B$	Disjoint union of A and B. In this thesis, the notation will imply that $A \cap B = \emptyset$.
X/Y	The quotient space of X by Y if X and Y are spaces. That is, collapsing the space Y to a point.
$\beta_k(A)$	The k-th Betti number of a space A for $k \in \mathbb{N}$. This is equal to $\text{rank}(H_k(A))$.
$\beta_k(X, A)$	The k-th Betti number of a pair of spaces (X, A) for $k \in \mathbb{N}$. This is equal to $\text{rank}(H_k(X, A))$.

Terminology

TDA	Topological data analysis
PHT	Persistent homology transform
XPHT	Extended persistent homology transform
XRPHT	Radial extended persistent homology transform
SEDT	Signed Euclidean distance transform
DMT	Distance to measure transform
LES	Long exact sequences

Introduction

Biometrics, i.e. biological measurements that can be used to identify individuals, are used frequently in security. Perhaps the most familiar example is that of smart phone users verifying their identities with their fingerprints or facial features on a daily basis. Improvements to analysis of biometrics are therefore crucial to enhancing the level of security within a system. Biometrics come in many forms. On top of the aforementioned image-based ones, there are also time-dependent ones such as voice or gait. However, when it comes to identity management, we favour biometrics that are stable with respect to multiple recordings from the same individual and have strong discriminatory power that can be used to distinguish between different individuals. Retinal fundus images are such a biometric. Since they are absolutely unique to each individual and are stable over time, they quickly become a popular choice of biometric in high-level security systems.

A retinal image is a 2D-mapping of the unique patterns of a person's retina using a low-intensity light source. Figure 1 is an example of such a scan. It reveals retina blood vessel structures as well as some non-vascular features.

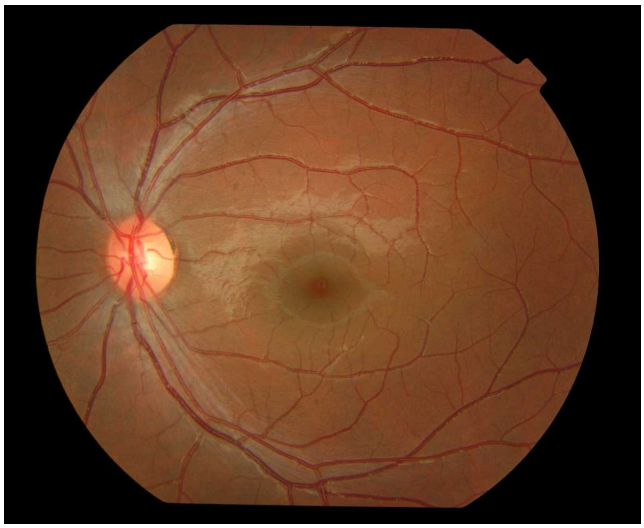


Figure 1: A retinal fundus image. Source: [1]

Current methods used to analyse retina can be divided into two main categories, depending on whether they are vascular based or non-vascular based. In either category, a combination of biological features that can be extracted from retinal scans are studied to generate a version of a similarity score between two samples. Such a score is then used to perform identification and verification tasks. Some examples of such biological traits include minutiae points (vascular based) such as bifurcation points where vessels split into two branches and ending points where vessels terminate, and luminance and structure (non-vascular based) of the scan. The main issues with these methods is that they generally rely on highly accurate segmentation, a process to classify and label each pixel of the image such that pixels with the same label share common characteristics. Automated segmentation can be extremely challenging to produce, especially when abnormalities appear in diseased individuals.

In this paper, we propose a novel approach to retina analysis using (extended) persistent homology, which adopts tools from algebraic topology to study shapes of proposed objects. We used a publicly available database called Retina Identification Database (RIDB). We start by translating given retinal images into a (piecewise-linear) manifold with boundary. While this requires segmentation that identifies the blood vessels in the samples, we find that no strict requirement on the level of accuracy of the segmentation is needed. By computing (extended) persistent homology of the vessel structure with respect to a chosen filtration, we obtain a summary statistic for each sample called the (extended) persistence diagram. A convenient metric defined on pairs of persistence diagrams, the Wasserstein distance, is then used to compare two samples. We summarise the procedure below:

(a) Preprocess image including performing image segmentation.

(b) Construct a filtration from an image, a delicate job that requires an expert understanding of how to capture the geometric and topological information of interest.

(c) Compute (extended) persistent homology with respect to the chosen filtration and obtain (extended) persistence diagrams.

(d) Compute pairwise 1-Wasserstein distance between (extended) persistence diagrams for different samples.

(e) Design a classifier to perform identification and verification tasks.

(f) Test the accuracy of the classifier.

The paper is divided into three chapters. Chapter 1 contains background material for the methods used in the thesis. In Section 1.1, we recall definitions and theorems in homology theory while we introduce the persistent homology and extended persistent homology in Sections 1.2 and 1.3. We finish off Chapter 1 with a close-up examination of (extended) persistent homology computed with respect to radial filtrations. We also discuss practical considerations required when analysing digital images in Section 1.4.

In Chapter 2, we perform the analysis of RIDB database following the steps mentioned above. In particular, in Section 2.3, we compare our methods to the ones in [30] since they also tested on the same RIDB database. We find that our methods achieve a comparable level of accuracy and take less time to complete the tasks.

Finally in Chapter 3, we prove a theorem that justifies an algorithm we used to compute extended persistent homology of the vessel structures in retinal scans with respect to radial filtrations in Section 2.2.3. The theorem allows us to recover the extended persistence pairings, with respect to a radial filtration, of a manifold with boundary embedded in $\mathbb{R}^n$ in dimensions 0 and $n-1$ from those of its boundary under mild assumptions. This is an extension of the result in [29], which proves an equivalent statement for height filtrations.

Chapter 1

A brief guide to (extended) persistent homology

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair
And having perhaps the better claim
Because it was grassy and wanted wear,
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black
Oh, I marked the first for another day!
Yet knowing how way leads on to way
I doubted if I should ever come back

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I,
I took the one less traveled by,
And that has made all the difference.

The road not taken, Robert Frost

It is both a blessing and a curse to have data of various forms bursting into our

lives everyday. We have more resources to gain a better understanding of ourselves and our surroundings, but the sheer volume of data can also be overwhelming – we need smarter and more efficient methods for analysis. It has also become apparent that the nature of data we collect nowadays is far more complicated than it used to be. Data points may appear in higher dimensions with stronger noise, and chances of missing data grow [4]. One important example of such data is biometrics , which are biological measurements that can be used for automatic verification and identification of individuals. Traditional data analysis techniques, e.g. regression models and clustering, might not work well with these geometric data that has much noise in the background. This is partially because the coordinate vectors to

Figure 1.1: An example of biometrics [7]: two retinal fundus images, the left one is from a healthy individual while the right one is from someone with eye diseases. Note that the data is given as a digital image, where we can extract information such as coordinate vectors and distance between vessels.

compare relevant information from two biometric samples do not arise naturally, and we do not care much about local parameters – what is of interest here is a global summary. In the example of retinal fundus images above (Figure 1.1), we may have a database where some images are from individuals with diseases and some without. What we are interested in is whether we can decide *automatically* from the discolouration in the non-vascular region or other features, as a whole, from a new individual the likelihood of them having certain diseases; we *do not* want parameters, for instance, associated with each pixel of the image. Thus, we need a way to study the data in a qualitative way while keeping its geometric nature. Topology becomes a natural choice. It captures the general geometry of the shape of the data points – for example, the connected components in a space,

and a notion of closeness between objects that does not rely on metrics.

A growing interest can be found across various disciplines in adapting the idea from algebraic topology to study the shape of a given set of data points in space. This relatively recent approach is formally known as *topological data analysis* (or TDA for short). So while topology can be sensitive to small pieces of extra noise, TDA allows us to quantify the topological structure with respect to a geometric parameter and provides a measure of significance for topological features. A 2017 paper by Kanari, et al. [16], for instance, invented a topological morphology descriptor based on TDA to encode and study the branching patterns of neuronal trees. Other applications of TDA can be found in the field of anthropology to analyse the shape of fossils or biology to quantify the behavior of biological aggregation and many more.

In this thesis, we propose a novel approach to personal identification and verification from retinal images using tools from TDA, more specifically, by computing (extended) persistent homology ((X)PH) on given input images. Waheed, et al. [30] performed the same task using biological features observed in retinal fundus images on a retinal identification database they created and made available online. This allows us to compare our method to theirs.

In fact, the methods we proposed here can be applied to a general class of data, namely (smooth or piecewise-linear) manifolds with boundary. Here, we revise some basic definitions about manifolds so that we can include useful results about (extended) persistence of some real-valued function over a manifold (with boundary) later in the chapter.

Definition 1.1 (Topological manifolds)**.** A *chart* (U, ϕ) for a topological space (which is a geometric space where we define what it means for a set to be open) M is a bijective and continuous function (a *homeomorphism*) $\phi : U \to \phi(U)$, where $U \subset M$ is an open set of M and $\phi(U)$ is an open subset of some Euclidean space.

A *topological n-manifold* is a Hausdorff, second countable topological space M such that for every point $x \in M$, there exists a chart $\phi_x : U_x \to \phi(U_x)$ where $U_x \subset M$ is an open neighbourhood of x and $\phi(U_x) \subset \mathbb{R}^n$ is an open subset.

A *topological n-manifold with boundary* is defined similarly to topological manifold of dimension n except we have $\phi(U_x) \subset \{(x_1, \cdots, x_n) \in \mathbb{R}^n | x_1 \geq 0\}$ is an open subset of the closed half-space of $\mathbb{R}^n$.

To make sense of smoothness and piecewise linearity on a manifold, we require some notion of compatibility between charts. To that end, consider two charts

(U_α, ϕ_α) and (U_β, ϕ_β) with $U_\alpha \cap U_\beta \neq \emptyset$. Then restricting ϕ_α and ϕ_β to the domain $U_\alpha \cap U_\beta$, we have two homeomorphisms. Combining them gives two new homeomorphisms: $\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \to \phi_\alpha(U_\alpha \cap U_\beta)$ and $\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$. We call them the *transition maps*. An *atlas* $\mathcal{A} = \{(U_\alpha, \phi_\alpha)\}$ for M is a collection of charts such that $\cup_\alpha U_\alpha = M$.

Definition 1.2 (Smooth/piecewise linear manifolds)**.** A topological n-manifold, with or without boundary, is *smooth* (or *piecewise linear* respectively) if its transition maps are smooth (or piecewise linear respectively). The corresponding charts are called *compatible*.

We can also define what it means for a function on a manifold to be smooth or piecewise linear.

Definition 1.3. Let M be a topological n-manifold, with or without boundary, with smooth (or piecewise linear) structure $\{(U_\alpha, \phi_\alpha)\}$. Then a function $f : M \to \mathbb{R}$ is *smooth* (or *piecewise linear*) if for all charts (U_α, ϕ_α), the map $f \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha) \to \mathbb{R}$ is a smooth (or piecewise linear) function.

In Chapter 3, we will consider a smaller subset of such functions called Morse functions, while in Chapter 2, we will dive into a more hands-on examination of the retinal images, which can be translated into manifolds with boundary embedded in $\mathbb{R}^2$. But before all of these, we start by introducing concepts relevant to (X)PH assuming only basic definitions from algebra and topology.

1.1 Simplicial complexes and homology

We are hoping to capture the topology of the shapes of given data. The data we are interested in are generally subsets of $\mathbb{R}^n$, and hence, we consider the subspace topology, where a subset is *open* if and only if it is an intersection of an open set in $\mathbb{R}^n$ with the given subset. The strongest notion of equivalence of topological spaces is called *homeomorphism*. Two topological spaces are *homeomorphic* if one can be continuously deformed into the other and back. However, deciding whether two manifolds are homeomorphic proves to be algorithmically unsolvable in general [20]. Thus, we resort to an easily computed, yet weaker notion, to capture the topology of a given shape, *homology*. It connects topological spaces with algebraic objects and offers us tools from both fields to study the subject more closely. To find the homology, and later on, the (extended) persistent homology, of a given space, we need to represent the space in terms of simplicial complexes.

We review the definitions of simplicial complexes and homology groups below. In particular, we distinguish between the notion of abstract simplicial complexes and geometric simplicial complexes, and connect the two notions via a construction called geometric realisation.

We mainly follow the definitions and constructions as in [13] and [18]. We refer the reader to [13] for a more detailed study of homology theories.

Definition 1.4. Given a finite number of points $x_0, x_1, \cdots, x_m$ in $\mathbb{R}^n$, an *affine combination* of these points is a point in $\mathbb{R}^n$ of the form

$$\sum_{i=0}^{m} \alpha_i x_i = \alpha_0 x_0 + \alpha_1 x_1 + \cdots \alpha_m x_m$$

such that $\sum_{i=0}^{m} \alpha_i = 1$. A *convex combination* of these points is an affine combination $\sum_{i=0}^{m} \alpha_i x_i$ such that $\alpha_i \geq 0$ for all i.

The *affine hull* of these points is the set of all affine combinations of these points. Similarly, the *convex hull* of these points is the set of all convex combinations of these points. Moreover, we call these $m+1$ points *affinely independent* if their affine hull is m-dimensional. Equivalently, the points $\{x_0, x_1, \cdots, x_m\}$ are affinely independent if the set of edges $\{x_1 - x_0, x_2 - x_0, \cdots, x_m - x_0\}$ is a linearly independent set.

Definition 1.5 (Geometric simplicial complexes)**.** A *geometric n-simplex* σ is the convex hull of $n+1$ affinely independent points (which are called *vertices*) in $\mathbb{R}^m$ with an ordering of its vertices. Illustrated in Figure 1.2 are examples of standard n-simplices defined by

$$\Delta^n = \left\{ (t_0, t_1, \cdots, t_n) \in \mathbb{R}^{n+1} \middle| \sum_{i=0}^{n} t_i = 1, t_i \geq 0 \text{ for all } i = 0, \cdots, n \right\}$$

for $n = 0, 1, 2$ with the natural ordering of vertices based on the subscripts. If $m < n$, the convex hull of $m+1$ of the $n+1$ points is a *m-face* of the n-simplex. The union of all the faces of Δ^n is the *boundary* of Δ^n, denoted by $\partial\Delta^n$. The open simplex $\Delta^n \backslash \partial\Delta^n$ is the *interior* of Δ^n.

A *Δ-complex* is a space $X \subset \mathbb{R}^d$ built iteratively as follows:

(0) X^0 is a discrete set of 0-simplices (vertices);

(n) Given a collection of maps $\sigma_\alpha^n : \partial\Delta_\alpha^n \to X^{n-1}$ such that for each $(n-1)$-face A of $\partial\Delta_\alpha^n$, the restriction $\sigma_\alpha^n|_A$ is the inclusion map into one of the $(n-1)$-simplices in X^{n-1} preserving the ordering of vertices in X^{n-1}. We then

glue Δ^n_α onto X^{n-1} using $\{\sigma^n_\alpha\}$to build X^n, i.e. $X^n = X^{n-1} \sqcup (\bigsqcup_\alpha \Delta^n_\alpha) / \sim$, where $\sim$ identifies $x \in \partial\Delta^n_\alpha$ with $\sigma^n_\alpha(x)$ for all α;

(∞) $X = \bigcup_n X^n$.

Finally, a *geometric simplicial complex* is a Δ-complex such that for each $n \in \mathbb{N}$, the map $\sigma^n_\alpha : \partial\Delta^n_\alpha \to X^{n-1}$ maps different $(n-1)$-faces of $\partial\Delta^n_\alpha$ to different $(n-1)$-simplices in X^{n-1} for all α.

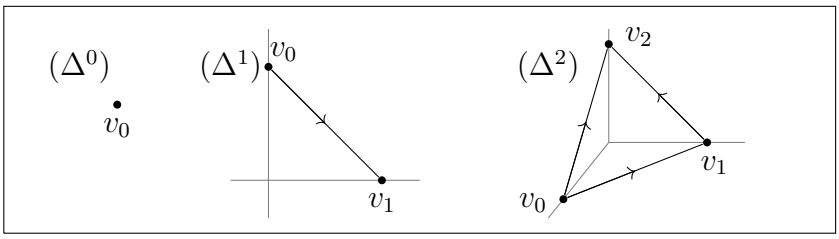


Figure 1.2: Standard n-simplices for $n = 0, 1, 2$ with orientation of edges given by ordering of the vertices. We can see that Δ^0 is just one vertex and Δ^1 is the edge connecting two 0-simplices. Finally, Δ^2 is a regular triangle in $\mathbb{R}^3$ with vertices at $(1, 0, 0), (0, 1, 0)$ and $(0, 0, 1)$ respectively. Denote by $[v_0, v_1]$ the edge facing the vertex v_2 at the bottom. Similarly, we use $[v_1, v_2], [v_0, v_2]$ to denote the edge facing vertices v_0, v_1 respectively. Note that the boundaries of an n-simplex looks like an $(n-1)$-simplex. This is true for higher n's as well.

As a comparison, we now give the definition of abstract simplicial complexes.

Definition 1.6 (Abstract simplicial complexes)**.** Given a collection of sets, any $n+1$ of them is called an *abstract n-simplex*. Any subset of a n-simplex is a *face* of the n-simplex.

A collection K of finite non-empty subsets of a set S is called an *abstract simplicial complex* if, for every set X in K, and every non-empty subset $Y \subseteq X$, the set Y also belongs to K. Any finite set that belongs to K is a *face* of K. Elements in S form the *vertex set* of K.

In fact, we can make one definition that describes both of them.

Definition 1.7 (Simplicial complexes)**.** A *simplicial complex* K is a set of simplices satisfying two conditions:

(i) Every face of a simplex σ from K also belongs to K;

(ii) Any two simplices σ_1, σ_2 from K are either disjoint, i.e. $\sigma_1 \cap \sigma_2 = \emptyset$ or intersect in a common face, i.e. $\sigma_1 \cap \sigma_2$ is a face of both σ_1 and σ_2.

If K is a simplicial complex in $\mathbb{R}^n$, then its *underlying space* is the union of its simplices equipped with the subspace topology inherited from $\mathbb{R}^n$.

In this thesis, we are only interested in *finite* simplicial complexes, i.e. simplicial complexes with a finite number of simplices.

From this point on, we assume all simplicial complexes mentioned are finite.

We will see that an abstract simplicial complex with n simplices can be made concrete into some geometric simplicial complex in $\mathbb{R}^n$. With this in mind, it makes sense to talk about an underlying space of an abstract simplicial complex since we can consider the underlying space of its associated geometric simplicial complex. We formalise this by introducing the definition of a geometric realisation for finite cases.

Definition 1.8 (Geometric realisations in finite cases)**.** Let K be a finite abstract simplicial complex. Then denote by $|K|$ its geometric realisation constructed as follows: consider the vertex set S of K and map vertex $v_k \in S$ to the unit coordinate of the k-th axis in $\mathbb{R}^n$, where n the number of vertices in K. Then every face σ of K can be naturally identified with a standard simplex. Finally, we take $|K|$ to be the union of these geometric simplices.

From this definition, it is easier to see that abstract simplicial complexes are a combinatorial generalisation of geometric simplicial complexes.

The primary form of data we study here are geometric objects in Euclidean spaces. Consider a topological space S. Denote by X its triangulation, which is a simplicial complex together with a homeomorphism $X \to S$. In other words, X is a simplicial complex whose underlying space is homeomorphic to S. Such a simplicial complex exists for all differentiable manifolds [15][31].

Define $C_n(X) \coloneqq G\{n\text{-simplices in } X\}$, where G is an abelian group. Elements of $C_n(X)$ are formal sums $\sum_{i \in I} a_i \sigma_i$, where I is a finite index set, $a_i \in G$ and σ_i are n-simplices in X. We call these elements *n-chains*. That is, $C_n(X)$ is a free abelian group generated by the n-simplices of X.

For computational simplicity, from this point onward, we will take $G = \mathbb{Z}_2$.

Remark 1.9. The reason we choose to work with $G = \mathbb{Z}_2$ is that an n-chain is then precisely the sum of n-simplices considered (since $\mathbb{Z}_2 = \{0, 1\}$). We can then ignore orientations (since $-1 = 1 \in \mathbb{Z}_2$), and adding two chains containing a common simplex means we are able to ignore the simplex (since $1 + 1 = 0$ in $\mathbb{Z}_2$). Furthermore, since $\mathbb{Z}_2$ is a field, $C_n(X)$ (and homology groups) becomes a vector space over $\mathbb{Z}_2$ instead of just an abelian group. This allows us to consider (extended) persistence modules later.

Define the *boundary homomorphism* $\partial_n : C_n(X) \to C_{n-1}(X)$ by

$$\sigma \mapsto \sum_{i=0}^{n} \sigma_{[v_0, v_1, \cdots, \hat{v}_i, \cdots, v_n]},$$

where $\hat{v}_i$ indicates omitting the i-th vertex and $\sigma_{[v_0, v_1, \cdots, \hat{v}_i, \cdots, v_n]}$ is the $(n-1)$-simplex that is opposite of the i-th vertex, or alternatively, the simplex containing all vertices but v_i. That is, an n-simplex σ is mapped to the sum of its boundary, a set of $(n-1)$-simplices.

Definition 1.10 (Chain complex and homology groups)**.** A *chain complex* $C_*(X)$ is a sequence of maps of abelian groups

$$\cdots \xrightarrow{\partial_{n+2}} C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} C_0 \longrightarrow 0$$

with $\partial_n \circ \partial_{n-1} = 0$, for all n.

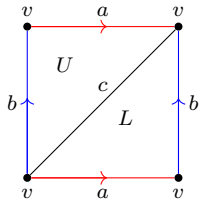
The *n-th homology group* is $H_n(X) = \frac{\ker(\partial_n)}{\mathrm{im}(\partial_{n+1})}$.

With some tedious computation, we can show that $\partial_n \circ \partial_{n-1} = 0$, which means the sequence of $C_n(X)$ for $n \in \mathbb{N}$ does form a chain complex. This allows us to compute the homology groups of a simplicial complex. Moreover, as noted above, $\mathrm{im}(\partial_{n+1})$ represents the boundary of an $(n+1)$-simplex, which is a set of its n-dimensional faces since we are working with $\mathbb{Z}_2$-coefficients. We call it the *n-boundary*, denoted by $B_n(X) = \mathrm{im}(\partial_{n+1})$. Similarly, we observe that $\ker(\partial_n)$ picks out the n-dimensional 'loops' built from n-simplices. Hence, we call $Z_n(X) = \ker(\partial_n)$ the *n-cycles*. Using this notation, we have $H_n(X) = \frac{Z_n(X)}{B_n(X)}$.

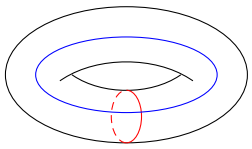
For small n, we can spell out exactly what the homology groups represent. For instance, when $n = 0$, by quotienting out the boundary of 1-simplices from the group generated by the 0-simplices (vertices), vertices that are connected by edges are regarded as one. That is, the zeroth homology group picks out how many connected components there are in the simplicial complex. In the same way, we can see that the first homology group detects how many 'holes' there are

on a 2D plane (think about a disc versus a circle). Moreover, higher dimensional homology detects holes of higher dimensions defined analogously.

Example 1.11 (Homology group of a torus)**.** We can represent a torus T with the following triangulation. Note that the first figure shows a Δ-complex structure on T. We identify vertices with the same label and edges of the same colour together with their orientations.



Note that it has one 0-simplex v, three 1-simplices a, b, c and two 2-simplices U, L.



Note that we can have the above representation because if we fold the rectangle vertically first so that the opposite sides b coincide to get a cylinder and then fold the cylinder horizontally so that the opposite sides a coincide, we get precisely the torus. The color in the representation matches the two cycles we have in the torus. The orientation in the representation indicates that we fold the rectangle without twisting it.

We then have the boundary maps:

$$\partial_1 : C_1(T) \to C_0(T) \text{ by } a \mapsto v + v = 0, b \mapsto v + v = 0, c \mapsto v + v = 0;$$

and

$$\partial_2 : C_2(T) \to C_1(T) \text{ by } U \mapsto a + b + c, V \mapsto a + b + c.$$

Then we can compute the homology groups

$$H_0(T) = \frac{\mathbb{Z}_2\{v\}}{\text{im}(\partial_1)} = \mathbb{Z}_2\{v\}.$$

This is consistent with our observation above: there is only one connected component in T, and thus the zeroth homology group is generated by only one element. Similarly, the torus has two holes, and we should suspect that the first homology group is generated by two elements. And this is precisely the case:

$$H_1(T) = \frac{\ker(\partial_1)}{\text{im}(\partial_2)} = \frac{\mathbb{Z}_2\{a, b, c\}}{\mathbb{Z}_2\{a + b + c\}} = \mathbb{Z}_2\{a, b\}.$$

Note that $\ker(\partial_1)$ is generated by a, b, c, which are precisely the three cycles (c corresponds to the cycle that goes around the torus in a spiral way) of the torus.

But since a, b, c form a boundary of faces of the torus, we would be overcounting the cycles if we include all of them since c can be obtained from a and b. Hence, taking the quotient by the boundary, we avoid overcounting. Finally, we have

$$H_2(T) = \ker(\partial_2) = \mathbb{Z}_2\{U + V\}.$$

A useful variation of the homology groups is called reduced homology groups.

Definition 1.12 (Reduced homology groups)**.** Let X be a simplicial complex. Define a map $\epsilon : C_0(X) \to \mathbb{Z}_2$ by $\sum_i a_i x_i \mapsto \sum_i a_i$. The *reduced homology groups* of X are the homology groups of

$$\cdots \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \xrightarrow{\epsilon} \mathbb{Z}_2 \longrightarrow 0.$$

That is,

$$\widetilde{H}_n(X) \coloneqq \begin{cases} H_n(X), & n > 0; \\ \frac{\ker(\epsilon)}{\mathrm{im}(\partial_1)}, & n = 0 \end{cases}.$$

Note that reduced and unreduced homology groups differ only at dimension 0. In particular, it satisfies the relation $\mathrm{rank}(H_0(X)) = \mathrm{rank}(\widetilde{H}_0(X)) + 1$. In the case of a connected space X, this means we have $\widetilde{H}_0(X) = 0$.

Moreover, we can consider the relative relation between a space and its subspace. This is particularly useful when we introduce extended persistent homology in section 1.3.

Definition 1.13 (Relative homology)**.** Given a space X and a subspace $A \subseteq X$. Let $C_n(X, A) \coloneqq C_n(X)/C_n(A)$. Note that this is doable since $C_n(X)$ is abelian. Since the boundary map $d_n : C_n(X) \to C_{n-1}(X)$ takes $C_n(A)$ to $C_{n-1}(A)$, this induces a boundary map $\partial_n : C_n(X, A) \to C_{n-1}(X, A)$, and $\partial^2 = 0$ holds as it holds for $C_*(X)$. Hence, we obtain a chain complex

$$\cdots \longrightarrow C_n(X, A) \xrightarrow{\partial_n} C_{n-1}(X, A) \longrightarrow \cdots.$$

The *relative homology groups* $H_n(X, A)$ are then defined in the usual way by taking $\ker(\partial_n)/\mathrm{im}(\partial_{n+1})$.

Below we include some useful theorems without proofs relating to relative homology and reduced homology groups. Detailed proof can be found in [13]. We will see them again when we introduce extended persistent homology later on in Section 1.3.

Lemma 1.14. *Let $f : X \to Y$ be a map between spaces. Then f induces a chain map*

$$\begin{aligned} f_{\#} : \ C_*(X) &\to \ C_*(Y) \\ \sigma &\mapsto \ f \circ \sigma. \end{aligned}$$

Moreover, let C_*, D_* be chain complexes and $g : C_* \to D_*$ be a chain map. Then g induces a map

$$\begin{aligned} g_* : \ H_n(C_*) &\to H_n(D_*) \\ [x] &\mapsto \ a\,[g_n(x)]\,. \end{aligned}$$

The proof of Lemma 1.14 boils down to verify well-definedness of the induced map and its requisite properties. Its implication, however, is of much interest. What Lemma 1.14 shows, is that if we are given a map on space level, it can induce a map on homology level. This idea is particularly useful, and Theorem 1.16 below is an occasion where we make use of this lemma. It describes a desirable algebraic relation between homology groups.

Definition 1.15 (Exactness)**.** A sequence of homomorphisms

$$\cdots \longrightarrow A_{n+1} \xrightarrow{\alpha_{n+1}} A_n \xrightarrow{\alpha_n} A_{n-1} \longrightarrow \cdots$$

is *exact* if $\ker(\alpha_n) = \operatorname{im}(\alpha_{n+1})$.

Theorem 1.16. *Let X be a space, and $A \subseteq X$ be a subspace such that (X, A) is a* good pair [†]*. Then there is an exact sequence*

$$\begin{aligned} \cdots \longrightarrow \widetilde{H}_n(A) \xrightarrow{i_*} \widetilde{H}_n(X) \xrightarrow{j_*} H_n(X, A) \xrightarrow{\partial} \widetilde{H}_{n-1}(A) \longrightarrow \cdots \\ \cdots \longrightarrow H_0(X, A) \longrightarrow 0, \end{aligned} \tag{1.1}$$

where i_* is induced by the inclusion map $i : A \to X$ and j_* is induced by the projection map $j : X \to X/A$. The map ∂ is called a connecting homomorphism*.*

The proof for Theorem 1.16 reduced to the observation that

$$0 \to C_*(A) \to C_*(X) \to C_*(X, A) \to 0$$

[†]The terminology of a "good pair" is also defined in [13]. A pair of spaces (X, A) is a *good pair* if A is a nonempty closed subspace that is a deformation retract of some neighborhood in X. The exact definition of a good pair is not of importance in this thesis. However, in this thesis, we will be working with manifolds with boundary embedded in $\mathbb{R}^n$. When A, X are both spaces in this class, the pair (X, A) is a good pair.

forms an exact sequence, which we call a *short exact sequence*. This, in turn, induces the long exact sequence (1.1).

Due to the nice properties of exact sequences, we can read off much useful information about relations of different homology groups. Here, we include another powerful tool that constructs long exact sequences, namely a *Mayer-Vietoris sequence*. We will use this in Chapter 3.

Theorem 1.17 (Mayer-Vietoris sequence)**.** *Suppose $A, B \subseteq X$ is a pair of subspaces such that $X = int(A) \cup int(B)$. Then there is a long exact sequence:*

$$\cdots \longrightarrow H_n(A \cap B) \xrightarrow{\Phi} H_n(A) \oplus H_n(B) \xrightarrow{\Psi} H_n(X) \xrightarrow{\partial} H_{n-1}(A \cap B) \longrightarrow \cdots \\ \cdots \longrightarrow H_0(X) \longrightarrow 0. \tag{1.2}$$

There is also a relative version of Mayer-Vietoris sequence. Suppose we have two pairs of subspaces $(A, C), (B, D) \subseteq (X, Y)$ such that $Y = int(C) \cup int(D)$. Then there is a long exact sequence:

$$\cdots \longrightarrow H_n(A \cap B, C \cap D) \xrightarrow{\Phi} H_n(A, C) \oplus H_n(B, C) \xrightarrow{\Psi} H_n(X, Y) \xrightarrow{\partial} \\ H_{n-1}(A \cap B, C \cap D) \longrightarrow \cdots \longrightarrow H_0(X, Y) \longrightarrow 0. \tag{1.3}$$

Another result in homology theory (that is useful for computing extended persistent homology) is the following:

Theorem 1.18. *Let (X, A) be a good pair. Then we have $H_n(X, A) \cong \widetilde{H}_n(X/A)$.*

1.2 Persistent homology

1.2.1 Persistence modules and interval decomposition

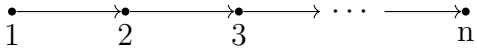
The concept of persistence, or rather persistence modules arises from quiver theory, which is the study of oriented multigraphs commonly used in representation theory [19].

Definition 1.19. A *representation* of a quiver Q is an assignment of

1. a vector space V_i for each vertex i in Q;
2. a linear map $x_{ij} : V_i \to V_j$ when there is an edge from vertex i to vertex j in Q.

The *dimension* of a representation of Q is the sum of the dimensions of the vector spaces it consists of.

Classifications of quiver representations of persistence modules have been studied rigorously. The term "persistence module" was first coined by Zomorodian and Carlsson [32], where they formally studied its connection to quiver representation. But the first account of "persistent homology groups" can be traced back to work of Robins [22] to study the topology of attractors in dynamical systems with experimental data and Edelbrunner, Letscher and Zomorodian [10] for automated topological simplification. Edelsbrunner et al. described a persistence module as a representation of a type L_n quiver (or its generalisation to an infinite diagram and/or diagrams with infinite-dimensional module), where a type L_n quiver is a directed graph of the following form:



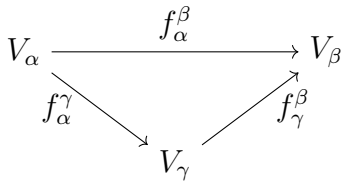
That is, a persistence module is a finite sequence of finite dimensional vector spaces indexed by natural numbers and connected by linear maps. In fact, persistence modules can be defined on a much more general parameter space namely a totally ordered set. That said, the parameter space is generally taken to be a subset of $\mathbb{R}$ in literature [29]. On the other hand, formulations such as persistence barcodes and diagrams, which summarise persistence modules for comparison, and stability results regarding different notions of distance drag persistence modules away from quiver theory. This allows it to become a field of interest in its own right. We will make this more concrete following the construction in [29], which adopts the more general definition for reasons that will become clearer in Section 1.3.

Definition 1.20. A totally ordered set $(\Theta, \leq)$ is a set Θ with a relation $\leq$ such that for all $\alpha, \beta, \gamma \in \Theta$, we have

- (Reflexivity) $\alpha \leq \alpha$;
- (Antisymmetry) if $\alpha \leq \beta$ and $\beta \leq \alpha$, then $\alpha = \beta$;
- (Transitivity) if $\alpha \leq \beta$ and $\beta \leq \gamma$, then $\alpha \leq \gamma$;
- (Comparability) either $\alpha \leq \beta$ or $\beta \leq \alpha$.

After defining the parameter space, we are now ready to define persistence modules.

Definition 1.21 (Persistence modules)**.** Let $(\Theta, \leq)$ be a totally ordered set. A *persistence module* V over Θ consists of a family $\{V_\alpha\}_{\alpha\in\Theta}$ of finite dimensional vector spaces over some field $\mathbb{F}$ and for all $\alpha \leq \beta$, a map $f_\alpha^\beta : V_\alpha \to V_\beta$ such that $f_\alpha^\alpha = id_{V_\alpha}$ and if $\alpha \leq \gamma \leq \beta$ the following diagram commute:



Categorically, that is, a persistence module is a functor $V : (\Theta, \geq) \to \mathrm{Vect}_\mathbb{F}$.

We can, in fact, drop the finite dimensionality condition in Definition 1.21 and achieve full generality. However, in this thesis, we will restrict our attention to the finite dimensional case. In fact, we will only be looking at a finite chain of vector spaces linked by some linear maps. This will become more apparent later on in the thesis.

From now on, a persistence module will be assumed to consist of a finite family of finite dimensional vector spaces over some field $\mathbb{F}$.

Two theorems from representation theories lay a theoretical foundation to what comes next.

Theorem 1.22 (Krull, Remak, Schmidt)**.** *Let Q be a finite quiver, and let $\mathbb{F}$ be a field. Then every finite-dimensional representation V of Q over $\mathbb{F}$ can be decomposed in to a direct sum*

$$V = V_1 \oplus V_2 \oplus \cdots V_r,$$

where V_i is indecomposable *for all $i = 1, \cdots, r$, i.e. if $V_i = U \oplus W$, where U, W are representations of Q, then either $U = 0$ or $W = 0$. Moreover, such a decomposition is unique up to isomorphism and permutation.*

Theorem 1.23 (Gabriel's theorem for L_n-type quivers)**.** *Let Q be a type L_n quiver, and let $\mathbb{F}$ be a field. Then every indecomposable representation of Q over $\mathbb{F}$ is isomorphic to some interval representation $\mathcal{I}_{[b,d]}$:*

$$\underbrace{0 \longrightarrow \cdots \longrightarrow 0}_{[1,b-1]} \longrightarrow \underbrace{\mathbb{F} \xrightarrow{\mathbb{1}} \cdots \xrightarrow{\mathbb{1}} \mathbb{F}}_{[b,d]} \longrightarrow \underbrace{0 \longrightarrow \cdots \longrightarrow 0}_{[d+1,n]}.$$

Remark 1.24. Theorem 1.23 holds for more general types of quivers. For example, it holds for type A_n quivers. These are the quivers with the same underlying graph as L_n-type quivers, except the arrows can now point at either directions. Representations of an A_n-type quiver are generally called *zigzag modules*.

Combining the two theorems, we know that every persistence module V has an *interval decomposition*, unique up to permutation. Recall that we defined persistence modules on a totally ordered set $(\Theta, \leq)$. It might not be immediately clear what we mean by intervals in an arbitrary totally ordered set, but they are analogous to intervals in $\mathbb{R}$. We make it precise in the following definition.

Definition 1.25 (Intervals)**.** An *interval* in a totally ordered space $(\Theta, \leq)$ is a subset $I \subseteq \Theta$ such that for all $\alpha \in \Theta$, either $\alpha \in I$ or $\alpha \leq \theta$ for all $\theta \in I$ or $\theta \leq \alpha$ for all $\theta \in I$. An *interval module* over an interval I is a persistence module $\mathcal{I}_I$ with vector spaces

$$V_\theta = \begin{cases} \mathbb{F}, & \theta \in I \\ 0, & \theta \notin I \end{cases},$$

and maps

$$\pi_\alpha^\beta = \begin{cases} id_\mathbb{F}, & \alpha, \beta \in I \\ 0, & \text{otherwise} \end{cases}.$$

We can switch our attention to elements in vector spaces in a persistence module.

Definition 1.26 (Birth and death)**.** Let V be a persistence module over a totally ordered set $(\Theta, \leq)$. We say a non-zero element $\lambda \in V_\alpha$ is *born* in V_α if for every $\alpha' < \alpha$, $\lambda \notin \text{im}(f_{\alpha'}^\alpha)$. Moreover, we say that λ *dies* entering V_β if for all $\alpha' < \alpha < \beta' < \beta$, $f_\alpha^{\beta'}(\lambda) \notin \text{im}(f_{\alpha'}^\beta)$, but $f_\alpha^\beta(\lambda) \in \text{im}(f_{\alpha''}^\beta)$ for some $\alpha'' < \alpha$. If no such β exists, we say λ *dies* at $V_{+\infty}$.

Figure 1.3 offers a nice illustration of Definition 1.5. We include it below:
What Theorem 1.22 and 1.23 indicate, in terms of Definition 1.5, is that

(a) every element in a given family of vector spaces in the persistence module has a (unique) birth and death time and thus, can be represented by an interval module over some interval I. So to link the two viewpoints, for each interval module $\mathcal{I}_I$, we call $\mathfrak{b}(\mathcal{I}_I) = \inf(I)$ and $\mathfrak{d}(\mathcal{I}_I) = \sup(I)$ the *birth and death parameters* respectively, and

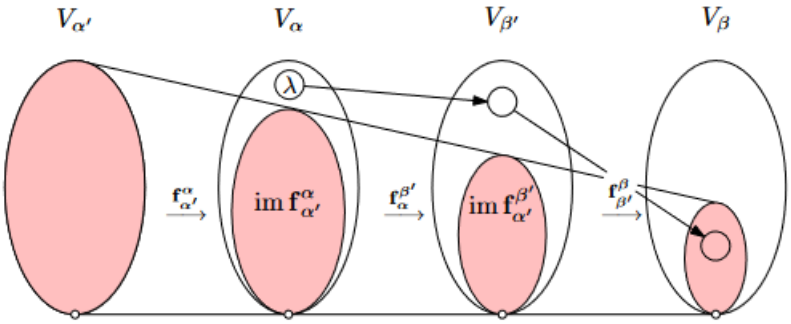


Figure 1.3: Element λ is born in V_α since it is not in the shaded area, which is the image of $V_{\alpha'}$ for all $\alpha' < \alpha$. It dies entering V_β since this is the first time it merges into the image of $V_{\alpha'}$ for some $\alpha' < \alpha$. [Source: [18]]

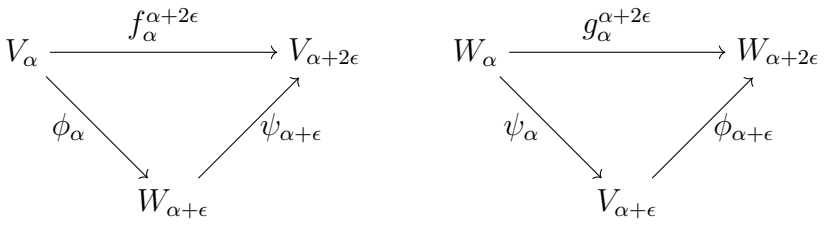
(b) if we know about the birth and death of the generators in the vector spaces in a persistence module, we know everything about the persistence module.

In light of this, we can summarise this information using diagrams, either persistence barcodes or persistence diagrams. *Persistence barcodes* are 2D diagrams whose horizontal axes represent time (e.g. a subset of $\mathbb{R}^*$) and the vertical axes denote the elements. For each generator in the persistence module, we connect its birth time and death time resulting in a collection of horizontal lines that together resemble barcodes. *Persistence diagrams*, on the other hand, are scatter plots that record the birth-death times as x-y coordinates (e.g. in $\mathbb{R}^{2+} = \{(a, b) \in \{-\infty\} \cup \mathbb{R} \times \mathbb{R} \cup \{\infty\} | a < b\}$). So each point on the persistence diagram represents a generator. It is not hard to see that both persistence barcodes and diagrams encode the exact same information. More importantly, they offer more intuitive ways to study and compare different persistence modules, in terms of information that is not easily extractable otherwise [14].

So it is natural to now raise the question: how do we compare different persistence modules? The short answer is that we want some notion of distance in order to quantify how different or similar two persistence modules are. A standard way to do so is to use a construction called the *interleaving distance*.

Definition 1.27 (Interleaving distance)**.** Let $(\{V_\alpha\}, \{f^\beta_\alpha\})$ and $(\{W_\alpha\}, \{g^\beta_\alpha\})$ be two persistence modules over $\mathbb{R}$. We say they are ϵ-*interleaved* for some $\epsilon > 0$ if there exist families of linear maps $\{\phi_\alpha : V_\alpha \to W_{\alpha+\epsilon}\}$ and $\{\psi_\alpha : W_\alpha \to V_{\alpha+\epsilon}\}$ such that the following triangles commutes:

The *interleaving distance* between $(\{V_\alpha\}, \{f^\beta_\alpha\})$ and $(\{W_\alpha\}, \{g^\beta_\alpha\})$ is taken to be the infimum of all $\epsilon > 0$ such that $(\{V_\alpha\}, \{f^\beta_\alpha\})$ and $(\{W_\alpha\}, \{g^\beta_\alpha\})$ are



ϵ-interleaved.

Remark 1.28. Definition 1.27 can be a bit technical to work with. This inspires us to make use of more tools available to persistence modules. For example, we can consider the comparison between persistence diagrams since they summarise information about corresponding persistence modules. Two popular notions of such distance are Wasserstein distance and bottleneck distance. The main idea is to consider a matching or a transportation plan between points in two persistence diagrams and calculate the distance with respect to the 'best' such matching. We define the terms here for general persistence modules over a totally ordered set following the setup in [29]. We will also include the specific definitions when we look at (extended) persistent homology modules in relevant sections.

Recall that Gabriel's theorem (Theorem 1.23) tells us that every finite persistence module can be decomposed into direct sums of interval modules. And we will see in Section 1.2.2 that the persistence modules we consider in this thesis are finite. So in the following definitions, we will write the persistence modules in this form.

Definition 1.29 (Transportation plan)**.** Let $(\Theta, \leq)$ be a totally ordered set and let $\mathcal{P} = \bigoplus_{I_i \in S^\mathcal{P}} \mathcal{I}_{I_i}$ and $\mathcal{Q} = \bigoplus_{I_j \in S^\mathcal{Q}} \mathcal{I}_{I_j}$ be persistence modules over Θ. Then a *transportation plan* between $\mathcal{P}$ and $\mathcal{Q}$ is a triple $T = (\hat{S}^\mathcal{P}, \hat{S}^\mathcal{Q}, \sigma)$ where $\hat{S}^\mathcal{P} \subseteq S^\mathcal{P}, \hat{S}^\mathcal{Q} \subseteq S^\mathcal{P}$, and $\sigma : \hat{S}^\mathcal{P} \to \hat{S}^\mathcal{Q}$ is a bijection. The intervals in $\hat{S}^\mathcal{P}$ and $\hat{S}^\mathcal{Q}$ are called *matched intervals* in T and their complements $S^\mathcal{P} \backslash \hat{S}^\mathcal{P}$ and $S^\mathcal{Q} \backslash \hat{S}^\mathcal{Q}$ are called *unmatched intervals* in T.

To each transportation plan, we need to associate a cost in order to talk about "optimal" matching. This requires a metric, and we define it in an analogous way to L^p norms.

Definition 1.30 (Totally ordered metric space)**.** We call $(\Theta, \leq, \text{dist})$ a *totally ordered metric space* if

1. $(\Theta, \leq)$ is a totally ordered set;

2. dist is a metric over Θ such that if $\alpha \leq \beta \leq \gamma$, then $\mathrm{dist}(\beta, \gamma) \leq \mathrm{dist}(\alpha, \gamma)$ and $\mathrm{dist}(\alpha, \beta) \leq \mathrm{dist}(\alpha, \gamma)$.

Definition 1.31 (Distance)**.** Let $(\Theta, \leq, \mathrm{dist})$ be a totally ordered metric space, and let $\mathcal{I}$ and $\mathcal{I}'$ be two intervals over Θ. Then for $p \in [1, \infty)$, their *p-distance* is

$$\mathrm{dist}_p(\mathcal{I}, \mathcal{I}') = \left(\mathrm{dist}(\mathfrak{b}(\mathcal{I}), \mathfrak{b}(\mathcal{I}'))^p + \mathrm{dist}(\mathfrak{d}(\mathcal{I}), \mathfrak{d}(\mathcal{I}'))^p\right)^{1/p}.$$

Their ∞*-distance* is

$$\mathrm{dist}_\infty(\mathcal{I}, \mathcal{I}') = \max\left\{\mathrm{dist}(\mathfrak{b}(\mathcal{I})), \mathfrak{b}(\mathcal{I}'), \mathrm{dist}(\mathfrak{d}(\mathcal{I}), \mathfrak{d}(\mathcal{I}'))\right\}.$$

To define the cost associated to a transportation plan, we need to define a set of "ephemeral" intervals. This is the diagonal line on the persistence diagram for persistent homology for example. In a general persistence module, we denote such a set by Eph.

Definition 1.32 (Cost of a transportation)**.** Let $\mathcal{P} = \bigoplus_{I_i \in S^{\mathcal{P}}} \mathcal{I}_{I_i}$ and $\mathcal{Q} = \bigoplus_{I_j \in S^{\mathcal{Q}}} \mathcal{I}_{I_j}$ be persistence modules over a totally ordered metric space $(\Theta, \leq, \mathrm{dist})$. Let Eph denote the set of ephemeral intervals over Θ and let $T = (\hat{S}^{\mathcal{P}}, \hat{S}^{\mathcal{Q}}, \sigma)$ be a transportation plan between $\mathcal{P}$ and $\mathcal{Q}$. Then for $p \in [1, \infty)$, the *p-cost of* T is given by

$$\begin{aligned} c_p(T) = &\sum_{I_i \in \hat{S}^{\mathcal{P}}} \mathrm{dist}_p(\mathcal{I}_{I_i}, \mathcal{I}_{\sigma(I_i)})^p + \sum_{I_i \in S^{\mathcal{P}} \setminus \hat{S}^{\mathcal{P}}} \inf_{\mathcal{I} \in \mathrm{Eph}} \{\mathrm{dist}_p(\mathcal{I}_{I_i}, \mathcal{I})^p\} \\ &+ \sum_{I_j \in S^{\mathcal{Q}} \setminus \hat{S}^{\mathcal{Q}}} \inf_{\mathcal{I} \in \mathrm{Eph}} \{\mathrm{dist}_p(\mathcal{I}_{I_j}, \mathcal{I})^p\}, \end{aligned}$$

and the ∞*-cost of* T is given by

$$\begin{aligned} c_\infty(T) = \max\Bigg\{ &\sup_{I_i \in \hat{S}^{\mathcal{P}}} \left\{\mathrm{dist}_\infty(\mathcal{I}_{I_i}, \mathcal{I}_{\sigma(I_i)})\right\}, \sup_{I_i \in S^{\mathcal{P}} \setminus \hat{S}^{\mathcal{P}}} \left\{ \inf_{\mathcal{I} \in \mathrm{Eph}} \{\mathrm{dist}_\infty(\mathcal{I}_{I_i}, \mathcal{I})\}\right\}, \\ &\sup_{I_j \in S^{\mathcal{Q}} \setminus \hat{S}^{\mathcal{Q}}} \left\{ \inf_{\mathcal{I} \in \mathrm{Eph}} \{\mathrm{dist}_\infty(\mathcal{I}_{I_j}, \mathcal{I})\}\right\}\Bigg\}. \end{aligned}$$

Definition 1.33 (p-Wasserstein distance and bottleneck distance)**.** Let $\mathcal{P} = \bigoplus_{I_i \in S^{\mathcal{P}}} \mathcal{I}_{I_i}$ and $\mathcal{Q} = \bigoplus_{I_j \in S^{\mathcal{Q}}} \mathcal{I}_{I_j}$ be persistence modules over a totally ordered metric space $(\Theta, \leq, \mathrm{dist})$. Then for $p \in [1, \infty)$, the *p-Wasserstein distance between* $\mathcal{P}$ *and* $\mathcal{Q}$ is

$$W_p(\mathcal{P}, \mathcal{Q}) = \inf\{c_p(T) \mid T \text{ is a transportation plan between } \mathcal{P} \text{ and } \mathcal{Q}\}.$$

The *bottleneck distance between* $\mathcal{P}$ *and* $\mathcal{Q}$ is

$$W_\infty(\mathcal{P}, \mathcal{Q}) = \inf\{c_\infty(T) \mid T \text{ is a transportation plan between } \mathcal{P} \text{ and } \mathcal{Q}\}.$$

1.2.2 Relating to topology

Adopting tools from algebraic objects such as persistence modules to treat topological data is of major interest to researchers, especially in an interdisciplinary setting. A natural construction that links persistence modules to topological information is called *filtration*, through which we set up a sequence of topological spaces connected via some continuous maps. These, in turn, induce maps on a algebraic level, e.g. homology.

Definition 1.34 (Filtration)**.** Given a subset $T \subseteq \mathbb{R}$, a filtration X over T is a family $\{X_a\}_{a \in T}$ of topological spaces, parameterised by $a \in T$, such that $X_a \subseteq X_b$ for all $a \leq b \in T$.

The filtration considered in this section is called *sublevel sets*, defined to be $X_a = f^{-1}(-\infty, a]$ for some function $f : X \to \mathbb{R}$ on a topological space X and some threshold $a \in \mathbb{R}$. Then the family $\{X_a\}_{a \in \mathbb{R}}$ of topological spaces forms a filtration. More importantly, there is a natural inclusion map $i : X_a \to X_b$ for $a \leq b$. From here, Lemma 1.14 tells us that i induces a map $f_k^{a \to b} : H_k(X_a) \to H_k(X_b)$ for dimension k, and this gives us a persistence module. We observe that for $a = b$, $f_k^{a \to b} = id$ is an isomorphism. Using the definition of homology, we have that

$$H_k(a, b) \coloneqq \mathrm{im}(f_k^{a \to b}) = Z_k(X_a) / \left(Z_k(X_a) \cap B_k(X_b)\right),$$

which we will call the *k-th persistent homology group*.

Definition 1.35 (Homological critical value)**.** Given the setup above, a value $a \in \mathbb{R}$ is a *homological critical value* of f if for all $\epsilon > 0$, the map $f_k^{a-\epsilon \to a+\epsilon}$ is *not* an isomorphism for some k. In other words, some birth or death occurs at the threshold a.

From this point onward, we will assume that f is *tame*, i.e. f has only finitely many homological critical values and the homology group corresponding to each sublevel set is finite-rank.

Let $a_1 < a_2 < \cdots < a_m$ be the critical values of f. Then we can consider sequence s_i for $0 \leq i \leq m$, where $s_{i-1} < a_i < s_i$. Let $X_i \coloneqq X_{s_i}$. Since, as mentioned before, the homology groups are computed with field coefficients $\mathbb{Z}_2$, which equip them with a vector structure, we actually have a persistence module:

$$H_*(X_0) \xrightarrow{f_*^{0 \to 1}} H_*(X_1) \xrightarrow{f_*^{1 \to 2}} \cdots \xrightarrow{f_*^{m-1 \to m}} H_*(X_m),$$

where the (linear) maps between consecutive homology groups (which can be regarded as vector spaces) are induced by inclusion maps of spaces. This is a *persistent homology module*. The upshot of choosing such a set of threshold values is that this is the smallest set that allows us to comprehend all the topological changes occurring along this filtration without any ambiguity. For instance, we can tell that a component $\alpha \in H_*(X_i)$ is born at a_i if it is not in $\text{im}(f_*^{i-1\to i})$. Moreover, we can know that it dies at a_j if $f_*^{i\to j-1}(\alpha) \notin \text{im}(f_*^{i-1\to j-1})$, but $f_*^{i\to j}(\alpha) \in \text{im}(f_*^{i-1\to j})$. With this assumption in mind, we can create a collection of persistence diagrams, one for each dimension and each consisting of a finite multiset of points.

To differentiate persistence diagrams constructed for the sublevel set filtration from those involving super-level sets (which we will see when we talk about extended persistence in Section 1.3), from this point onwards, the term "persistence diagrams/barcodes" is reserved for the ones for the sublevel set filtration, whereas its counterpart in Section 1.3 will be called *extended persistence diagrams/barcodes*.

One of the most popular such functions f considered in literature are height functions. In fact, Turner et al. defined the *persistent homology transform* with filtration given by height functions with some fixed direction $v \in S^{n-1}$ in [28]. We will consider multiple choices of f in Chapter 2 as well. It might be intuitive to think that different filtrations corresponding to various functions should capture different features in topology. That is indeed the case, as we will explore more in Section 2.2.3. Thus, it is surprising that this is often overlooked in practice [27].

Before that, let us turn our attention back to comparisons between persistence diagrams. Recall that persistence diagrams can be thought of as multisets sitting inside $\mathbb{R}^{2+}$. Equivalently, we can consider them as subsets of $\mathbb{R}^{2+} \times \mathbb{N}$. The extra information carried by $\mathbb{N}$ is called the *multiplicity* of a point. To define a notion of distance between two persistence diagrams, we need to think about how points in one set relate to those in the other. This motivates the following definition of a transportation plan.

Definition 1.36 (Transportation plans between persistence diagrams)**.** Suppose $\mathcal{D} = \{(I, m) \in \mathbb{R}^{2+} \times \mathbb{N}$ and $\mathcal{D}' = \{(I', m') \in \mathbb{R}^{2+} \times \mathbb{N}\}$ are two finite persistence diagrams. Then a *transportation plan* between $\mathcal{D}$ and $\mathcal{D}'$ is a bijection

$$\widetilde{\sigma} : \mathcal{D} \cup \Delta((\mathbb{R}^*)^2) \times \mathbb{N} \to \mathcal{D}' \cup \Delta((\mathbb{R}^*)^2) \times \mathbb{N},$$

where $\Delta((\mathbb{R}^*)^2) \times \mathbb{N}$ is the set of points on the diagonal line in $\mathbb{R}^{2*}$, each with countably infinite copies. We define $\widetilde{\sigma}$ in the following process:

(1) Fix a bijection $\sigma : \mathcal{D} \supseteq \mathcal{B} \to \mathcal{B}' \subseteq \mathcal{D}'$, that is we partially match up points in $\mathcal{D}$ with points in $\mathcal{D}'$;

(2) For points in $\mathcal{D}\backslash\mathcal{B}$ and $\mathcal{D}'\backslash\mathcal{B}'$, we match them to the closest points in $\Delta(\mathbb{R}^{2*}) \times \mathbb{N}$. That is, for $(\mathfrak{b}, \mathfrak{d}) \in \mathcal{D}\backslash\mathcal{B} \cup \mathcal{D}'\backslash\mathcal{B}'$, we match them with $\left(\frac{\mathfrak{b}+\mathfrak{d}}{2}, \frac{\mathfrak{b}+\mathfrak{d}}{2}\right)$. Note that here we allow $\mathfrak{b} = -\infty$ or $\mathfrak{d} = \infty$. (3) For the rest of points in $\Delta(\mathbb{R}^{2*}) \times \mathbb{N}$, we map them to itself. Note that this is possible since we are allowing countably many points on the diagonal and $\mathcal{D}$, $\mathcal{D}'$ only contain finitely many points.

Remark 1.37. One should think of Definition 1.36 as moving points from one diagram to another and trying to capture the optimal cost of transporting these points. Hence the name.

Definition 1.38 (Wasserstein distance and bottleneck distance)**.** Consider $\widetilde{\sigma}$ transportations plans between persistence diagrams $\mathcal{D}$ and $\mathcal{D}'$. For $p \in [1, \infty)$, we define the *Wasserstein distance* to be

$$W_p(\mathcal{D}, \mathcal{D}') = \left(\inf_{\widetilde{\sigma}} \sum_{x \in \mathcal{D}} \|x - \widetilde{\sigma}(x)\|_p^p \right)^{1/p}. \tag{1.4}$$

The *bottleneck distance* is defined analogously for when $p \to \infty$:

$$d_B(\mathcal{D}, \mathcal{D}') = \inf_{\widetilde{\sigma}} \left(\sup_{x \in \mathcal{D}} \|x - \widetilde{\sigma}(x)\|_\infty \right). \tag{1.5}$$

Remark 1.39. One should notice from Definition 1.38 that the notions of distance here are analogous to L_p or L_∞ norms on function spaces defined on discrete sets. Furthermore, it is possible that for each transportation plan, the corresponding cost, i.e. $\sum_{x \in \mathcal{D}} \|x - \widetilde{\sigma}(x)\|_p^p$ for $p \in [1, \infty)$ and $\sup_{x \in \mathcal{D}} \|x - \widetilde{\sigma}(x)\|_\infty$ is infinite since we allow points to be born or die at $-\infty$ and ∞ respectively.

Remark 1.40. There is a more general form of Wasserstein distance, where we take different metrics for the plane (q) and the power (p). That is we have

$$W_{p,q}(\mathcal{D}, \mathcal{D}') = \left(\inf_{\widetilde{\sigma}} \sum_{x \in \mathcal{D}} \|x - \widetilde{\sigma}(x)\|_q^p \right)^{1/p}. \tag{1.6}$$

The corresponding bottleneck distance will be defined in the same way as in Equation 1.5 since that is what we will get if we take p, q approaching infinity.

Turner argued in [2] that the choice of p is directly related to the feature we care about when comparing diagrams (for instance taking p approach infinity means we care less about points that are extremely close to each other), while the choice of q does not matter as much. Moreover, we should take $p = q$ for several reasons including that it gives a more natural stability result. Furthermore, it has been shown empirically that taking $p = 1$ performs better than other choices of p when modelling and comparing shapes in $\mathbb{R}^2$ or objects and surfaces in $\mathbb{R}^3$ [28]. This is precisely the kind of data we are working with.

So in what follows, we will use W_1 to compare different persistence diagrams (or extended persistence diagrams for that matter).

The classic stability theorem is related to the bottleneck distance. It was proven by Cohen-Steiner, Edelsbrunner, and Harer [5], and we include it here.

Theorem 1.41 (Stability of sublevel set filtration)**.** *Suppose $f, g : X \to \mathbb{R}$ are continuous tame functions defined on a triangulable space X, and consider the L_∞-norm*

$$\|f - g\|_\infty = sup_{x \in X} |f(x) - g(x)|.$$

Then we have

$$d_B(Dgm_k(f), Dgm_k(g)) \leq \|f - g\|_\infty,$$

where $Dgm_k(f), Dgm_k(g)$ are the k-dimensional persistence diagrams of f and g respectively.

1.3 Extended Persistence

On top of considering only the sublevel sets, we can also include the superlevel sets and consider relative homology when we build our persistence modules. This gives us more information about the topology of a given space. More importantly, we will see that the Wasserstein distance and bottleneck distance between two extended persistence diagrams will always be finite, which makes comparisons easier in practice.

Intuitively, we can think of extended persistent homology as first scanning outwards and gradually seeing more information, and then moving back down until we get back to the starting point. In practice, this is achieved by involving relative homology. Denote by $X^a = f^{-1}[a, \infty)$ the superlevel set of some tame function f at some threshold $a \in \mathbb{R}$. Again suppose $a_1 < a_2 < \cdots < a_m$ are the

homological critical values of f. We can find the same sequence $\{s_i\}_{i=0}^m$ as before. Now consider the following persistence module:

$$\begin{array}{ccccccc} 0 = H_*(X_0) & \to & H_*(X_1) & \to \cdots \to & H_*(X_m) & = & H_*(X) \\ & & & & \downarrow & & \\ 0 = H_*(X, X^0) & \leftarrow & H_*(X, X^1) & \leftarrow \cdots \leftarrow & H_*(X, \emptyset) & = & H_*(X) \end{array} \tag{1.7}$$

Since the beginning and the end of the chain are both 0, all components that are born along the way will eventually die. In fact, there are three different possibilities of such components depending on their birth and death times. We call a component that is born and die scanning outwards (i.e. in the homology groups of sublevel sets) *ordinary*. One with birth and death time both occurring tracing inwards (i.e. in the relative homology groups) is called *relative*. Finally, we follow the exposition in [29] and call the ones that are born going outwards and die coming inwards *essential*. These components are often referred to as *extended* in the literature as well.

Following the setup given in Section 1.2.1, we can rewrite the above in terms of persistence modules. In particular, we can think of Θ as two copies of $\mathbb{R}$. More precisely, we follow the exposition in [29] and define $O = \{(t, \mathrm{Ord}) : t \in \mathbb{R}\}$ and $R = \{(t, \mathrm{Rel})t \in \mathbb{R}\}$. Then we set $\Theta = O \cup R$ and define a total order as follows:

- $(s, \mathrm{Ord}) \leq (t, \mathrm{Ord})$ if $s \leq t$;
- $(t, \mathrm{Rel}) \leq (s, \mathrm{Rel})$ if $s \leq t$;
- $(s, \mathrm{Ord}) < (t, \mathrm{Rel})$ for all $s, t \in \mathbb{R}$.

For the *k-th extended persistence module* V_k over Θ, we assign vector spaces $V_{(t,\mathrm{Ord})} = H_k(X_t) = H_k(X_t, \emptyset)$ and $V_{(t,\mathrm{Rel})} = H_k(X, X^t)$. The maps between vector spaces are induced by inclusion maps between corresponding pairs of spaces. Then linking back to the terminologies above, we have that the ordinary and relative classes are the ones whose birth and death are both in O and R respectively. The essential classes, on the other hand, would have birth (s, Ord) and death (t, Rel). We can further classify the essential classes into two subcategories, Ess^+ and Ess^-, where the former has $s < t$ and the latter satisfies $t < s$. Moreover, we have that

$$V_k = \mathrm{Ord}_k(X, f) \oplus \mathrm{Rel}_k(X, f) \oplus \mathrm{Ess}^+(X, f) \oplus \mathrm{Ess}^-(X, f),$$

where each summand is a submodule cover a copy of $\mathbb{R}$ corresponding to the labelled class [29]. We can again visualise the birth and death of the components

in each of these classes as a point in $\mathbb{R}^2$ (note that here we do not need $\mathbb{R}^{2+}$ since no component has non-finite birth or death time). The resulting diagram is the *extended persistence diagram*.

With the newly introduced setup, we need to redefine the Wasserstein distance and bottleneck distance. This is because

(a) our totally ordered set has changed, which means we need to redefine a metric on $\Theta = O \cup R$; and

(b) it no longer makes sense to match points in an extended persistence diagram to points on the 'diagonal' without specifying the class they belong to. This motivates the definition of 'ephemeral' points or 'ephemeral' interval modules.

To address point (a), we impose a natural metric on Θ:

$$\begin{aligned} \operatorname{dist}((t,\mathrm{Ord}),(s,\mathrm{Ord})) &= |t-s|, \text{ for all } s,t \in \mathbb{R}; \\ \operatorname{dist}((t,\mathrm{Rel}),(s,\mathrm{Rel})) &= |t-s|, \text{ for all } s,t \in \mathbb{R}; \\ \operatorname{dist}((t,\mathrm{Ord}),(s,\mathrm{Rel})) &= \infty, \text{ for all } s,t \in \mathbb{R}. \end{aligned}$$

For point (b), we define *ephemeral interval modules* as a union of 'diagonal' points:

$$\begin{aligned} \mathrm{Eph} =& \{I_{[(t,\mathrm{Ord}),(t,\mathrm{Ord}))} | t \in \mathbb{R}\} \cup \{I_{[(t,\mathrm{Rel}),(t,\mathrm{Rel}))} | t \in \mathbb{R}\} \\ & \cup \{I | \mathfrak{b}(I) = (t,\mathrm{Ord}), \mathfrak{d}(I) = (t,\mathrm{Rel}), t \in \mathbb{R}\} \end{aligned}$$

In particular, one can see that the last class of interval modules whose birth is in Ord and death in Rel corresponds to the essential class (regardless of positive or negative signs). We can think of them as tiny specks in the space that contribute to essential classes yet so small that we can disregard them.

With the two problems taken care of, a *transportation plan* between extended persistence modules is then similar to that of persistence modules/diagrams. We still want to match up the points on the extended persistence diagrams. Furthermore, within each class (Ord, Rel, Ess^+ and Ess^-), a transportation plan is exactly the same as that for persistence modules except instead of matching points to the diagonals, we can match them to the ephemeral points of the relevant class. In fact, [29] shows that the Wasserstein/bottleneck distances between extended persistence modules/diagrams are exactly the sum/maximum of the Wasserstein/bottleneck distances for each class.

Theorem 1.42 (Wasserstein/bottleneck distance for extended persistence modules [29])**.** *Let $\mathcal{P}$ and $\mathcal{Q}$ be two extended persistence modules of the same dimension. Then we have $\mathcal{P} = Ord(\mathcal{P}) \oplus Rel(\mathcal{P}) \oplus Ess^+(\mathcal{P}) \oplus Ess^-(\mathcal{P})$, and similarly $\mathcal{Q} = Ord(\mathcal{Q}) \oplus Rel(\mathcal{Q}) \oplus Ess^+(\mathcal{Q}) \oplus Ess^-(\mathcal{Q})$.*

*It then follows that the p-*Wasserstein distance *($1 \leq p < \infty$) between $\mathcal{P}$ and $\mathcal{Q}$ is:*

$$\begin{aligned} W_p(\mathcal{P}, \mathcal{Q})^p =& W_p(Ord(\mathcal{P}), Ord(\mathcal{Q}))^p + W_p(Rel(\mathcal{P}), Rel(\mathcal{Q}))^p \\ &+ W_p(Ess^+(\mathcal{P}), Ess^+(\mathcal{Q}))^p + W_p(Ess^-(\mathcal{P}), Ess^-(\mathcal{Q}))^p. \end{aligned}$$

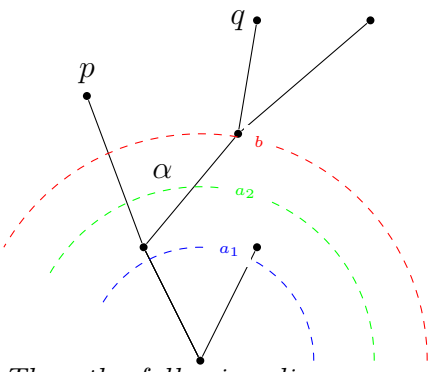
Furthermore, the bottleneck distance *between $\mathcal{P}$ and $\mathcal{Q}$ is:*

$$\begin{aligned} W_\infty(\mathcal{P}, \mathcal{Q}) = \max \big\{ & W_\infty(Ord(\mathcal{P}), Ord(\mathcal{Q})), W_\infty(Rel(\mathcal{P}), Rel(\mathcal{Q})), \\ & W_\infty(Ess^+(\mathcal{P}), Ess^+(\mathcal{Q})), W_\infty(Ess^-(\mathcal{P}), Ess^-(\mathcal{Q})) \big\}. \end{aligned}$$

Many duality results also emerge when we deal with extended persistence. Below we prove a 'mini' duality result for when the data is given by a tree diagram just to give some flavour. The proof relies on the fact that a tree is a connected component without cycles. However, thanks to Poincaré duality and Lefschetz duality theorems for ordinary homology, we have duality results in much more general settings. The former gives an isomorphism between the k-th homology and $(n-k)$-th cohomology groups of a manifold, while the latter connects the k-th cohomology (and respectively, homology) groups of a manifold M with boundary ∂M to the $(n-k)$-th relative homology (and respectively, cohomology) groups of the pair $(M, \partial M)$. Excision theorem says that under certain conditions, we are allowed cut off a subspace $U \subseteq A \subseteq X$ from a pair (X, A) of spaces such that $H_k(X, A) \cong H_k(X \backslash U, A \backslash U)$. So combining excision theorem with Lefschetz duality theorem, we can get back from cohomology groups to relative homology groups. This offers us insights on how Ordinary and Relative classes from the extended persistence modules of different dimensions are related. We will also include some of these duality results that appeared in [25] and [6].

Definition 1.43. A *finite tree* $T \subseteq \mathbb{R}^n$ is an undirected graph embedded in $\mathbb{R}^n$ with finitely many vertices that is connected and acyclic.

Theorem 1.44 ('Mini' duality theorem)**.** *Let $T \subseteq \mathbb{R}^n$ be a finite tree and let $0 < a < b$.*



Denote by T^a and T^b the superlevel set $\phi^{-1}[a,\infty)$ and $\phi^{-1}[b,\infty)$, where $\phi : T \to \mathbb{R}$ is the radial function centred at the root r of the tree defined by $\phi(x) = \|x - r\|_2$ and $\|\cdot\|_2$ is the Euclidean distance inherited from $\mathbb{R}^n$. See the illustrated graph on the left.

Then the following diagram commutes:

$$\begin{array}{ccc} H_1(T,T^b) & \xrightarrow{\cong} & \widetilde{H}_0(T^b) \\ \downarrow{\scriptstyle i_*} & & \downarrow{\scriptstyle i_*} \\ H_1(T,T^a) & \xrightarrow{\cong} & \widetilde{H}_0(T^a) \end{array}$$

Proof. Since (T, T^t) is a good pair, we have the following LES:

$$\cdots \to \widetilde{H}_1(T) \xrightarrow{j_1} H_1(T,T^t) \xrightarrow{\partial} \widetilde{H}_0(T^t) \xrightarrow{i_0} \widetilde{H}_0(T) \to \cdots .$$

Since T is acyclic and connected, we have that $\widetilde{H}_1(T) = \widetilde{H}_0(T) = 0$. This implies that $H_1(T,T^t) \overset{\phi_t}{\cong} \widetilde{H}_0(T^t)$.

In fact, we can write down an explicit isomorphism: we first label the nodes from 1 to n. Observe that for each element α in $H_1(T,T^t)$, we have $\phi^{-1}(t)$ intersect with α at precisely two points. These trace to two connected components in T^t, call them σ_p and σ_q, where $p, q \in \{1, 2, \cdots, n\}$ correspond to the vertex labels contained in σ_p and σ_q respectively. Note that there could be multiple choices of such p and q, but they represent the same connected components. Then we define $\phi_t(\alpha) \coloneqq \sigma_p + \sigma_q$. It is clear that this is a well-defined isomorphism.

Taking $t = a, b$ respectively, we have for $\alpha \in H_1(T,T^b)$, there can only be two possibilities. Either $i_*(\alpha) = 0 \in H_1(T,T^a)$ (See the dotted line labeled a_1 in the figure above), or $i_*(\alpha) = \alpha \in H_1(T,T^a)$ (See the dotted line labeled a_2 in the figure above).

Let $\phi_b(\alpha) = \sigma_p + \sigma_q$ for some p, q. In the first case, we then have σ_p and σ_q become the same connected component in T^a. Hence, we have

$$i_*(\sigma_p + \sigma_q) = 2\sigma_p = 0 \in \widetilde{H}_0(T^a),$$

since we take coefficients in $\mathbb{Z}_2$. This is equal to

$$\phi_a(i_*(\alpha)) = \phi_a(0) = 0.$$

In the second case, we have that σ_p and σ_q remain two separate connected components in T^a, and hence,

$$i_*(\sigma_p + \sigma_q) = \sigma_p + \sigma_q \in \widetilde{H}_0(T^a).$$

This is equal to

$$\phi_a(i_*(\alpha)) = \phi_a(\alpha) = \sigma_p + \sigma_q.$$

We have shown that in both cases, we get $\phi_a \circ i_* = i_* \circ \phi_b$. Therefore, the given diagram commutes. □

This phenomenon occurs in higher dimensions as well.

Theorem 1.45 (General duality theorem [25])**.** *Let $\mathbb{X} = \{X_t\}$ be a filtration of a simplicial complex X. Denote by $PD_k(\mathbb{X})$ the corresponding k-dimensional persistence diagram. Define $PD_k^0(\mathbb{X})$ to be the restriction of $PD_k(\mathbb{X})$ to the classes with finite lifetime, and define $PD_k^\infty(\mathbb{X})$ to be the restriction of $PD_k(X)$ to the classes with infinite lifetime. Furthermore, let $PD_k(X, \mathbb{X})$ be the k-dimensional persistence diagram of the relative homology groups $H_k(X, X_t)$ and define $PD_k^0(X, \mathbb{X})$ and $PD_k^\infty(X, \mathbb{X})$ in a similar way. Then we can have*

$$PD_k^0(\mathbb{X}) = PD_{k+1}^0(X, \mathbb{X}),$$

and

$$PD_k^\infty(\mathbb{X}) = PD_k^\infty(X, \mathbb{X}).$$

Corollary 1.46. *Let M be a finite simplicial complex with vertex set S. Then denote by $|M|$ its geometric realisation. Let $|f| : |M| \to \mathbb{R}$ be a function such that the interior of each simplex in $|M|$ is given by linear interpolation of the values of f on its vertices. Then there is a bijection $\phi : Ord_k(M, f) \to Rel_{k+1}(M, -f)$ given by*

$$\phi\left(\mathcal{I}_{[(b,Ord),(d,Ord))}\right) = \mathcal{I}_{[(-d,Rel),(-b,Rel))}.$$

Another useful result for reducing the complexity of computing extended persistence modules of a manifold is given by the symmetry theorem below.

Theorem 1.47 (Symmetry theorem in [29])**.** *Let M be a triangulated n-manifold, and let $f : M \to \mathbb{R}$ be a piecewise-linear function defined on the vertices of M and extended linearly on the edges by interpolation. Then we have*

$$Ord_k(M, f) \xrightarrow{\cong} Ord_{n-k-1}(M, -f) : \mathcal{I}_{[(b,Ord),(d,Ord))} \mapsto \mathcal{I}_{[(-d,Ord),(-b,Ord))},$$
$$Ess_k(M, f) \xrightarrow{\cong} Ess_{n-k}(M, -f) : \mathcal{I}_{[(b,Ord),(d,Rel))} \mapsto \mathcal{I}_{[(-d,Ord),(-b,Rel))},$$
$$Rel_k(M, f) \xrightarrow{\cong} Rel_{n-k+1}(M, -f) : \mathcal{I}_{[(b,Rel),(d,Rel))} \mapsto \mathcal{I}_{[(-d,Rel),(-b,Rel))}.$$

Remark 1.48. In the case where M is an n-manifold and $f : M \to \mathbb{R}$ is a real-valued function, Theorem 1.45 and 1.47 combined gives a manifold symmetry result in [6], which says that the k-th dimensional extended persistence diagram $\mathrm{XPH}_k(M, f)$ and the $(n-k)$-th dimensional extended persistence diagram $\mathrm{XPH}_{n-k}(M, f)$ are reflections of each other along the line of ephemeral points for all dimensions $k \geq 0$. Since for an $n-$manifold M with boundary, its boundary ∂M is an $(n-1)$-manifold. This symmetry theorem then allows us to compare $\mathrm{Ess}_0(\partial M, f)$ and $\mathrm{Ess}_{n-1}(\partial M, f)$, which we will do in Chapter 3.

1.4 Radial distance filtration

In this section, we develop some theories and discuss some properties regarding (extended) persistence homology with respect to radial filtrations. This allows us to explore a concrete example to gain a better understanding of the somewhat abstract definitions given in Section 1.2 and 1.3. Furthermore, as one can see in Figure 1.1, optical discs, the bright spots in the retinal images, can be thought of as the "root" of vessels expanding outwards. This shape naturally leads us to consider using radial filtration centred at the centre of the optical disc, and this is what we trialled in Section 2.2.3. Therefore, the second subsection (Section 1.4.2) is devoted to describe how we interpret binary digital images in practice and address some practical concerns when computing radial extended persistence homology.

1.4.1 Radial (extended) persistence

Consider radial functions on geometric simplicial complexes X embedded in $\mathbb{R}^n$. Fix a centre $c \in \mathbb{R}^n$, we define the *radial function*

$$\begin{aligned} \rho : X &\to \mathbb{R} \\ x &\mapsto \|x - c\|_2 = \sqrt{(x-c) \cdot (x-c)}, \end{aligned}$$

where $\cdot$ is the inner product inherited from $\mathbb{R}^n$. Suppose X can be described as a finite simplicial complex, the filtration $X(c)$ is a one-parameter family of subspaces of X parameterised by a radius r, i.e.

$$X(c)_r \coloneqq \left\{x \in X \middle| \|x - c\|_2 \leq r\right\}.$$

We denote by $\mathrm{RPD}_k(X(c))$ the k-dimensional radial persistence diagram obtained by computing how the k-dimensional homology groups change along the filtration over radius r. Note that the set $X(c)_r$ can be deformation retracted to the set $\{\Delta \in X | \forall x \in \Delta, |x - c| \leq r\}$ of simplices in X, and hence they have the same homology groups. This means that our persistence diagram is well-defined as homology groups are computed based on the simplicial structure defined in the latter set.

Lemma 1.49. *Fix a finite simplicial complex X embedded in $\mathbb{R}^n$. The map from the set of centres to k-dimensional radial persistence diagrams $v \mapsto RPD_k(X(c))$ is Lipschitz with respect to Euclidean metric on $\mathbb{R}^n$ and bottleneck distance on persistence diagrams.*

Proof. Let $\rho_c(x) = \|x - c\|_2$ be the radial function centred at $c \in \mathbb{R}^n$, and let $c_1, c_2 \in \mathbb{R}^n$. Then we have

$$|\rho_{c_1}(x) - \rho_{c_2}(x)| = \left|\|x - c_1\|_2 - \|x - c_2\|_2\right| \leq \|(x - c_1) - (x - c_2)\|_2 = \|c_1 - c_2\|_2,$$

where the inequality comes from reverse triangle inequality. Since this holds for all $x \in X$, we have that

$$|\rho_{c_1} - \rho_{c_2}|_\infty \leq \|c_1 - c_2\|_2. \tag{1.8}$$

Since X is finite, there exist finitely many homological critical points, and hence any function defined on X is tame. This means we can apply Theorem 1.41. By the stability theorem, we have that the bottleneck distance between $\mathrm{RPD}_k(X(c_1))$ and $\mathrm{RPD}_k(X(c_2))$ satisfies

$$d_B(\mathrm{RPD}_k(X(c_1)), \mathrm{RPD}_k(X(c_2))) \leq |\rho_{c_1} - \rho_{c_2}|_\infty. \tag{1.9}$$

Combining Equation (1.8) and (1.9), we get

$$d_B(\mathrm{RPD}_k(X(c_1)), \mathrm{RPD}_k(X(c_2))) \leq \|c_1 - c_2\|_2, \tag{1.10}$$

as required. □

Remark 1.50. Since Lipschitz continuity implies continuity, we have shown that the map $c \mapsto \mathrm{RPD}_k(X(c))$ is continuous. This means the *radial persistent homology transform*

$$\begin{aligned}\mathrm{RPHT} : \mathbb{R}^n &\to \mathrm{RPD}(X) \\ v &\mapsto (\mathrm{RPD}_0(X(c)), \mathrm{RPD}_1(X(c)), \cdots, \mathrm{RPD}_n(X(c)))\end{aligned}$$

is continuous function. Moreover, the equivalent statement, replacing persistence diagrams with extended persistence diagrams, holds as well with little modification. And we can define the *extended radial persistent homology transform*

$$\begin{aligned}\mathrm{XRPHT} : \mathbb{R}^n &\to \mathrm{XRPD}(X) \\ v &\mapsto (\mathrm{XRPD}_0(X(c)), \mathrm{XRPD}_1(X(c)), \cdots, \mathrm{XRPD}_n(X(c))),\end{aligned}$$

where $\mathrm{XRPD}_k(X(c))$ is the k-dimensional extended persistence diagram with respect to the radial filtration centred at $c \in \mathbb{R}^n$.

Another useful result that justifies the use of radial filtration is the fact that radial persistent homology transform is injective. The proof is straightforward: setting the centre in different locations inside or outside the space X will result in different 0-dimensional essential classes. So in fact, simply checking the 0-dimensional radial persistence diagram is enough to differentiate radial filtrations centred at two different locations. Since the radial persistent homology transform includes information about persistence diagrams of all possible dimensions, including the 0-dimensional one, we can conclude that RPHT is indeed injective. The same statement jumping to extended radial persistent homology transform holds as well, since extended persistence diagrams encode all (and more) information contained in the ordinary persistence diagram.

Below we include an example as an illustration of how we compute the persistence diagram (and barcode) with respect to radial filtration.

Example 1.51. Consider Diagram 1.4:

Setting v_0 as the centre, we can obtain a filtration on the top right using the radial function. The containment symbol $\subseteq$ shows that as the radius increases, we see more and more information (vertices and edges). We can also calculate the homology for each sublevel set. For instance, we can see that vertex v_2 was born at (radius) b and dies at (radius) d when it merges into the same connected component as v_1 and v_4 via the edge connecting v_3 and v_4. The point (b, d) on the persistence diagram describes that. Similarly, we can describe the birth and

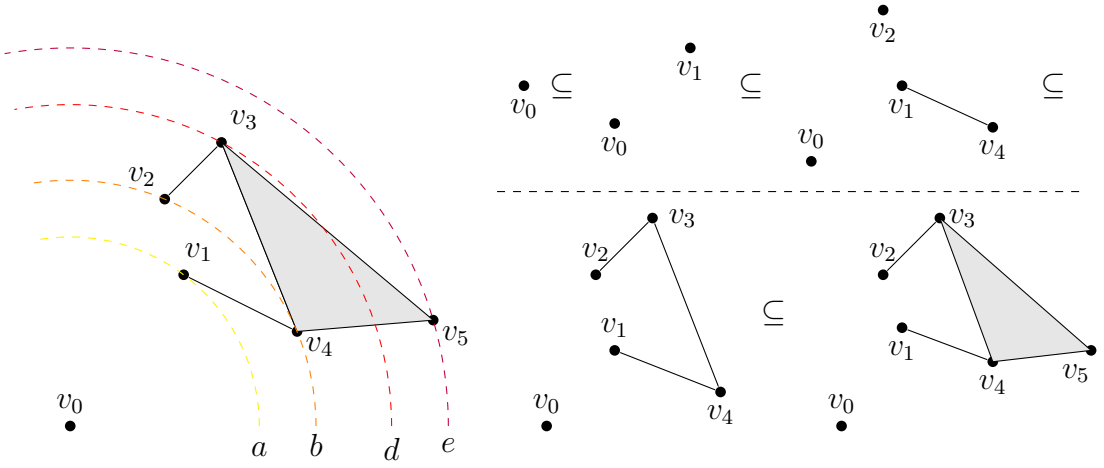


Figure 1.4: A 2D diagram (left) with 6 vertices and edges connecting some of them. The shaded area means the triangle formed by v_3, v_4 and v_5 is filled. Treating v_0 as the centre, we used different colors to indicate different distances each vertex is away from v_0. The letters a, b, d, e at the bottom of each curved line are the radii of the corresponding arcs centred at v_0. On the right is the corresponding filtration based on the radial function.

death of other components in this filtration and obtain the persistence diagram and summary in Figure 1.5.

Remark 1.52. What we saw in Example 1.51 is that all the homological critical points (or births and deaths) occur at the vertices. This is a desired property because then we only need to check if there is any homology change at each vertex without worrying about the edges or the interiors of formed triangles. In [28], it has been shown that height filtration mentioned in Section 1.2.2 satisfies this property. However, this fails to be true in general when we use radial functions. We illustrate this with a slightly modified version of Example 1.51.

Example 1.53. Consider Diagram 1.6:

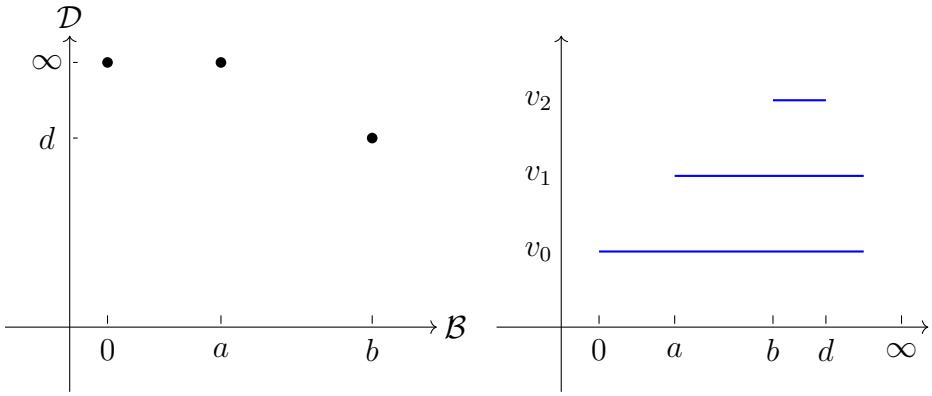


Figure 1.5: The persistence diagram (left) for the diagram in Figure 1.4 using the radial function. Since there is no cycle in the digram, we only need to consider the 0-dimensional persistence diagram. The $\mathcal{B}$-axis indicates the birth time, i.e. the radius at which a component is born, and the $\mathcal{D}$-axis is for the death time. Note that two points on the persistence diagram with infinity death time refer to the two connected components in the original diagram. The persistence barcode, encoding the same information, is given on the right.

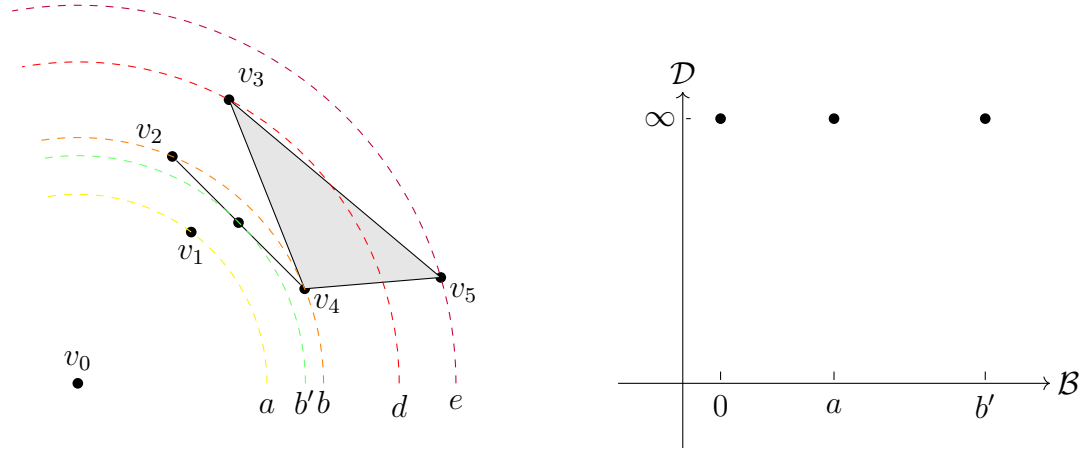


Figure 1.6: A modified diagram compared to the one in Figure 1.4. Note that it has the same vertices as in Figure 1.4, but the edges are connected slightly differently.

While the two diagrams look similar, by connecting vertex v_4 to v_2 instead of v_1, the filtration will see the midpoint (highlighted in the diagram) of the edge $[v_2, v_4]$ first before seeing the endpoints. This is because v_2 and v_4 are the same distance b away from v_0, and we note that this is the only scenario for $X \subseteq \mathbb{R}^2$

where the filtration sees the midpoint of an edge before its endpoint. Therefore, if we apply R(X)PHT on X, then to each edge in X, there will be a line of centre points c for which the end-points of that edge are equidistant to c. Things will become more complicated when we move to higher dimensions.

Now to upgrade everything to extended persistent homology, we make use of Theorem 1.18, which says that under mild conditions, a relative homology group is isomorphic to the reduced homology group of the quotient space, when we move down the chain (1.7) and consider relative homology. For example, a connected component, born at the smallest value it can take (i.e. when the connected component is first spotted by the filtration), will die at the largest value. So in Example 1.4, the connected component involving v_1 is born at a and will die at d in the extended persistent homology. Its corresponding extended persistence diagram will then have two points of the essential class, a point $(0, 0)$ on the diagonal for the connected component v_0, and a point (a, e) above the diagonal. Note that while v_0 "lived" for a long time, it is only a point in the space. This, in a way, justifies the definition of ephemeral interval modules.

1.4.2 Working with binary digital images

Binary images are $m \times n$ matrices or 2-dimensional arrays, with entries called *pixels* taking two values, 0 (black) or 1 (white). In our case, the structure we are interested in are the white vessels in the segmentation. Therefore, *foreground* is defined to be the set of pixels with value 1.

To implement radial filtration in our retinal images, we modified the algorithm described in [29]. The original algorithm [‡] is implemented in R and takes binary images as inputs and outputs the extended persistent homology of the foreground with respect to height filtrations for an even number of directions. We adjust it so that it can compute radial extended persistent homology given a single radial filtration. Figure 1.7 displays the workflow of the algorithm. In what follows, we will discuss some necessary theoretical justifications and highlight some practical considerations. Justifications for other parts of the algorithm can be found in Section 6 of [29].

To extract foreground information, it suffices to extract its boundary. There are many ways in image analysis to translate pixel information into structures

[‡]The corresponding R package is available at `https://github.com/james-e-morgan/xpht`.

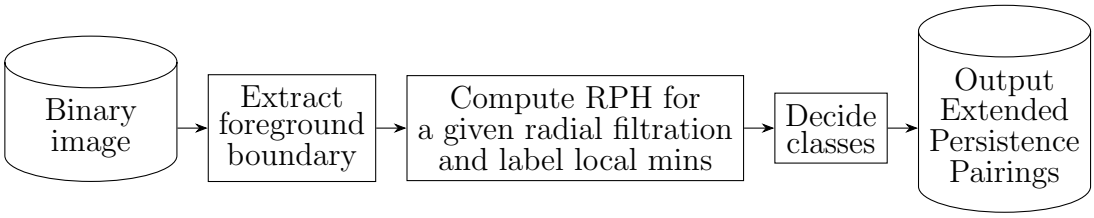


Figure 1.7: Workflow diagram to compute RXPH with respect to a given radial filtration on the foreground of an input binary image.

required, some may define vertices to be centres of foreground pixels while others choose the vertices of pixels. In our algorithm, we trace the midpoints of foreground pixels anticlockwise with respect to the foreground image, with some rules imposed, to encode the boundary information (see Figure 1.8). This way, the boundary will be a piecewise-linear curve, and if we follow the orientation on the boundary, the image will be always on our left hand side.

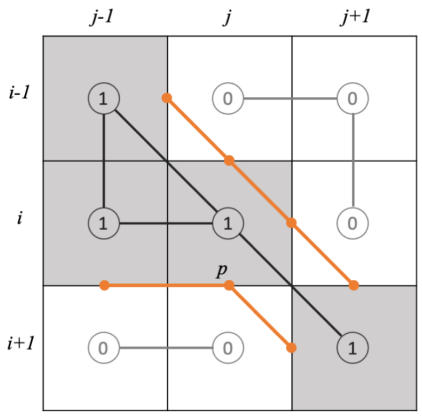


Figure 1.8: An example in [29] showing allowed edges for foreground boundary. Pixels labeled 1 are foreground pixels and orange lines are the traced boundary. Note that we do not allow connecting midpoints of edges vertically or horizontally through the pixels. For example, we cannot connect $p = \left(i + \frac{1}{2}, j\right)$ and $\left(i - \frac{1}{2}, j\right)$.

Now suppose for a given binary image input A, we have traced its boundary $\mathcal{B}$ by recording the midpoints of pixel boundaries it passes through. Fix a point c, and consider the radial function ρ_c centred at the point c. Since a circle can intersect with a straight line at at most two points, unlike in [29], we do not need to consider the case where more than three vertices on the same straight line take the same value ("co-linear" case in [29]). Furthermore, if two vertices on

the same straight line have the same value, then the filtration will see the middle of the straight line first (when the circle is tangent to the line) before seeing the vertices. We then make the following definition:

Definition 1.54. Let γ be a piecewise-linear curve in the boundary $\mathcal{B}$ of A with m vertices ordered cyclically $x_0, x_1, \cdots, x_m = x_0$. Then a vertex x_i is a *0-critical point* if $\rho_c(x_{i-1}) > \rho_c(x_i)$ and $\rho_c(x_i) < \rho_c(x_{i+1})$, where the subscripts are taken modulo m. We say x_i is a *local minimum* or alternatively, a *(+)-critical point* if there exists some $\epsilon > 0$ such that for all $a \in B(x_i, \epsilon) \cap A$, we have $\rho_c(x_i) \leq \rho_c(a)$.

The major changes for the algorithm to adopt radial filtration occurs in the process of labelling local minimums. Recall that in the first step, we translate the foreground image into a set of boundary components, the algorithm then computes only the values of the radial functions on the vertices of the foreground and hence is able to decide whether a vertex on the boundary is a homological critical point. By deciding local minimality, we can then translate the k-dimensional homological critical points of the boundary into birth and death parameters of different classes in dimension k. We will justify this in Chapter 3.

The reason we want to pick out local minimums is that for all 0-critical points, (+)-critical points are related Ordinary classes, while non-local minimums are related to Relative classes. To differentiate between these two types of critical points algorithmically, we use the following fact about orientation of triangles:

Lemma 1.55. *Let ΔABC be a triangle with positive area. Denote by* $\det(x, y)$*, where x, y are two-dimensional column vectors, the determinant of a 2×2 matrix whose columns are x and y. Then its vertices A, B, C are in an anticlockwise order if* $\det(C-B, A-B) > 0$*, and they are in a clockwise order if* $\det(C-B, A-B) < 0$.

Now we use Lemma 1.55 to decide local minimality of the 0-critical points.

Theorem 1.56. *Let $A \subseteq \mathbb{R}^2$ be a bounded set whose boundary is a disjoint union of piecewise-linear closed curves. Let γ be a piecewise-linear curve of A with vertices $x_0, x_1, \cdots, x_m = x_0$ traversed counterclockwise with respect to A. Denote by B the enclosed region of A by γ. Fix a point $c \in \mathbb{R}^2$ and consider the radial function ρ_c centred at c. A 0-critical point x_i is a (+)-critical point for ρ_c if and only if* $\det(x_{i-1} - c, x_i - c) < 0$.

Proof. Let x_i be a 0-critical point. Consider the triangle $\Delta x_i c x_{i-1}$. Then every point p on the interior of the edge $x_i c$ satisfies $\rho_c(p) < \rho_c(x_i)$.

Figure 1.9: An illustration of the two cases discussed in the proof: the first case (top left) where x_i is a (+)-critical point and the second case (top right) where x_i is not. The shaded area indicates where B is.

Suppose x_i is a (+)-critical point for ρ_c. Then there exists some $\epsilon > 0$ such that for all $a \in B(x_i, \epsilon) \cap A$, we have $\rho_c(a) \geq \rho_c(x_i)$. It follows that $p \notin B(x_i, \epsilon) \cap A$ for all p on $x_i c$. Hence, $\Delta x_i c x_{i-1}$ does not contain a subset of B, which means c is on the opposite side of $x_i x_{i-1}$ compared to B. Since γ is traversed in a counterclockwise order, we must have x_i, c, x_{i-1} in a clockwise order. Therefore, by Lemma 1.55, we have $\det(x_{i-1} - c, x_i - c) < 0$ as required.

Conversely, suppose x_i is not a (+)-critical point for ρ_c. Then the triangle $\Delta x_i c x_{i-1}$ contains a subset of B, i.e. c is on the same side of $x_i x_{i-1}$ as B. Again, since γ is traversed counterclockwise, we must now have x_i, c, x_{i-1} in a counterclockwise order, which by Lemma 1.55, we know that $\det(x_{i-1} - c, x_i - c) > 0$. □

Recall that in Example 1.6, we saw a case where the filtration sees the midpoint of an edge before it sees its endpoints. Since the point c is chosen to be in the middle of a pixel, such a case would occur if and only if a boundary segment is tangent to some circle centred at c (see Figure 1.10). However, in our algorithm, the persistent homology computation assumes that all homological critical values are at vertices of the simplicial complex. This means potential errors may occur in the extended persistence diagram.

To work around this issue, we propose three ways:

(i) for edges whose positions are similar to the ones in Figure 1.10, we add an extra vertex in the middle of it;

(ii) instead of considering normal radial filtration, we consider filtration induced by values of the vertices. That is, the filtration can only see an edge at value r if both vertices take values less than or equal to r.

(iii) instead of choosing the centre as the centre of a pixel, we add a small offset. That is, instead of using (x, y), which sits in the centre of a pixel, we

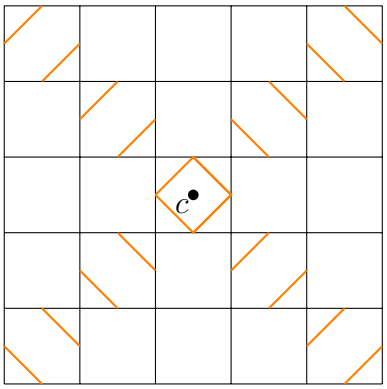


Figure 1.10: Edge positions that will lead to the case in Example 1.6.

consider the radial filtration computed with the centre being $(x + \epsilon, y + \epsilon)$ instead for some small ϵ.

Both suggestion (i) and (iii) should give us the actual persistence diagram. However, we can also show that option (ii) will lead to theoretically negligible error. This is because we only have finitely many vertices in the boundary of our images. For each of the edge in the case of Figure 1.10, the error in the birth/death value (taking the midpoint of an orange edge as the birth/death point versus using one of its endpoints) is bounded by $\frac{2-\sqrt{2}}{4}$. Hence, the total error between the genuine persistence diagram and the estimated persistence diagram is going to be relatively small.

Chapter 2

XPH in action: identification and verification based on retinal images

> "Men have forgotten this truth," said the fox. "But you must not forget it. You become responsible, forever, for what you have tamed. You are responsible for your rose..."
>
> *The petit prince*, Antoine de Saint-Exupéry

Cybersecurity has become one of the most rigorously studied topics in the current data driven world. Tasks such as data identification and verification are crucial in keeping systems safe. The two words might sound similar, but they actually refer to approaches to assessing different levels of security of some systems. Identification tasks generally assume a lower level of security required. Imagine you are a resident in an apartment in a modern building with a fingerprint door lock. To make sure that you are able to enter your own home, the property manager asked every resident of the building to record a few fingerprint samples in their database. Upon your arrival, the manager compares your face with the list of photos he has of the residents and matches you with a name on the list. This process is what we call *identification*. It assumes we are confident that the system is safe enough and no one is going to commit fraud. So a new sample is definitely going to match up with exactly one identity from the database, and

a simple piece of evidence, like your appearance, from the users is sufficient for a 1-to-n match-up process, but no in-depth verification of the validity of the evidence is required. For example, if you put on make up and try to pretend to be someone else, it is possible that the manager would not have noticed, After logging in your fingerprints, every time you get home and lay your finger on the door knob, it immediately recognised your fingerprints, and voila, "welcome home". On the other hand, to feel safe at your apartment, you would want no one other than you to be able to enter the property. That is, if someone else, accidentally or not, present their fingerprints at the door, the door should not open for them. To do this, the door system (whatever software it depends on) needs to run the input fingerprint through the system and recognise whether this fingerprint matches that of the person they claimed to be. This is verification. Unlike asking a question of "who are you?" as in the identification task, it begs the question: "are you actually who you said you are?".

Biometrics are data used for automatic verification and identification of individuals. Among them, retinal images are the most trustworthy characteristics for personal authentication and identification due to their uniqueness and stability over time [12]. Thus, they will be the focus of this thesis. We define our *identification and verification problems* as follows:

Problem 2.1. Consider a given database $\mathcal{D}$ of retinal images from m individuals, each individual providing n samples.

- *Identification problem:* We create a list $\mathcal{L}$ of retinal fundus images such that each of the m individuals provides exactly ℓ retinal image samples ($\ell < n$). Then $\mathcal{L}$ contains $m\ell$ samples. We randomly pick a image $\mathfrak{s}$ from $\mathcal{D}\backslash\mathcal{L}$, and compare it with the images in $\mathcal{L}$ to *identify* which individual matches the sample $\mathfrak{s}$. We are interested in the accuracy of our identification algorithm as well as the time it takes to make an identification.

- *Verification problem:* We randomly fix an individual i and pull out all the samples in $\mathcal{D}$ that belong to individual i. Denote by $\mathcal{I}$ the set of all such samples except for k of them ($k < n$), i.e. $\mathcal{I}$ contains $n - k$ samples all provided by individual i. Then we randomly choose a sample $\mathfrak{t}$ from $\mathcal{D}\backslash\mathcal{I}$ and compare it with the samples in $\mathcal{I}$ to *verify* whether $\mathfrak{t}$ comes from individual i. We are interested in the accuracy of our verification algorithm as well as the time it takes to complete the verification.

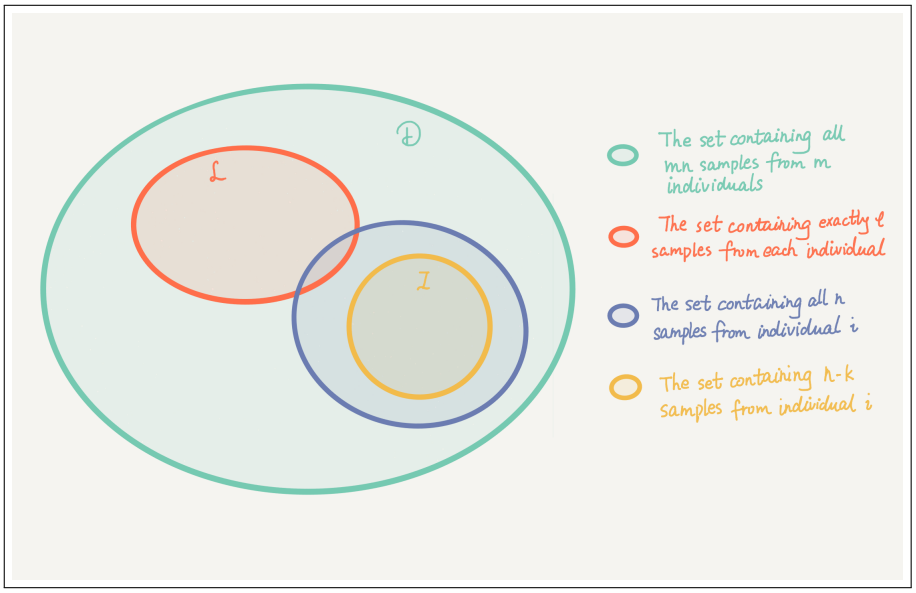


Figure 2.1: A Venn diagram giving an example of the set relations of $\mathcal{D}, \mathcal{L}$ and $\mathcal{I}$ defined in Problem 2.1.

The mainstream techniques for analysing retinal analysis and solving Problem 2.1 use biological feature extraction. Here, we propose an alternative approach using purely mathematics-based techniques, namely by computing (X)PH with respect to various filtrations and comparing resulting (extended) persistence diagrams. We note that the paper [30] proposed a method that combined both vascular and non-vascular features to perform personal verification, and achieved a high level of accuracy testing with the Retina Identification DataBase (RIDB). The article also recorded the average lapsed time to perform the task. This allows us to compare our method with the one proposed in [30]. What we discovered is that our method, with suitably chosen filtration, obtained a comparable accuracy while taking significantly shorter time to complete the verification task.

2.1 RIDB database

Using retinal images to perform personal identification and verification tasks has become a popular research topics over the past few years. Researchers have found that retinal images can also be used for early detections of cardiovascular diseases and other potential health problems [21].

Most methods proposed in the literature utilised certain features in retinal images. For instance, [8] uses intersections of blood vessels (corner detection)

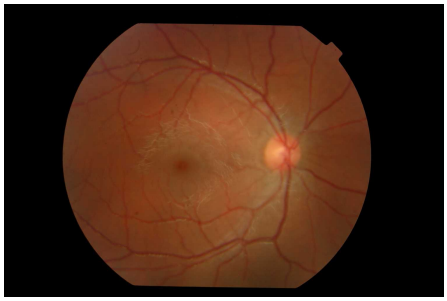
since they are invariant under eye movements, and [21] considers the optic disc and optic cup ratio. These methods are examples of two categories of commonly used techniques summarised in [30], vascular based and non-vascular based, with the former being more commonly considered by research groups. Extracting these features generally relies heavily on highly accurate segmentation, which is extremely difficult to achieve. Moreover, such methods can be computationally heavy and take a long time to perform.

This motivates us to consider an alternative method, which focuses more on the topology of retinal shapes so that it should theoretically be less sensitive to segmentation errors. Furthermore, we can easily choose a filtration that is rotation and translation invariant. This means we do not need to specifically look for features that are stable relative to eye movements. We will also see that our method performs the verification task significantly faster than the method in [30]. Finally, we note that most papers in the literature, including [30] despite its title, only considered the verification task defined in this thesis.

Retina Identification DataBase (RIDB) contains retinal fundus images obtained using the TOPCON-TRC camera [1]. Abdul Salam, A, et al. collected 100 healthy retinal images from 20 individuals, with each person contributing five samples. Each image has name of the form "IM00000i_j", which indicates the i-th sample from individual j for $1 \leq i \leq 5$ and $1 \leq j \leq 20$. The JPEG images have resolution 1504×1000 and can be from either a left eye or a right eye depending on the orientation. Below are two images from the RIDB Database as an illustration.

One advantage of the RIDB Database, as mentioned above, is that it contains multiple samples from each individual. This allows us to perform both identification and verification tasks. For identification, we take an arbitrary retinal image from the RIDB database and test if and how long it takes our method to find the person it matches to. For verification, we fix a person and see if our method can determine correctly whether or not a randomly chosen retinal image is consistent with the stored images for this person. To that end, we will eventually compute (extended) persistent homology with respect to some filtration to capture the topology of vessel structures in RIDB images and compare resulting persistence diagrams. But before this, we need to preprocess the images.

(a) Retinal fundus image IM000001_4



(b) Retinal fundus image IM000001_1

Figure 2.2: Two sample images from the RIDB Database from person 1 and person 4. Note that the rectangular tag on the top right of the images provides an orientation and gives a way to differentiate images captured from left (left image) and right (right image) eyes. Another noticeable feature is the bright circular spot around where the blood vessels converge on one side of the retina; this is the optic disc. The size of the discs are roughly equal for both eyes and across all individuals. Later, we will use the centre of the optic disc as a guide for image alignment in order to compare different persistence diagrams. The slightly darker circular parts, without visible vessels, in the middle of both images are the maculae – they appear in the middle as the individuals were required to gaze into the camera. Finally, the red-yellowish colour of the retinal images is typical among healthy individuals, with potential variations due to complex reasons that are out of the scope of this project [11]

2.2 Proposed analysis method

The flow diagram of our method is shown in Figure 2.3. The method has two main parts:

- the middle flow chart represents an registration module, where a database is created; and
- the top and bottom flow charts represent the tasks we perform using the database.

The registration module is an offline process where we encode the topological structure of given retinal images with (extended) persistence diagrams after first preprocessing the input images. The identification/verification module is an online process where images in question undergo a similar preprocessing and feature encoding procedure, and then compare with the existing database to decide whether the individual is genuine or an imposter.

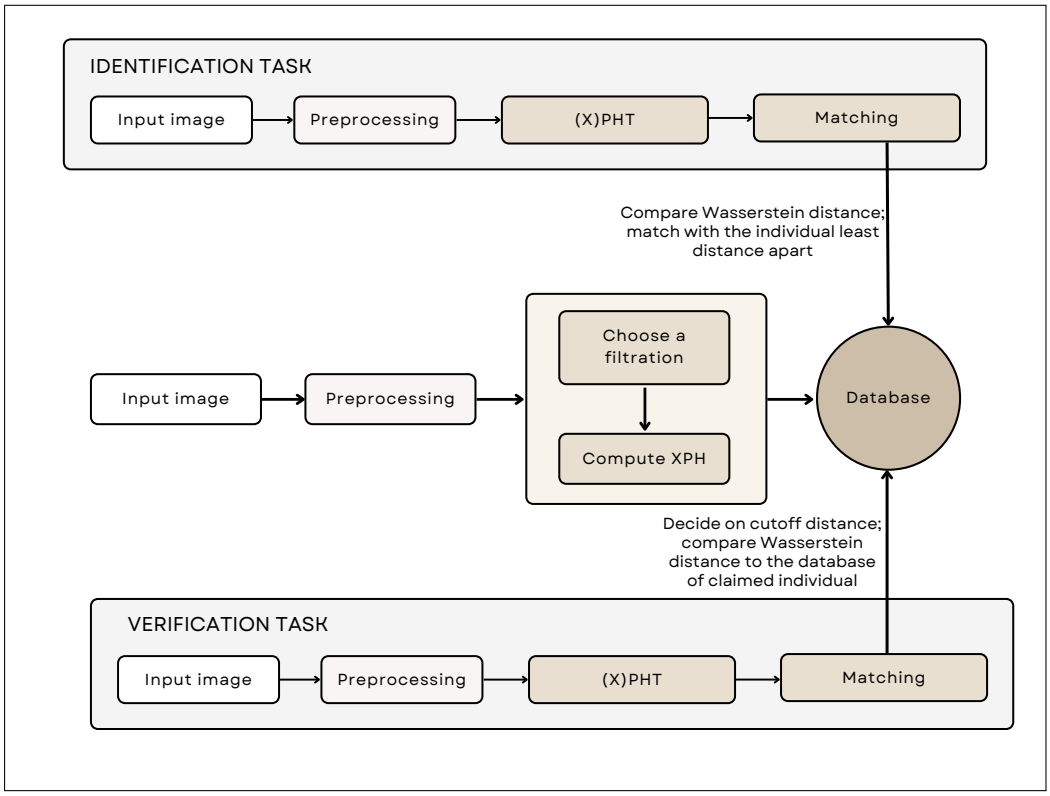


Figure 2.3: Flow chart of our method

2.2.1 Data preprocessing and image segmentation

The shapes that we are interested in and hoping to classify are the vessel structures in retinal images. Moreover, the code we use for computing extended persistent homology asks for a binary image as input. Thus, image segmentation, a process that classifies each image pixel as either vessel or non-vessel and therefore partitions the input image into two classes (black and white), is required as a first step.

Since the original database does not come with segmented images, we had to come up with our own. The method we adopted is called U-Net, a convolutional neural network (CNN) framework that is specifically designed for biomedical image segmentation. The original version and detailed explanation of the architecture of U-Net can be found in [24], and we modified it slightly for our purpose (see Figure 2.4 for the modified architecture) [†].

Our algorithm takes in an input image of size 192×192 and outputs a clas-

[†]Codes used can be accessed via: `https://github.com/JencyJ12/imageSegmentation.git`.

sified image of the same size after a sequence of encoder and decoder blocks. It comprises of two main processes: a contracting network and an upsampling process. In each encoder block (of the contracting network), we use repeated 3×3 convolution layers, each followed by a batch normalisation process (this is not included in the original network) for more stable convergence and a rectified linear unit (ReLU) as the activation function. A 2×2 max pooling operation is then performed to extract the most significant features, during which stage the image is downsized by a factor of 4. We also double the number of feature channels after each encoder block starting from 64. The upsampling process concatenates the image from a 2×2 up-convolution (which halves the number of filter channels), and the corresponding image from the contracting network to gain more information about the features. Each decoder block also includes two 3×3 convolution layers, each followed a batch normalisation and a ReLU.

Note that the original images have resolution 1504×1000 pixels, but the input images for our U-Net algorithm are of size 192×192. This is because the original images are too big for the algorithm to learn the detailed features. Moreover, the original images have some black pixels surrounding the retinal fundus scan (the region of interest), which means a large area of the original image does not contribute much information to the machine learning process. Therefore, we trimmed the original image to have size 1152×960 so that it leaves out some of the unimportant black pixels and then divided the trimmed images into 212 patches with step size 60 (that is two adjacent patches will have 132×192 or 192×132 overlaps).

To train our U-Net, we are fortunate to have Dr. Katharine Turner to trace the vessels (ground truths) on 52 retinal images using Procreate (see Figure 2.5). We also separate the RGB channels of the original images. What we notice is that the vessel structures are most apparent when we restrict to the green channels of original images (see Figure 2.6). Hence, we decide to train our model based on the ground truths and the green channels of original images to better assist our algorithm to learn which features to extract.

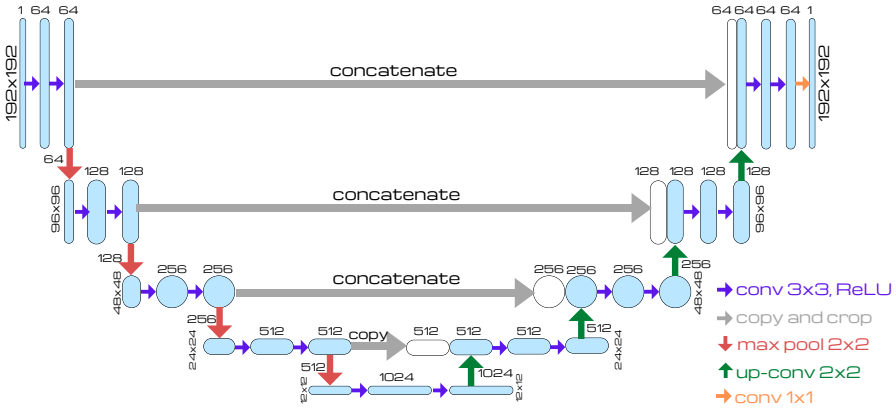
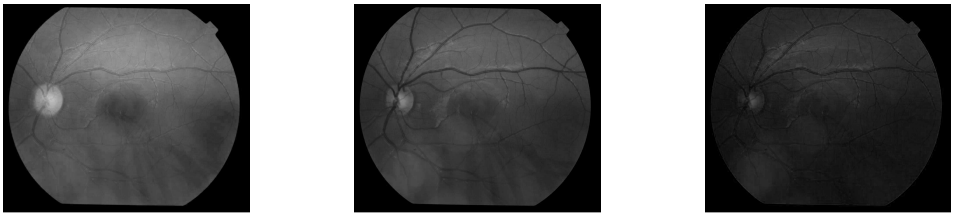


Figure 2.4: U-Net architecture modified to suit our purpose. The height of each blue/white rectangle represents the image size and the width represent the number of feature channels. Each group of 3 rectangles represent either an encoder block (going down with the red arrows) or a decoder block (going up with the green arrows).



(a) Red channel (b) Green channel (c) Blue channel

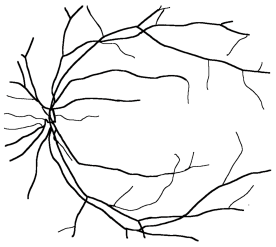
Figure 2.6: The red, green and blue channels of the retinal fundus image IM000004_8.

We train our model on a total of 11024 patches of original images together with their corresponding segmentation. After 20 epochs with batch size 16, we obtain a validation accuracy of 0.9900 and a mean Intersection-Over-Union (IoU) value of 0.8937125, where

$$\text{IoU} = \frac{\text{true\_positives}}{\text{true\_positives} + \text{false\_positives} + \text{false\_negatives}}.$$

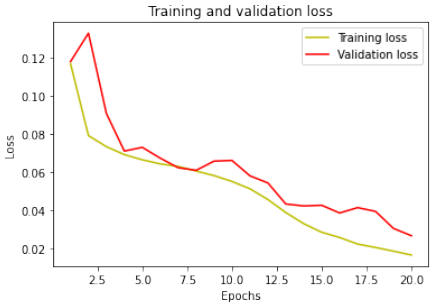
IoU is a standard metric to test the accuracy of semantic segmentation, e.g. an IoU score above 0.5 is generally considered a good score, with 1 being the theoretical maximum.

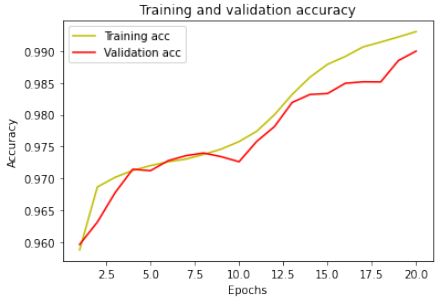
(a) IM000001_1



(b) IM000001_14

Figure 2.5: Examples of manually segmented ground truths using Procreate.





(a) U-Net training and validation loss. (b) U-Net training and validation accuracy.

Figure 2.7: Training and validation loss and accuracy at each epoch.

After training the models, we run it on patches (size 192×192) of original images restricted to the green channels and then concatenate the predicted segmentation back to images of original size (1152×960), averaging the overlapping regions. Figure 2.9 shows two examples of such segmentation. Observe that while they do capture the main vessel structures of images in Figure 2.2, some specks and discontinuities are visible. We suspect two reasons for this phenomenon:

(a) the manual segmentation is not consistent as it is not done by experts and has specks to start with (e.g. some can be spotted in Figure 2.5) – many biometric segmentation tasks in the literature start with ground truths traced by experts in the field;

(b) the original images are highly compressed in jpeg format. The compression artefacts are visible at 20×20 pixel level. This makes it difficult for machines to differentiate vascular versus non-vascular pixel values.

Since (extended) persistent homology is sensitive to discontinuities that lead to unwanted births of a new class, we first adopt some morphological operations to 'fix' the discontinuities to some extent. We use a combination of image dilation

and erosion to remove unwanted holes (discontinuities) in the image while keeping the modified object as close (topologically) to the original one as possible.

Definition 2.2. Let $A \subseteq \mathbb{R}^2$ be an object (set) and let B be a structuring element in a 2D plane. Then *dilation* of A by$B$ is given by

$$A \oplus B := \{x : B_x \cap A \neq \emptyset\},$$

where

$$B_x := \{b + x | b \in B\}$$

is the translation of B by x. Now *erosion* of A by B is given by

$$A \ominus B := \{x : B_x \subseteq A\}.$$

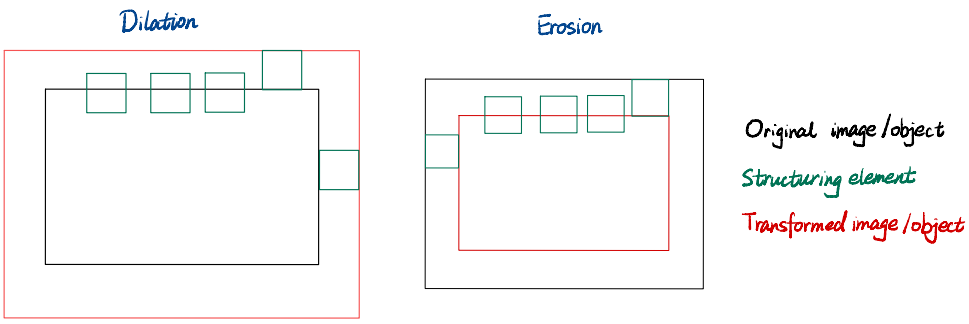
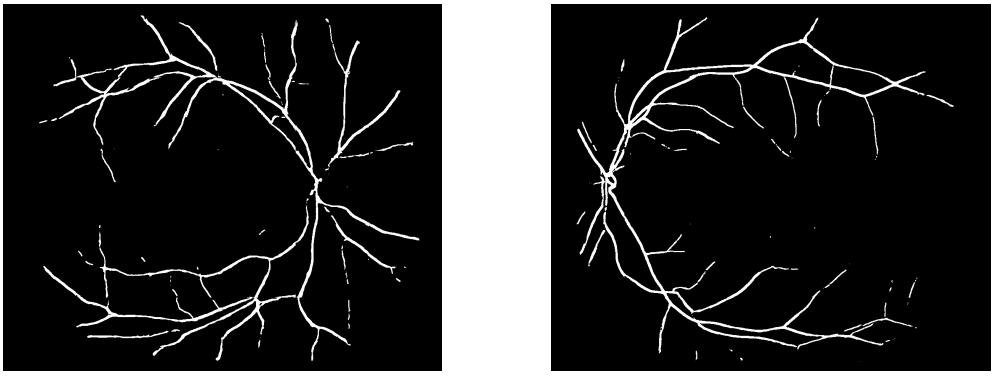


Figure 2.8: An example of image dilation and erosion using square shaped structuring elements.

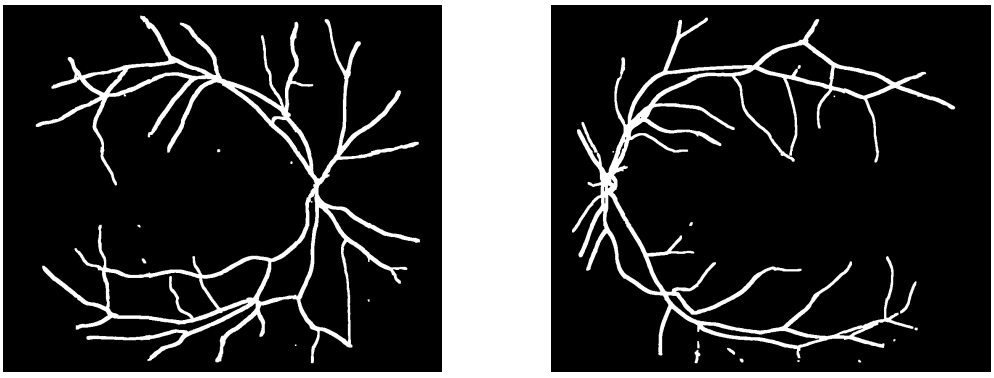
Commonly chosen structuring elements are discs centred at the origin. In our case, we find that a dilation using a square of size 5×5 followed by an erosion with a square of size 2×2 works reasonably well on the segmented retinal images.

Figure 2.10 shows the modified segmentation in Figure 2.9. The vessels are notably thicker, but the holes in the original images, while smaller, persist to exist. More importantly, some discontinuities disappeared as we desired. At the same time, the random noise/specks are more visible. However, this should not cause too much problem theoretically since the specks would have short lifespans and be close to the diagonal line on the (extended) persistence diagram. That means they should not contribute much to the Wasserstein distance when we compare two images.



(a) Segmented IM000001_1 (b) Segmented IM000001_4

Figure 2.9: Two examples of segmentation predicted by U-Net model.



(a) Modified segmented IM000001_1 (b) Modified segmented IM000001_4

Figure 2.10: The two examples of segmentation in Figure 2.9 after image dilation and erosion.

2.2.2 Measure of success

Permutation test

To check that we can use persistence diagrams to differentiate individuals, some statistical tests are required to test the hypotheses:

$\mathcal{H}_0$: the persistence diagrams for all individuals have the same distribution;

v.s.

$\mathcal{H}_a$: at least one of the individuals has persistence diagrams coming from a different distribution.

However, the complexity of the space of persistence diagrams means traditional hypothesis testing methods that assume underlying parametric models are not available [23]. Therefore, we opt for an alternative approach, which examines the exact kind of hypothesis as above.

To perform a permutation test, we use the sum of within group distance, that is the sum of all 1-Wasserstein distances between persistence diagrams computed using samples from the same individual. Then we run 1000 simulations that randomly assign an identity to each persistence diagram. So the persistence diagram corresponding to the first sample from person 1 might be labeled as a sample from person 19 instead. In each simulation, we compute the sum of within group distances based on the new assignment. The p-value is then the fraction of the 1000 simulations whose total within group distances are less than or equal to the true sum. We will retain the null hypothesis $\mathcal{H}_0$ if the p-value is greater than or equal to 0.05 and reject it if the p-value is less than 0.05. The latter case indicates that there is a statistically significant difference between persistence diagrams from different individuals, and that it makes sense to perform identification and verification tasks by comparing persistence diagrams.

Indicators of model performance

We also want to measure the performance of our approach to personal identification and verification tasks. For the identification task, this is relatively straightforward. We run simulations, where a database of 80 images with four samples (i.e. $\ell = 4$ in Problem 2.1) from each individual is selected randomly in each simulation, and test using the remaining 20 images whether we can match them to the correct identity. We choose to have four samples per person in the database since this matches up with the reality (consider recording multiple sentences when setting up audio assistance on smartphones).

It is slightly harder to pick a measure for the verification task since there is much choice involved. For instance, we need to define a threshold distance as a measure of "similar enough" to the claimed identity. To show that our method can be used as a classification tool, we look at the receiver operating characteristic (ROC) curve, which is a 2D-plot of true positive rate (TPR) against the false positive rate (FPR) at different threshold values, where

$$\text{TPR} = \frac{\#\text{True Positive classifications}}{\#\text{True Positive classifications} + \#\text{False Negative classifications}},$$

and

$$\text{FPR} = \frac{\#\text{False Positive classifications}}{\#\text{False Positive classifications} + \#\text{True Negative classifications}}.$$

The most ideal ROC curve should look like a steep stair up from 0 to 1 and remains at TPR = 1 for FPR going from 0 to 1. This is an indication of a perfect binary classifier. That said, a classifier is considered good if the area under the ROC curve (AUC) is greater than 0.5, or equivalently, most of the ROC curve sits above the diagonal line.

Other measures for the behaviour of a verification model include recognition rates (RR), false acceptance rate (FAR) and false rejection rate (FRR), where

$$\text{RR} = \frac{\#\text{ True Positive classifications} + \#\text{ True Negative classifications}}{\#\text{ Total classifications}},$$

$$\text{FAR} = \frac{\#\text{ False Positive classifications}}{\#\text{ Total classifications}},$$

and

$$\text{FRR} = \frac{\#\text{ False Negative classifications}}{\#\text{ Total classifications}},$$

From the formula, it is not hard to see that a high RR and low FAR and FRR values are preferred. Since these are also the performance measures used in [30], we compute them for our approaches following the same procedure to compare the models.

2.2.3 Radial distance results

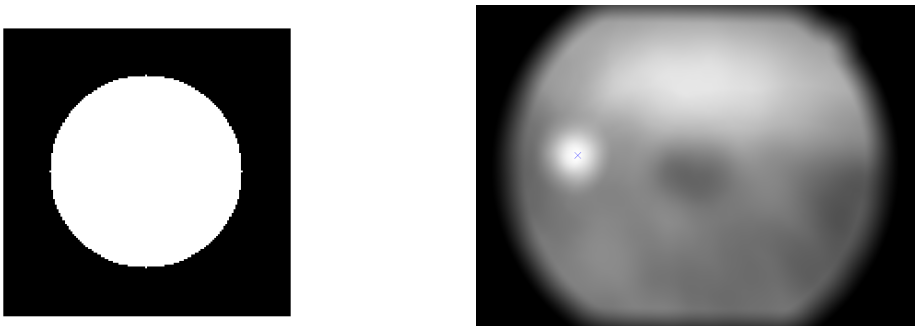
As motivated by [27], we consider different choices of filtration and examine how robust they are to the noise in the binary segmentation. We also note that the filtrations we choose are rotation and translation invariant. Hence, we do not need to align the images or consider potential effects of eye movements on the original retinal images. In this section, we restrict our attention to radial filtrations.

As mentioned above, we choose c to be the centre of an optical disc. We obtain the coordinate of the centre using image convolution. The idea is that we create a mask (also known as a kernel or a convolution matrix) and apply it to each pixel and their neighbouring pixels across the whole image so that the desired features can be easily identified. We now formalise the process. Consider an image as an $m \times n$ matrix whose entries are the corresponding pixels values. Denote by $f(x, y)$ the entry in row x, column y. Now let $w(x, y)$ be a $(2k+1) \times (2l+1)$ filter

matrix. Then the transformed matrix will be an $m \times n$ matrix whose (x,y)-entry is

$$g(x,y) = \sum_{dx=-k}^{k} \sum_{dy=-l}^{l} w(dx+k+1, dy+l+1) f(x-dx, y-dy).$$

Due to the circular shape of optical discs, we created a circular mask (of radius 150 pixels based on a rough measurement of the optical disc) so that the transformed image should highlight the disc with the centre taking the maximum value. As we can see from Figure 2.7, the optical discs are most visible in the red channel of original images. Therefore, we ran our mask on the red channels (see Figure 2.11) to obtain a list of coordinates for the centre. Note that the coordinates here are actually in the centre of a pixel and we include the relevant code in Appendix A.



(a) Kernel for picking out the bright spot (b) Transformed image

Figure 2.11: An illustration of a transformed image after running the kernel across the red channel. The blue cross on the transformed image corresponds to the centre of the optical disc.

After recording the centre coordinates, we can use the modified algorithm described in Section 1.4.2 to compute the radial extended persistent homology with respect to the radial filtrations defined using these centres. Note that this means we choose to use option (ii) in Section 1.4.2 to deal with the potential errors due to properties of radial functions. Figure 2.12 shows four resulting extended persistence diagrams. We can see that the two persistence diagrams for individual 4 are reasonably similar, but it seems hard to differentiate the first samples from individual 1 and 4. We also note that most of the points are extremely close to the diagonal line, and this is mainly due to the small specks that we can see in the segmentation.

As we have seen before, we can quantify the differences between two extended

persistence diagrams by computing their 1-Wasserstein distance with the codes provided in [29].

To avoid notational confusion, in what follows, we compute *within group distance* $\mathcal{D}$ by first taking all the sums of pairwise 1-Wasserstein distances of the extended persistence diagrams of 5 samples from the same individuals and then adding the 20 sums together. That is,

$$\mathcal{D} = \sum_{i=1}^{20} \sum_{j=2}^{5} \sum_{k=1}^{j-1} W_1((\mathbf{X})\mathbf{RPD}(R_{ij}), (\mathbf{X})\mathbf{RPD}(R_{ik})),$$

where $(\mathbf{X})\mathbf{RPD}(R_{il})$ is the (extended) persistence diagram computed for the l-th sample from individual i with respect to the filtration given in the context. Furthermore, the *cutoff distance* for the verification task refers to the 1-Wasserstein distance between (extended) persistence diagrams.

We found that the minimum difference between any two extended persistence diagrams of the five from the same individual is 5.243 and the average within group difference is 479.2246. To test whether the within group differences are actually significantly lower than between group differences, we perform a permutation test. We did 1000 simulations and obtained a p-value of 0.007 ± 0.004, which is strictly less than 0.05. Hence, there is a statistically significant evidence that we can differentiate the individuals by comparing their radial extended persistence diagrams.

We then proceed to perform identification tasks by comparing radial extended persistence diagrams of each individual to the ones in a database. In each simulation, the database consists of four randomly selected samples from each individual, and hence has 80 diagrams in total. We then test the accuracy of our identification module using the remaining 20 diagrams. We match the test image using closest two classification, and obtained an accuracy of 0.25 ± 0.05.

Figure 2.13 summarises the performance of the verification tasks using radial filtration. The recognition rate reaches a maximum of 95.98% when the cutoff distance is chosen to be 12. However, the cutoff value is too close to the minimum pairwise 1-Wasserstein distance for extended persistence diagrams from the same individual. This means that the system would reject the identity in most cases regardless of whether the person is genuine or an imposter, and the reason

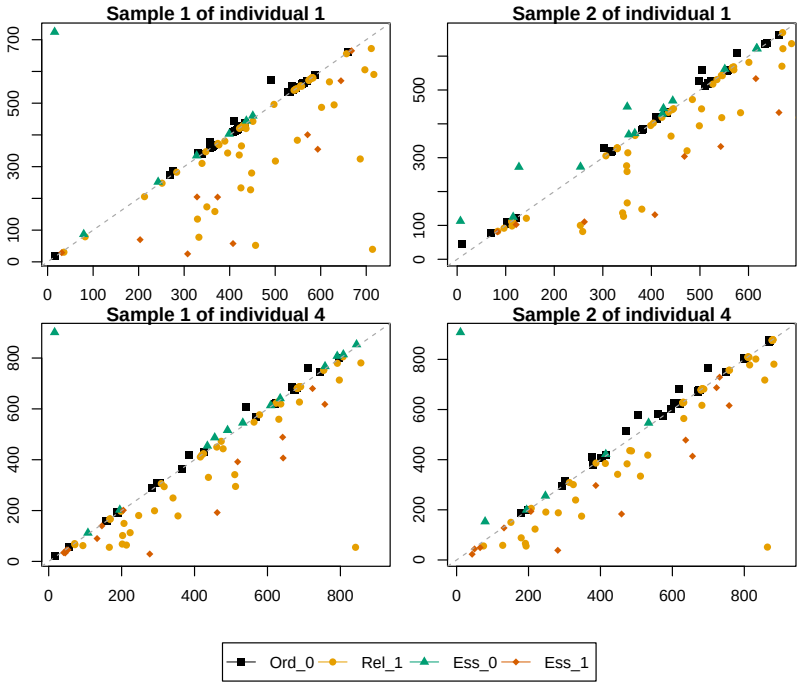


Figure 2.12: Extended persistence diagrams for two samples from two individuals.

this cutoff distance corresponds to a high accuracy is because most samples we are testing are imposters. So the high recognition rate does not imply that the system is secure. Moreover, we can see that the ROC curve is not ideal, and we suspect that this is due to the random specks in the segmentation and therefore led to many unwanted points in the extended persistence diagram. While theoretically they should only affect the Wasserstein distances by a small amount, the combination of a relatively large number of the specks may have caused some statistically significant effects. Therefore, we proceed to find ways that should dial down the influence of the noise. In particular, we consider other choices of filtrations.

2.2.4 SEDT and DMT results

SEDT

Let Ω be a subset of a metric space X with distance function d. Then the *Signed Euclidean function* $f(x)$ on X is defined to be

$$f(x) = \begin{cases} d(x, \partial\Omega), & x \in \Omega; \\ -d(x, \partial\Omega), & x \in \Omega^c, \end{cases}$$

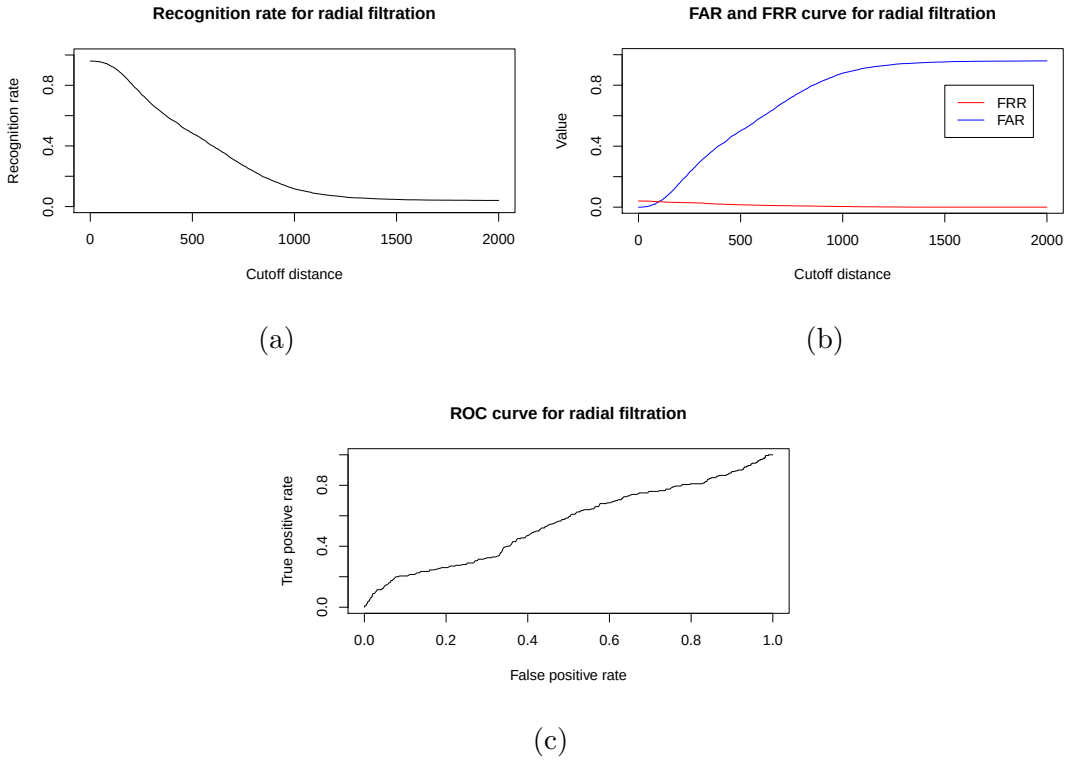


Figure 2.13: (a) RR curve, (b) FAR and FRR curves and (c) ROC curve using radial filtration.

where $\partial\Omega$ is the boundary of Ω, and $d(x, \partial\Omega) = \inf_{y\in\partial\Omega} d(x, y)$. In our case, we choose Ω to be the set of white pixels, i.e. the vessels, in the segmented images. We then define our filtration based on the value of f each point $x \in X$ takes and compute persistent homology accordingly. This process is what we call *signed Euclidean distance transform* (SEDT).

The reason we consider such a filtration is because they are less sensitive to noise than radial filtration. Consider a speck with small area. Then the maximum value a point inside the speck can take will be closer to zero than the lifetime of the speck when using radial filtration. Therefore, the homology class representing the speck should be closer to the diagonal line on the persistence diagram compared to when using radial filtration. Moreover, it takes into account both vessels and non-vessel areas, which should provide more information about the topology of given shapes.

Diamorse is an image analysis software that uses discrete Morse theory and can compute persistent homology of 2D and 3D grayscale images on top of performing SEDT of segmented images [9]. The top two images in Figure 2.16 is generated

using Diamorse and offer a visualisation of two examples after transformation. Pixels of the same value (colour) are in the same level set. Using Diamorse, we compute the persistent homology with respect to signed Euclidean distance filtration and Figure 2.14 illustrates four examples of the resulting persistence diagrams. Compared to Figure 2.12 for radial filtration, we can see that samples from the same individual seem to have more similar persistence diagrams than ones from different individuals.

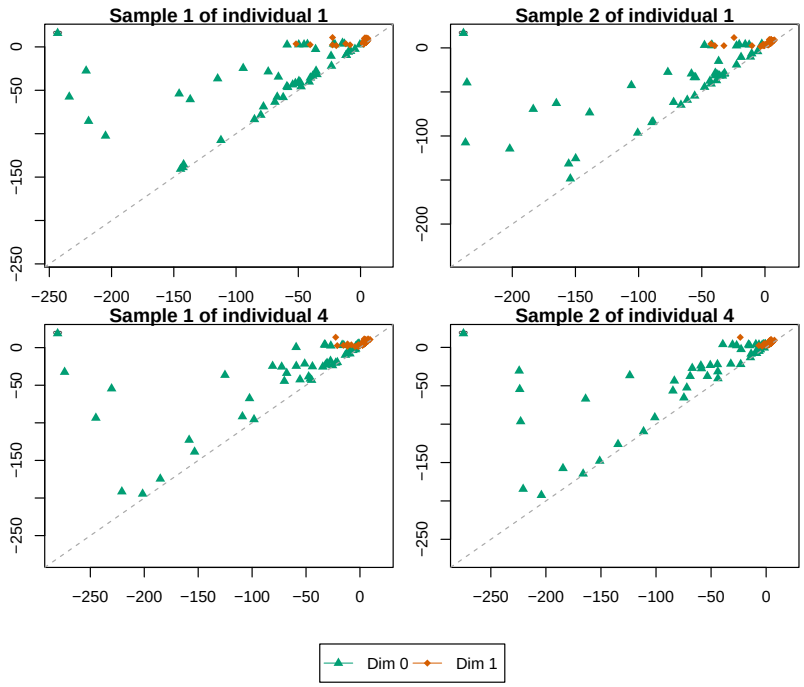
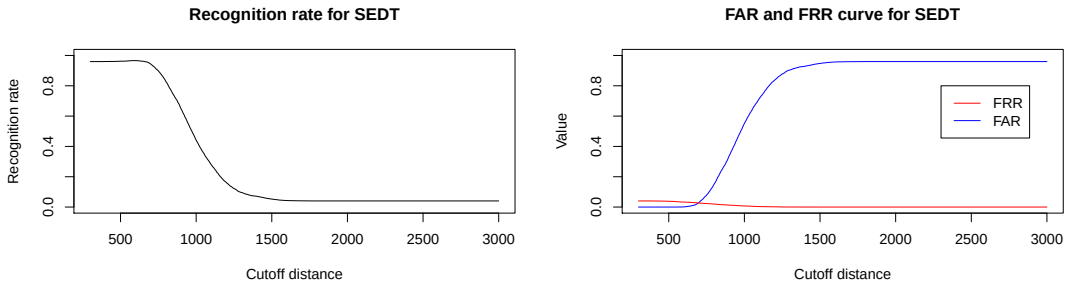


Figure 2.14: Persistence diagrams for two samples from two individuals using signed Euclidean distance filtration. The top left points correspond to the infinite-lifetime classes.

A permutation test verified this observation. To compare persistence pairings obtained from Diamorse, we modified the distance computation function in [17] to be able to compute distances between these persistence diagrams [‡] and saw that the within group distances are between 369.9748 and 695.8304. Like radial filtration, we run 1000 simulations that randomly assign each persistence pairing to an individual and obtain a p-value of 0.001, which is smaller than the p-value we obtained for radial filtration. Furthermore, a similar test for the identification task gives a 0.60 ± 0.05 accuracy. Finally, Figure 2.15 shows the performance of

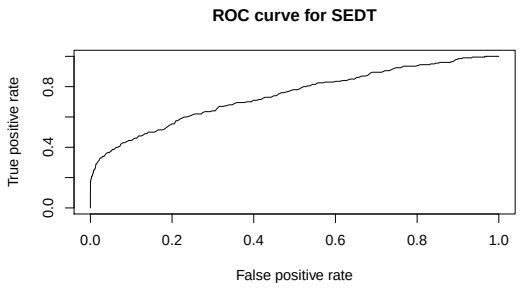
[‡]Codes in Appendix B.

performing the verification task. We can see that the ROC curve is much better than the one using radial filtration. The maximum recognition rate is 96.65% at cutoff distance 581.0811, a much more reasonable choice compared to the one for radial filtration.



(a)

(b)



(c)

Figure 2.15: (a) RR curve, (b) FAR and FRR curves and (c) ROC curve using signed Euclidean distance filtration.

DMT

While SEDT shows reasonable results for verification and identification tasks, it is not hard to see in Figure (a),(b) of 2.16 that noises in original segmentation are still present and lead to many points close to the diagonal line on the persistence diagrams (Figure 2.14). To further dial down the impact of the noise, we consider an alternative filtration, distance to measure transformation. Define a measure μ on a subset of a segmented image A given by the number of white pixels in the subset. Then for the centre x of each pixel, we define a function g_x for an input radius r:

$$g_x(r) = \mu\left(A \cap B(x, r)\right).$$

Now setting a threshold value k, we can assign to each pixel whose centre is x the value

$$\theta_k(x) = -\inf\{r | g_x(r) \geq k\}.$$

Translating and rescaling these values to between 0 and 255, we then obtain a grayscale image.

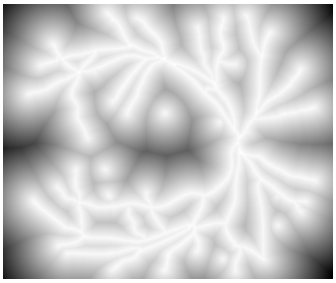
The upshot of such a transform is that by setting a relatively large threshold value, we are able to ignore small components while keeping desired structure of the original image. The code implementation of this method is inspired by [26] and uses image convolution to compute the measure at each point for various radii. We include relevant code in Appendix C. After some trial and error, we decide that $k = 200$ is a reasonable choice. The bottom two images of Figure 2.16 gives two examples of such a transform. Note that we can no longer see the specks as we do in SEDT of the same images, and at the same time, the vessel shapes are topologically similar to the ones in the top two images of Figure 2.16. The former point is also verified in the persistence pairings computed by Diamorse. As we can see in Figure 2.17, only a few points are close to the diagonal, and it is even more apparent that samples from the same individual are more similar than those from different people.

To further test our hypothesis, we again run a permutation test and obtained a p-value of 0.001 after 1000 trials. Accuracy for the identification task also improved to a rate of 0.75 ± 0.05. Finally, Figure 2.18 reveals the performance of the verification task, and better performance than previous filtration choices can be seen in both curves. Moreover, the highest recognition rate is 96.38% at a cutoff distance 161.9119 (the within group distance ranges from 70 to 233 as a reference).

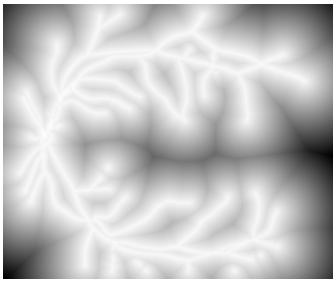
2.2.5 Threshold selection in practice

As discussed above, we can see that distance to measure filtration produces the best result. Here, we demonstrate how we can select the cutoff distance in reality. To that end, we partition our dataset of 100 retinal images into three disjoint sets: a database containing 60 images with 3 samples from each individual, a training set containing 20 images with one sample from each individual and the remaining 20 images as a testing set. This way, the testing process is completely blind to the training process.

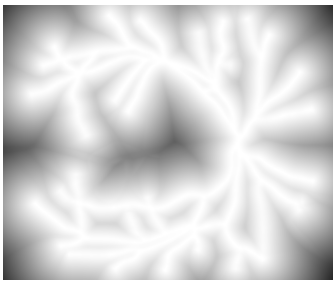
In each simulation, we compare the training set to the database by computing pairwise 1-Wasserstein distances to obtain a cutoff distance that gives the best



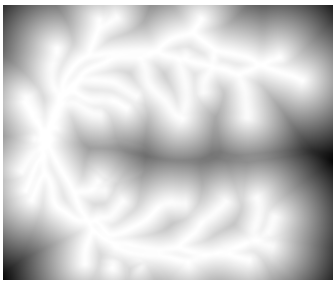
(a) IM000001_1



(b) IM000001_4



(c) IM000001_1



(d) IM000001_4

Figure 2.16: Side-by-side comparison of the two transforms: (a) (b) Signed Euclidean distance transform of two samples; (c) (d) Distance to measure transform of two samples.

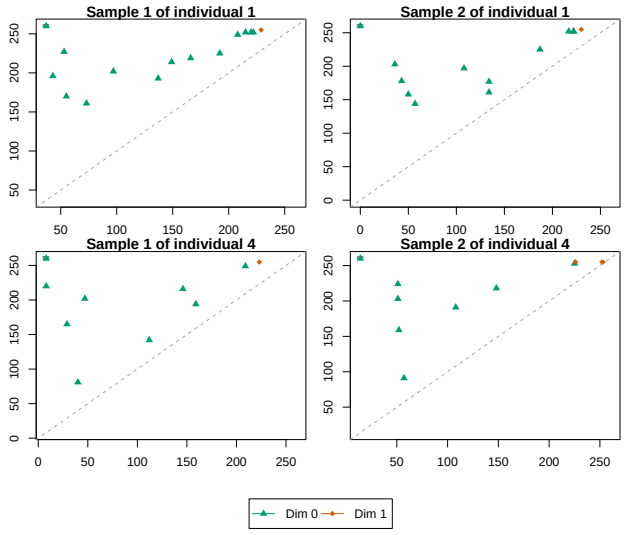
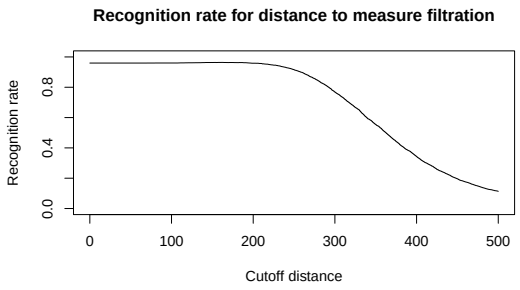
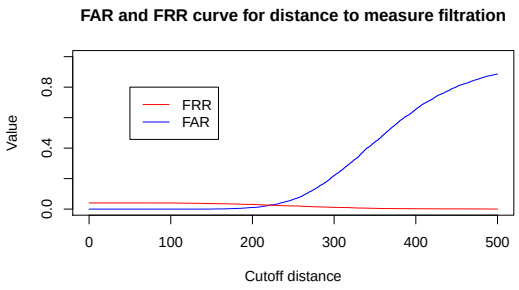


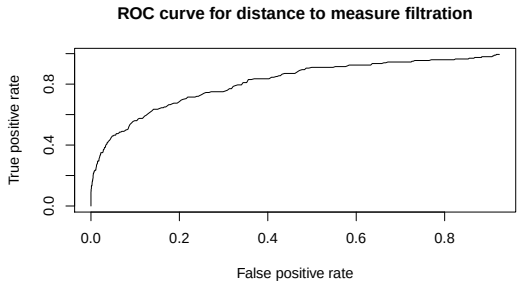
Figure 2.17: Persistence diagrams for two samples from two individuals using distance to measure filtration. The top left points correspond to the infinite-lifetime classes.



(a)



(b)



(c)

Figure 2.18: (a) RR curve, (b) FAR and FRR curves and (c) ROC curve using distance to measure filtration.

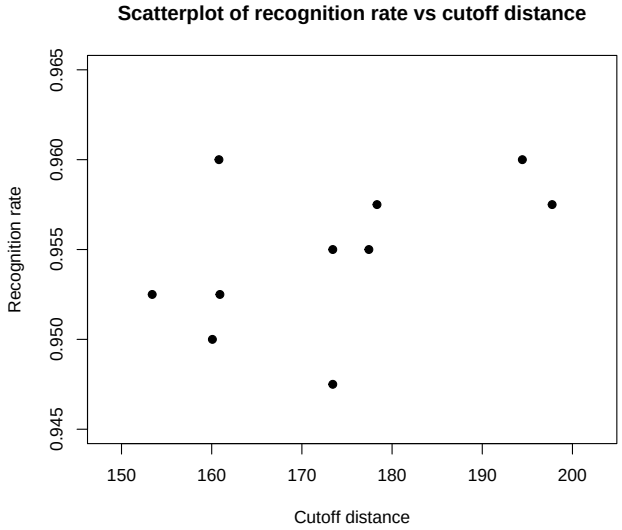


Figure 2.19: Scatter plot of recognition rate and cutoff distance compared from 10 simulations.

recognition rate. If multiple cutoff distances give the same recognition rate, we take the median of them. Then we use the obtained cutoff distance to verify the testing sets, that is we authenticate the individual if its average distance to the samples from the corresponding individual is less than the cutoff distance, and reject otherwise.

Figure 2.19 displays the accuracy of authentication with 10 simulations. We can see that the accuracy is relatively stable around 95.5% and the cutoff distances chosen are reasonably similar with different simulations (i.e. different partitions of the RIDB dataset). This justifies the validity of such a selection process. The code used to generate Figure 2.19 is included in Appendix D.

2.3 Conclusion

Table 2.1[§] compares the performance of our proposed methods and the one in [30] since the same retinal database was utilised. Note that while the title of [30] referenced personal identification tasks, they actually referred to verification tasks

[§]The paper [30] does not explicitly explain how they obtain their average elapsed time. The time we use here is the total time used to produce persistence pairings $\mathcal{P}$ for the test image, compute 1-Wasserstein distances between $\mathcal{P}$ and the persistence pairings for the images in the database, compare the average of the distances with the cutoff distance.

as defined in this thesis. We can see that while a vascular based method in [30] produces the highest recognition rate, using a distance to measure (dist2meas) filtration gives the best overall accuracy when we take into account measures such as FRR and FAR. Furthermore, methods involving persistent homology take significantly less time to carry out than methods in [30]. That said, there is still room for improvement. For instance, we could use a distance to measure function to filter out the specks in original segmentation and recompute the radial extended persistent homology on the cleaner segmentation using the same choice of centres. We could also try combining the results from different filtrations.

	Identification	Verification			
Method	Accuracy(%)	RR(%)	FRR	FAR	Avg time (sec)
Radial filtration	25	95.98	0.0	0.04	14.06
SEDT	60	96.65	0.0	0.04	26.28
Dist2meas	75	96.38	0.0	0.03	30.66
Vascular based [30]	-	100	0.52	0.16	127.9
Non-vascular based [30]	-	92.5	0.06	0.80	71.0

Table 2.1: A comparison table for the performance of methods used in this thesis and in previous work.

2.4 Future directions

Another heated research topic revolving retinal fundus images is disease detection. Early signs of cardiovascular diseases are visible in non-vascular areas of retinal fundus images. Current methods for automatic detection of such signs include extracting features such as the distance of optic disks to some minutiae points. We could explore the use of persistent homology to analyse the irregular shapes of non-vascular regions.

The proposed methods can also be adapted to other biometrics such as fingerprints and facial geometry, etc.

Chapter 3

Computing XRPH for manifolds with boundary

> "It's easy to get the feeling that you know the language just because when you order a beer they don't bring you oysters," [Paul said.]
>
> *My life in France*, Julia Child

In this chapter, we will show that we can recover the extended persistence homology of a manifold with boundary with respect to radial filtration from the persistence homology of the boundary curves. This will further justify our modification of the algorithm used to compute XRPH of the segmented retinal image embedded in $\mathbb{R}^2$ in Section 1.4.

3.1 Preliminary

To prove the aforementioned relation, we will use Morse theory, which is the study of topology of a manifold by analysing differentiable functions on the manifold, in both smooth and piecewise linear scenarios. The former is more for the theoretical development while the latter is directly related to the numerical algorithms such as the one used in Section 1.4. Note that piecewise linear case is not the same as discrete Morse theory, but rather functions that are defined on vertices and interpolated linearly on the edges (consider the radial function defined on the extracted piecewise linear boundary). Here, we will follow the setup in [29] and give necessary background information to keep the thesis self-contained.

We will eventually restrict our attention to the radial function ρ_c for a fixed centre $c \in \mathbb{R}^n$. This is a Morse function in the scenarios we considered in Section 1.4. We will give a formal definition of Morse functions, but we need to first define what it means for a point in a manifold to be a critical point. As noted in [29], the standard way to define critical points is in terms of taking derivatives with respect to some charts. Here we adopt an equivalent definition justified using Morse Lemma.

Definition 3.1 (Critical points of smooth functions)**.** Let M be a smooth n-manifold without boundary and $f : M \to \mathbb{R}$ is a smooth function. Then $p \in M$ is a *regular point* of f if there is a chart (U, ϕ) such that $\phi(p) = 0$ and for $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ in a neighbourhood of 0, we have

$$f \circ \phi^{-1}(x) = f(p) + x_n.$$

We call $p \in M$ a *non-degenerate critical point of f with Morse index k* if there is a chart (U, ϕ) such that $\phi(p) = 0$ and $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ in a neighbourhood of 0, we have

$$f \circ \phi^{-1}(x) = f(p) - x_1^2 - x_2^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_n^2.$$

A similar definition for piecewise linear functions can be made replacing taking squares with taking absolute values.

Definition 3.2 (Critical points of piecewise linear functions)**.** Let M be a piecewise linear n-manifold without boundary and $f : M \to \mathbb{R}$ is a piecewise linear function. Then $p \in M$ is a *regular point* of f if there is a chart (U, ϕ) such that $\phi(p) = 0$ and for $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ in a neighbourhood of 0, we have

$$f \circ \phi^{-1}(x) = f(p) + x_n.$$

We call $p \in M$ a *non-degenerate critical point of f with Morse index k* if there is a chart (U, ϕ) such that $\phi(p) = 0$ and $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ in a neighbourhood of 0, we have

$$f \circ \phi^{-1}(x) = f(p) - |x_1| - |x_2| - \cdots - |x_k| + |x_{k+1}| + \cdots + |x_n|.$$

Since we will be considering manifolds with boundary, we define critical points on the boundary as well using a similar formulation as above. Note that if the critical point is in the interior of the manifold, then we can treat it the same way as critical points in manifolds without boundary. We will also differentiate between (+)- and (−)-critical points (compare with Definition 1.54).

Definition 3.3 (Boundary critical points of smooth functions)**.** Let $(M, \partial M)$ be a smooth n-manifold with boundary and $f : M \to \mathbb{R}$ a smooth function. Then $p \in \partial M$ is a *non-degenerate critical point of f with index* (k, η) if there exists a chart (U, ϕ) such that $\phi(p) = 0$ and for $x = (x_1, \cdots, x_n)$ in a neigbourhood of 0 in the closed half-space $\{(x_1, \cdots, x_n) \in \mathbb{R}^n | x_1 \geq 0\}$ of $\mathbb{R}^n$, we have

$$f \circ \phi^{-1}(x) = f(p) + \eta x_1 - x_2^2 - \cdots - x_{k+1}^2 + x_{k+2}^2 + \cdots x_n^2,$$

where $\eta \in \{-1, 1\}$. We call p a $(+)$-*critical point* if $\eta = 1$ and $(-)$-*critical point* if $\eta = -1$.

Similarly, the analogous definition in the piecewise linear case is given as expected:

Definition 3.4 (Boundary critical points of piecewise linear functions)**.** Let $(M, \partial M)$ be a piecewise linear n-manifold with boundary and $f : M \to \mathbb{R}$ a piecewise function. Then $p \in \partial M$ is a *non-degenerate critical point of f with index* (k, η) if there exists a chart (U, ϕ) such that $\phi(p) = 0$ and for $x = (x_1, \cdots, x_n)$ in a neigbourhood of 0 in the closed half-space $\{(x_1, \cdots, x_n) \in \mathbb{R}^n | x_1 \geq 0\}$ of $\mathbb{R}^n$, we have

$$f \circ \phi^{-1}(x) = f(p) + \eta x_1 - |x_2| - \cdots - |x_{k+1}| + |x_{k+2}| + \cdots |x_n|,$$

where $\eta \in \{-1, 1\}$. We call p a $(+)$-*critical point* if $\eta = 1$ and $(-)$-*critical point* if $\eta = -1$.

We are now ready to formally define Morse functions. It is easy to see from the definition that the radial function defined on a finite simplicial complex for some fixed centre does satisfy the assumptions of Morse functions.

Definition 3.5 (Morse functions)**.** Let $(M, \partial M)$ be a smooth (or piecewise linear) manifold. Then a function $f : M \to \mathbb{R}$ is a *Morse function* if

(i) f is smooth (or piecewise linear);

(ii) All of the critical points of $f|_{\text{int}(M)}$ and $f|_{\partial M}$ combined are non-degenerate;

(iii) The combined number of critical points is finite and they all take distinct values.

3.2 Relating XRPH of a manifold to that of its boundary

From this point on, the only Morse functions we will be considering are radial functions on a compact n-manifold $(M, \partial M)$ with boundary embedded in $\mathbb{R}^n$. That is, fix $c \in \mathbb{R}^n$, and consider $\rho_c : \mathbb{R}^n \to \mathbb{R}$ by $\rho_c(x) = \|x - c\|_2$.

For ease of notation, for $S \subset \mathbb{R}^n$, we write $\rho_c^S \coloneqq \rho_c|_S$, the restriction of ρ_c to S. Then we use $\mathrm{Crit}(\rho_c^M, k)$ to denote the set of critical points of $\rho_c^{\mathrm{int}(M)}$ with index k. Furthermore, denote by $\mathrm{Crit}(\rho_c^M, (k, \eta))$ the set of critical points of $\rho_c^{\partial M}$ with index (k, η). For a non-degenerate critical point $x \in \partial M$, we will call η its *sign*, and denote by $\mathrm{sgn}(\rho_c^M, x)$. Finally, if

$$\mathrm{XRPH}(M, \rho_c) = \mathrm{Ord}_*(M, \rho_c) \oplus \mathrm{Rel}_*(M, \rho_c) \oplus \mathrm{Ess}_*(M, \rho_c)$$

is the radial extended persistence diagram of M, we let

$$\mathfrak{b}_k^{\mathrm{ord}}(M, \rho_c) \coloneqq \mathfrak{b}\left(\mathrm{Ord}_k(M, \rho_c) \oplus \mathrm{Ess}_k(M, \rho_c)\right),$$

and

$$\mathfrak{b}_k^{\mathrm{rel}}(M, \rho_c) \coloneqq \mathfrak{b}\left(\mathrm{Rel}_k(M, \rho_c)\right).$$

Similarly, we define the set of death parameters based on the class of the endpoint of an interval is in. That is,

$$\mathfrak{d}_k^{\mathrm{ord}}(M, \rho_c) \coloneqq \mathfrak{d}\left(\mathrm{Ord}_k(M, \rho_c)\right),$$

and

$$\mathfrak{d}_k^{\mathrm{rel}}(M, \rho_c) \coloneqq \mathfrak{d}\left(\mathrm{Rel}_k(M, \rho_c) \oplus \mathrm{Ess}_k(M, \rho_c)\right).$$

The notations are also consistent with ones in [29].

Theorem 3.6 is key to proving our main results. Its proof can be found in [29] (Corollary 4.14).

Theorem 3.6. *Let $(M, \partial M)$ be an n-manifold with boundary. Consider the radial function $\rho_c : \mathbb{R}^n \to \mathbb{R}$ with some fixed centre c such that ρ_c^M is a Morse function. Then for all $k \geq 0$, we have*

$$\mathfrak{b}_k^{ord}(M, \rho_c) \cup \mathfrak{d}_{k-1}^{ord}(M, \rho_c) = \{(\rho_c(p), Ord) | p \in Crit(\rho_c^M, k) \cup Crit(\rho_c^M, (k, +1))\},$$

and

$$\mathfrak{b}_k^{rel}(M, \rho_c) \cup \mathfrak{d}_{k-1}^{rel}(M, \rho_c) = \{(\rho_c(p), Rel) | p \in Crit(\rho_c^M, n-k) \cup Crit(\rho_c^M, (n-k-1, -1))\},$$

Remark 3.7. Note that Theorem 3.6 also works for general Morse functions, as shown in [29]. That is, we can replace ρ_c in the statement with any Morse function f.

We now use Theorem 3.6 to further compare the birth and death parameters of M in different classes with those of the boundary ∂M of M. Before that, we will include a useful theorem from linear algebra that will be applied multiple times to obtain relations of Betti numbers of different dimensions for various spaces. The *k-th Betti number* represents the rank of the k-th homology group. For notational simplicity, in what follows, we will let $\beta_k(X) \coloneqq \text{rank}(H_k(X))$ for a topological space X and $\beta_k(X, A) \coloneqq \text{rank}(H_k(X, A))$ for a pair of topological spaces (X, A) with $A \subseteq X$.

Theorem 3.8. *Consider a finite exact sequence of vector spaces:*

$$0 \to A_1 \to A_2 \to \cdots \to A_n \to 0.$$

Then we have

$$\sum_{i=1}^{n} (-1)^i rank(A_i) = 0.$$

Proposition 3.9. *Let $(M, \partial M)$ be a n-manifold with boundary embedded in $\mathbb{R}^n$. Consider the radial function $\rho_c : \mathbb{R}^n \to \mathbb{R}$ for the centre $c \in \mathbb{R}^n$ such that ρ_c^M is a Morse function. Then for each critical value, there exists a unique critical point, either in the interior or on the boundary of M, that achieves that value. Then if $c \notin M$, for all $k \geq 0$, we have*

$$\begin{aligned}
\mathfrak{b}_k^{ord}(M, \rho_c) &= \{(\rho_c^{\partial M}(p), Ord) \in \mathfrak{b}_k^{ord}(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = +1\};\\
\mathfrak{b}_k^{rel}(M, \rho_c) &= \{(\rho_c^{\partial M}(p), Rel) \in \mathfrak{b}_k^{rel}(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = -1\};\\
\mathfrak{d}_k^{ord}(M, \rho_c) &= \{(\rho_c^{\partial M}(p), Ord) \in \mathfrak{d}_k^{ord}(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = +1\};\\
\mathfrak{d}_k^{rel}(M, \rho_c) &= \{(\rho_c^{\partial M}(p), Rel) \in \mathfrak{d}_k^{rel}(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = -1\}.
\end{aligned}$$

If $c \in M$, let $r = \min_{p \in \partial M} \rho_c(p)$. Then all the birth and death parameters of M and ∂M satisfy the same relations as above for $k \geq 0$ except the following:

$$\begin{aligned}
\mathfrak{b}_0^{ord}(M, \rho_c) &= \{(\rho_c^{\partial M}(p), Ord) \in \mathfrak{b}_0^{ord}(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = +1\} \cup \{(0, Ord)\};\\
\mathfrak{d}_{n-1}^{rel}(M, \rho_c) &= \{(\rho_c^{\partial M}(p), Rel) \in \mathfrak{d}_{n-1}^{rel}(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = -1\} \backslash \{(r, Rel)\};\\
\mathfrak{b}_n^{rel}(M, \rho_c) &= \{(r, Rel)\};\\
\mathfrak{d}_n^{rel}(M, \rho_c) &= \{(0, Rel)\}.
\end{aligned}$$

Proof. The proof will use Mayer-Vietoris sequence (Theorem 1.17) to relate the homology of sublevel sets of ∂M with that of sublevel sets of M. To that end, choose $R > 0$ such that $M \subset B(c, R)$ where $B(c, R)$ is the open ball of radius R centred at $c \in \mathbb{R}^n$. We can extend the domain of ρ_c to $\overline{B(c, R)}$. Let $A = \overline{B(c, R)} \backslash \text{int}(M)$. Since ρ_c is a Morse function on M, there exists some $\epsilon > 0$ such that all critical values of ρ_c^M are at least ϵ apart and that the largest critical value is strictly less than $R - \epsilon$. Then we have $[\inf(\rho_c(M)), \sup(\rho_c(M))] \subseteq [0, R)$ since ρ_c takes non-negative values.

For $s > 0$, we consider the sublevel sets ρ_c restricted to three subsets $M, \partial M$ and L of $\mathbb{R}^n$, call them M_s, $(\partial M)_s$ and L_s respectively. Then we have $(\partial M)_s = M_s \cap L_s$ and $M_s \cup L_s = \rho_c^{-1}(-\infty, s] \cap \overline{B(c, R)} = \overline{B(c, s)}$. By Theorem 1.17, we have the LES:

$$\begin{aligned} \cdots \longrightarrow H_{k+1}(M_s \cup L_s) \longrightarrow H_k((\partial M)_s) \longrightarrow H_k(M_s) \oplus H_k(L_s) \longrightarrow \\ H_k(M_s \cup L_s) \longrightarrow \cdots \longrightarrow H_0(M_s \cup L_s) \longrightarrow 0. \end{aligned} \tag{3.1}$$

Now note that for $k > 0$, we have $H_{k+1}(M_s \cup L_s) = H_k(M_s \cup L_s) = 0$. Hence, we have for $k > 0$,

$$H_k((\partial M)_s) \cong H_k(M_s) \oplus H_k(L_s),$$

which implies

$$\beta_k((\partial M)_s) = \beta_k(M_s) + \beta_k(L_s), \tag{3.2}$$

by taking the ranks. Now for $k = 0$, we have $H_0(M_s \cup L_s) = \mathbb{Z}_2$ since $s > 0$. Hence, using the first isomorphism theorem and the final parts of LES (3.1), we have that

$$\left(H_0(M_s) \oplus H_0(L_s)\right) / H_0((\partial M)_s) \cong H_0(M_s \cup L_s),$$

which implies

$$\beta_0(M_s) + \beta_0(L_s) = \beta_n(M_s \cup L_s) + \beta_0((\partial M)_s) = \beta_0((\partial M)_s) + 1. \tag{3.3}$$

A critical observation here is that ρ_c defined over all of $\mathbb{R}^n$ has only one critical point at c. So if $c \notin M$, there will be no critical points in the interior of M. In particular, c cannot be a critical point of ρ_c^M, and thus, 0 cannot be a critical value. However, if $c \in M$, then c will be the only critical point with critical value 0 and gives birth to a 0-dimensional homology class. We consider the two cases separately.

First consider the case where the centre $c \notin M$. Suppose $(t, \text{Ord}) \in \mathfrak{b}_k^{\text{ord}}(M, \rho_c)$. That is $\beta_k(M_{t+\epsilon}) - \beta_k(M_{t-\epsilon}) = 1$ since we chose our ϵ so that there exists exactly

one critical point with critical value between $t-\epsilon$ and $t+\epsilon$. In fact, that critical value would be t. By Theorem 3.6, we have that $p = \rho_c^{-1}(t)$ is the unique critical point with index k (since we choose $\epsilon > 0$ so that t is the only critical value between $[t-\epsilon, t+\epsilon]$) and that $\operatorname{sgn}(\rho_c^M, p) = +1$. Then if we consider the restriction of ρ_c to L, we would have $\operatorname{sgn}(\rho_c^L, p) = -1$ by construction. Then again by Theorem 3.6, we know that $(t, \mathrm{Ord}) \notin \mathfrak{b}_k^{\mathrm{ord}}(L, \rho_c)$ and $(t, \mathrm{Ord}) \notin \mathfrak{b}_k^{\mathrm{ord}}(L, \rho_c)$. Hence, $\beta_k(L_{t+\epsilon}) = \beta_k(L_{t-\epsilon})$. For $k > 0$, we can then use Relation (3.2) and obtain

$$\begin{aligned}\beta_k((\partial M)_{t+\epsilon}) - \beta_k((\partial M)_{t-\epsilon}) \\ &= (\beta_k(M_{t+\epsilon}) + \beta_k(L_{t+\epsilon})) - (\beta_k(M_{t-\epsilon}) + \beta_k(L_{t-\epsilon})) \\ &= 1. \end{aligned} \tag{3.4}$$

For $k = 0$, we would use Relation (3.3) and obtain

$$\begin{aligned}\beta_0((\partial M)_{t+\epsilon}) - \beta_0((\partial M)_{t-\epsilon}) \\ &= (\beta_0(M_{t+\epsilon}) + \beta_0(L_{t+\epsilon}) - 1) - (\beta_0(M_{t-\epsilon}) + \beta_0(L_{t-\epsilon}) - 1) \\ &= 1. \end{aligned} \tag{3.5}$$

Hence, in either case, $(t, \mathrm{Ord}) \in \mathfrak{b}_k^{\mathrm{ord}}(\partial M, \rho_c^{\partial M})$.

Conversely, if $(t, \mathrm{Ord}) \in \mathfrak{b}_k^{\mathrm{ord}}(\partial M, \rho_c^{\partial M})$ with $\operatorname{sgn}(\rho_c^M, \rho_c^{-1}(t)) = +1$. Then this implies that $\beta_k((\partial M)_{t+\epsilon}) - \beta_k((\partial M)_{t-\epsilon}) = 1$ and that $\operatorname{sgn}(\rho_c^L, \rho_c^{-1}(t)) = -1$. Again, by Theorem 3.6, we have that $\beta_k(L_{t+\epsilon}) = \beta_k(L_{t-\epsilon})$. So using Relation (3.2) for $k > 0$, we have that

$$\begin{aligned}\beta_k(M_{t+\epsilon}) - \beta_k(M_{t-\epsilon}) \\ &= (\beta_k((\partial M)_{t+\epsilon}) - \beta_k(L_{t+\epsilon})) - (\beta_k((\partial M)_{t-\epsilon}) - \beta_k(L_{t-\epsilon})) \\ &= 1. \end{aligned} \tag{3.6}$$

Similarly, for $k = 0$, we use Relation (3.3) and get

$$\begin{aligned}\beta_k(M_{t+\epsilon}) - \beta_k(M_{t-\epsilon}) \\ &= (\beta_k((\partial M)_{t+\epsilon}) - \beta_k(L_{t+\epsilon}) + 1) - (\beta_k((\partial M)_{t-\epsilon}) - \beta_k(L_{t-\epsilon}) + 1) \\ &= 1. \end{aligned} \tag{3.7}$$

Again, in either case, we have shown that $(t, \mathrm{Ord}) \in \mathfrak{b}_k^{\mathrm{ord}}(M, \rho_c)$.

The statement about death parameters in the ordinary class can be shown following an analogous argument except now $(t, \mathrm{Ord}) \in \mathfrak{d}_k^{\mathrm{ord}}(S, \rho_c)$ if and only if $\beta_k(S_{t+\epsilon}) - \beta_k(S_{t-\epsilon}) = -1$ for $S = \partial M$ or M. We can then use Theorem 3.6 along with Relations (3.2) and (3.3) to reach the desired conclusion.

Now assume $c \in M$. Then we have $M_{-\epsilon}$, $(\partial M)_m$, and L_m are all empty sets and have homology groups zero for all dimensions for $m = \pm\epsilon$. But $M_\epsilon = M_\epsilon \cup L_\epsilon$ are both a single connected component since 0 is the only critical value in $[-\epsilon, \epsilon]$ and has homology group $\mathbb{Z}_2$ in dimension 0 and homology group 0 in all other dimensions. Hence, Relation (3.2) still holds for $s \geq -\epsilon$ and thus, the above arguments hold for all $k > 0$. However, in the case where $k = 0$, we have

$$\beta_0(M_\epsilon) - \beta_0(M_{-\epsilon}) = \beta_0(M_\epsilon \cup L_\epsilon) = 1,$$

but

$$\beta_0((\partial M)_\epsilon) - \beta_0((\partial M)_{-\epsilon}) = 0.$$

Hence, we have $(0, \mathrm{Ord}) \in \mathfrak{b}_k^{\mathrm{ord}}(M, \rho_c)$, but $(0, \mathrm{Ord}) \notin \mathfrak{b}_k^{\mathrm{ord}}(\partial M, \rho_c^{\partial M})$. It follows that

$$\mathfrak{b}_k^{\mathrm{ord}}(M, \rho_c) = \{(\rho_c^{\partial M}(p), \mathrm{Ord}) \in \mathfrak{b}_k^{\mathrm{ord}}(\partial M, \rho_c^{\partial M}) : \mathrm{sgn}(\rho_c^M, p) = +1\} \cup \{(0, \mathrm{Ord})\}$$

instead.

The proof for $\mathfrak{d}_k^{\mathrm{ord}}(M, \rho_c)$ for the case $c \in M$ is the same as the case when $c \notin M$. So the death parameters in the Ordinary class for dimension 0 remain the same.

We now consider the relative class. We use the relative version of Mayer-Vietoris sequence:

$$\begin{aligned}\cdots \longrightarrow H_{k+1}(M \cup L, M^s \cup L^s) \longrightarrow H_k((\partial M), (\partial M)^s) \longrightarrow H_k(M, M^s) \oplus H_k(L, L^s) \\ \longrightarrow H_k(M \cup L, M^s \cup L^s) \longrightarrow \cdots \longrightarrow H_0(M \cup L, M^s \cup L^s) \longrightarrow 0.\end{aligned} \tag{3.8}$$

Now for $k \geq 0$, $k \neq n$ and $0 \leq s \leq R$, we have $H_k(M \cup L, M^s \cup L^s) = 0$. For $k = n$, we have $H_n(M \cup L, M^s \cup L^s) = \mathbb{Z}_2$ if $0 < s \leq R$ and $H_n(M \cup L, M^0 \cup L^0) = 0$. From LES (3.8), we have that for $k \neq n - 1$ or n,

$$H_k((\partial M), (\partial M)^s) \cong H_k(M, M^s) \oplus H_k(L, L^s),$$

which implies that

$$\beta_k((\partial M), (\partial M)^s) = \beta_k(M, M^s) + \beta_k(L, L^s). \tag{3.9}$$

Suppose $(t, \mathrm{Rel}) \in \mathfrak{b}_k^{\mathrm{rel}}(M, \rho_c^M)$. Then we have $\beta_k(M, M^{t-\epsilon}) - \beta_k(M, M^{t+\epsilon}) = 1$. By Theorem 3.6, we know $\mathrm{sgn}(\rho_c^M, \rho_c^{-1}(t)) = -1$ and therefore, $\mathrm{sgn}(\rho_c^L, \rho_c^{-1}(t)) = +1$. Hence, we have $\beta_k(L, L^{t-\epsilon}) = \beta_k(L, L^{t+\epsilon})$. This implies that

$$\beta_k((\partial M), (\partial M)^{t-\epsilon}) - \beta_k((\partial M), (\partial M)^{t+\epsilon}) = 1$$

from Relation (3.9), which means

$$(t, \text{Rel}) \in \mathfrak{b}_k^{\text{rel}}(\partial M, \rho_c^{\partial M}).$$

Conversely, if $(t, \text{Rel}) \in \mathfrak{b}_k^{\text{rel}}(\partial M, \rho_c^{\partial M})$ such that $\text{sgn}(\rho_c^M, \rho_c^{-1}(t)) = -1$. Then it follows that $\beta_k(\partial M, \partial M^{t-\epsilon}) - \beta_k(\partial M, \partial M^{t+\epsilon}) = 1$ and $\text{sgn}(\rho_c^L, \rho_c^{-1}(t)) = +1$. Then Theorem 3.6 tells us that $\beta_k(L, L^{t-\epsilon}) = \beta_k(L, L^{t+\epsilon})$. Combining with Relation (3.9), we have

$$\beta_k(M, M^{t-\epsilon}) - \beta_k(M, M^{t+\epsilon}) = 1,$$

and hence we have shown that

$$(t, \text{Rel}) \in \mathfrak{b}_k^{\text{rel}}(M, \rho_c)$$

as required.

Note that since LES (3.8) holds for $s = 0$, we do not need to consider the two cases separately. Again, the death parameters for the relative class follow from an analogous argument.

However, we do need to consider the position of c for the case $k = n - 1$ and $k = n$. Since ∂M is an $(n-1)$-manifold, we have

$$H_n(\partial M, (\partial M)^s) = 0,$$

for all $0 \leq s \leq R$. So the relevant part of LES (3.8) is

$$\begin{aligned} 0 \longrightarrow H_n(M, M^s) \oplus H_n(L, L^s) \hookrightarrow H_n(M \cup L, M^s \cup L^s) \longrightarrow H_{n-1}((\partial M), (\partial M)^s) \\ \twoheadrightarrow H_{n-1}(M, M^s) \oplus H_{n-1}(L, L^s) \longrightarrow 0. \end{aligned} \tag{3.10}$$

Let $r = \min_{p \in \partial M} \rho_c(p)$. We have the following observation: if $c \in M$, then we have that

$$H_n(M, M^s) = \begin{cases} 0, & r < s \leq R \\ \mathbb{Z}_2, & 0 < s \leq r \\ 0, & s = 0 \end{cases}, \tag{3.11}$$

and

$$H_n(L, L^s) = 0, \tag{3.12}$$

for all $0 \leq s \leq R$. Hence, there are only two critical points in dimension n: a birth of relative class at r and a death of relative class at 0. Now when $0 \leq s \leq r$, note that $(\partial M)^s = \partial M$ by the definition of r. Then the (3.10) tells us that

$$H_{n-1}(M, M^s) = H_{n-1}(L, L^s) = 0.$$

In particular, this satisfies

$$\beta_{n-1}(\partial M, (\partial M)^s) = \beta_{n-1}(M, M^s) + \beta_{n-1}(L, L^s). \tag{3.13}$$

When $r < s \leq R$, on the other hand, we have $H_n(M, M^s) \oplus H_n(L, L^s) = 0$, and hence

$$\beta_{n-1}(\partial M, (\partial M)^s) = \beta_{n-1}(M, M^s) + \beta_{n-1}(L, L^s) + 1. \tag{3.14}$$

This means that for critical values $t \in (0, r)$ or $t \in (r, R)$, we can use the same argument as above to show that

$$\begin{aligned}
&\mathfrak{b}^{\mathrm{rel}}_{n-1}(M, \rho_c) \backslash \{(r, \mathrm{Rel})\} \\
&\qquad = \{(\rho_c^{\partial M}(p), \mathrm{Rel}) \in \mathfrak{b}^{\mathrm{rel}}_{n-1}(\partial M, \rho_c^{\partial M}) : \mathrm{sgn}(\rho_c^M, p) = -1\} \backslash \{(r, \mathrm{Rel})\}; \\
&\mathfrak{d}^{\mathrm{rel}}_{n-1}(M, \rho_c) \backslash \{(r, \mathrm{Rel})\} \\
&\qquad = \{(\rho_c^{\partial M}(p), \mathrm{Rel}) \in \mathfrak{d}^{\mathrm{rel}}_{n-1}(\partial M, \rho_c^{\partial M}) : \mathrm{sgn}(\rho_c^M, p) = -1\} \backslash \{(r, \mathrm{Rel})\}.
\end{aligned}$$

Observe that (r, Rel) correspond to a death of relative class in ∂M of dimension $n-1$, but it is not a critical point for M or L since ρ_c^M is Morse and thus $H_{n-1}(M, M^s) = H_{n-1}(L, L^s) = 0$ for $s < r + \epsilon$. This agrees with the computation below:

$$\begin{aligned}
&\beta_{n-1}(\partial M, (\partial M)^{r-\epsilon}) - \beta_{n-1}(\partial M, (\partial M)^{r+\epsilon}) \\
=&(\beta_{n-1}(M, M^{r-\epsilon}) + \beta_{n-1}(L, L^{r-\epsilon})) - (\beta_{n-1}(M, M^{r+\epsilon}) + \beta_{n-1}(L, L^{r+\epsilon}) + 1) \\
=&-1.
\end{aligned}$$

Now since $c \in M$, the point $\rho_c^M(r)$ is a $(-)$-critical point, but it should not be included in $\mathfrak{d}^{\mathrm{rel}}_{n-1}(M, \rho_c)$. Hence, we have the relation in the statement.

If $c \notin M$, then $c \in L$. So we interchange the role of M and L in the homology groups (3.11) and (3.12). In particular, we have that $H_n(M, M^s) = 0$ for all $0 \leq s \leq R$. Hence, no birth or death of relative class occur in dimension n in M, i.e. $\mathfrak{b}^{\mathrm{rel}}_n(M, \rho_c^{\partial M}) = \mathfrak{d}^{\mathrm{rel}}_n(M, \rho_c^{\partial M}) = \emptyset$. Since $H_n(\partial M, (\partial M)^s) = 0$, for all $0 \leq s \leq R$, there is no birth or death on ∂M of relative class of dimension n. So we have $\mathfrak{b}^{\mathrm{rel}}_n(\partial M, \rho_c^{\partial M}) = \mathfrak{d}^{\mathrm{rel}}_n(\partial M, \rho_c^{\partial M}) = \emptyset$. For dimension $n-1$, relations (3.13)

and (3.14) still hold and hence we have the same relation for $\mathfrak{b}^{\mathrm{rel}}_{n-1}(M,\rho_c)$ and $\mathfrak{d}^{\mathrm{rel}}_{n-1}(M,\rho_c)$ as the case $c \in M$ when the critical value is not (r, Rel). However, since not $c \notin M$, the critical point $\rho^M_c(r)$ is $(+)$-critical, so we have the relation in the statement. □

In fact, the proof of Theorem 3.9 provides much information about the birth and death parameters of the space L constructed within the proof. We will eventually use information about both M and L to recover XRPH of M from ∂M. So it makes sense to take a closer look at L and summarise its birth and death parameters.

Corollary 3.10. *Suppose we have the same set up as Proposition 3.9. Choose $R > 0$ large enough such that $M \subseteq B(c,R)$ and that there exists $\epsilon > 0$ such that all critical values of M are ϵ apart and the largest critical value is less than $R-\epsilon$. Consider $L = \overline{B(c,R)} \backslash int(M)$.*

Then if $c \notin L$, we have

$$\mathfrak{d}^{rel}_0(L,\rho_c) = \{(\rho_c^{\partial M}(p), Ord) \in \mathfrak{b}^{ord}_0(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = +1\} \cup \{(R, Rel)\};$$
$$\mathfrak{d}^{rel}_{n-1}(L,\rho_c) = \{(\rho_c^{\partial M}(p), Rel) \in \mathfrak{d}^{rel}_{n-1}(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = +1\} \sqcup \{(R, Rel)\};$$

and for $k \geq 0$ in all other cases,

$$\mathfrak{b}^{ord}_k(L,\rho_c) = \{(\rho_c^{\partial M}(p), Ord) \in \mathfrak{b}^{ord}_k(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = -1\};$$
$$\mathfrak{b}^{rel}_k(L,\rho_c) = \{(\rho_c^{\partial M}(p), Rel) \in \mathfrak{b}^{rel}_k(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = +1\};$$
$$\mathfrak{d}^{ord}_k(L,\rho_c) = \{(\rho_c^{\partial M}(p), Ord) \in \mathfrak{d}^{ord}_k(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = -1\};$$
$$\mathfrak{d}^{rel}_k(L,\rho_c) = \{(\rho_c^{\partial M}(p), Rel) \in \mathfrak{d}^{rel}_k(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = +1\}\}.$$

If $c \in L$, on the other hand, let $r = \min_{p\in\partial M} \rho_c(p)$. Then all the birth and death parameters remain the same for $k \geq 0$ except the following:

$$\mathfrak{b}^{ord}_0(L,\rho_c) = \{(\rho_c^{\partial M}(p), Ord) \in \mathfrak{b}^{ord}_0(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = -1\} \cup \{(0, Ord)\};$$
$$\mathfrak{d}^{rel}_{n-1}(L,\rho_c) = \big(\{(\rho_c^{\partial M}(p), Rel) \in \mathfrak{d}^{rel}_{n-1}(\partial M, \rho_c^{\partial M}) : sgn(\rho_c^M, p) = +1\} \sqcup \{(R, Rel)\}\big) \backslash \{(r, Rel)\};$$
$$\mathfrak{b}^{rel}_n(L,\rho_c) = \{(r, Rel)\};$$
$$\mathfrak{d}^{rel}_n(L,\rho_c) = \{(0, Rel)\}.$$

Proof. The proof for the birth and death parameters of L within $B(c,R)$, for both $c \in L$ case and $c \notin L$ case, uses the same techniques as the proof of Proposition 3.9. Note that LES (3.1) and (3.8) still hold, and so do the relations. Hence, we

get the same relations when the critical value is less than R after swapping M for L in the statement of Proposition 3.9. Moreover, note that for critical points $p \in B(c, R)$, $\mathrm{sgn}(\rho_c^M, p) = -\mathrm{sgn}(\rho_c^L, p)$ by construction. This is why we switch the signs in the relations given in Corollary 3.10.

So it remains to consider what happens at $\partial B(c, R)$. Note that we have for all $k > 0$ and $R - \epsilon < s < R + \epsilon$,

$$H_k((\partial M)_{R-\epsilon}) \cong H_k((\partial M)_s) \cong H_k(M_s) \oplus H_k(L_s).$$

Since $H_k(M_{R-\epsilon}) \cong H_k(M_{R+\epsilon}) \cong H_k(M)$ by construction, we know that (R, Ord) cannot be a critical value for L. For similar reason, by considering LES (3.8), we know that (R, Rel) cannot be a critical value for L in dimension k for $0 < k < n - 1$. The computation of n-th relative homology groups in the proof of Proposition 3.9 also shows that (R, Rel) cannot be a critical value for L in dimension n. However, the cases when $k = 0$ or $k = n - 1$ are worth considering. For $k = 0$, consider the following part of LES (3.8):

$$0 \to H_0((\partial M), (\partial M)^s) \to H_0(M, M^s) \oplus H_0(L, L^s) \to H_0(M \cup L, M^s \cup L^s) \to 0.$$

Now since $H_0(M \cup L, M^{R+\epsilon} \cup L^{R+\epsilon}) = H_0(M \cup L) = \mathbb{Z}_2$ and $H_0(M \cup L, M^{R-\epsilon} \cup L^{R-\epsilon}) = 0$, and $(\partial M)^s = M^s = \emptyset$ for $R - \epsilon \leq s \leq R + \epsilon$, it follows that

$$\beta_0(\partial M) = \beta_0(M) + \beta_0(L, L^{R+\epsilon}) - 1,$$

and

$$\beta_0(\partial M) = \beta_0(M) + \beta_0(L, L^{R-\epsilon}).$$

Taking the difference of the two equations, we get

$$\beta_0(L, L^{R-\epsilon}) - \beta_0(L, L^{R+\epsilon}) = -1.$$

Hence, we have shown that $(R, \mathrm{Rel}) \in \mathfrak{d}(L, \rho_c)$. For $k = n - 1$, consider the following commuting diagram:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & 0 & \longrightarrow & H_{n-1}((\partial M)) & \longrightarrow & H_{n-1}(M) \oplus H_{n-1}(L, L^{R+\epsilon}) & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle i_*} & & \downarrow{\scriptstyle i_*} & & \downarrow{\scriptstyle i_*} & & \\
0 & \longrightarrow & \mathbb{Z}_2 & \longrightarrow & H_{n-1}((\partial M)) & \longrightarrow & H_{n-1}(M) \oplus H_{n-1}(L, L^{R-\epsilon}) & \longrightarrow & 0
\end{array}$$

The first row is obtained by substituting $s = R + \epsilon$ in (3.10) and the second row is obtained by substituting $s = R - \epsilon$ in (3.10). The arrows in between rows are induced by inclusion maps. This tells us that

$$\beta_{n-1}(\partial M) = \beta_{n-1}(L, L^{R+\epsilon}) + \beta_{n-1}(M),$$

and

$$\beta_{n-1}(\partial M) = \beta_{n-1}(L, L^{R-\epsilon}) + \beta_{n-1}(M) + 1.$$

Hence, taking the difference of the two equations, we get

$$\beta_{n-1}(L, L^{R-\epsilon}) - \beta_{n-1}(L, L^{R+\epsilon}) = -1,$$

and hence, we have $(R, \mathrm{Rel}) \in \mathfrak{d}^{\mathrm{rel}}_{n-1}(L, \rho_c)$, as required. □

The assumptions in the proof of Proposition 3.9 will carry on for the rest of this section. So we listed here.

In what follows, we will assume that $(M, \partial M)$ is a compact n-manifold with boundary such that the radial function ρ_c^M is Morse for some fixed centre $c \in \mathbb{R}^n \backslash \partial M$. Choose $R > 0$ large enough such that $M \subseteq B(c, R)$ and that there exists $\epsilon > 0$ such that all critical values of M are ϵ apart and the largest critical value of ρ_c^M is less than $R - \epsilon$. Define $L \coloneqq \overline{B(c, R)} \backslash \mathrm{int} M$. Then $M \cup L = \overline{B(c, R)}$ and $M \cap L = \partial M$. Finally, let $r \coloneqq \min_{p \in \partial M} \rho_c(p)$, and $\hat{r} \coloneqq \max_{p \in \partial M} \rho_c(p)$.

Corollary 3.11. *We have equality of the following disjoint unions:*

$$\begin{aligned}
\mathfrak{b}_0(M, \rho_c) \sqcup \mathfrak{b}_0(L, \rho_c) &= \mathfrak{b}_0(\partial M, \rho_c) \sqcup \{(0, Ord)\};\\
\mathfrak{d}_0(M, \rho_c) \sqcup \mathfrak{d}_0(L, \rho_c) &= \mathfrak{d}_0(\partial M, \rho_c) \sqcup \{(R, Rel)\};\\
\mathfrak{d}_{n-1}(M, \rho_c) \sqcup \mathfrak{d}_{n-1}(L, \rho_c) &= \big(\mathfrak{d}_{n-1}(\partial M, \rho_c) \sqcup \{(R, Rel)\}\big) \backslash \{(r, Rel)\};\\
\mathfrak{b}_n(M, \rho_c) \sqcup \mathfrak{b}_n(L, \rho_c) &= \{(r, Rel)\};\\
\mathfrak{d}_n(M, \rho_c) \sqcup \mathfrak{d}_n(L, \rho_c) &= \{(0, Rel)\};
\end{aligned}$$

and for $k > 0$ in all other cases:

$$\begin{aligned}
\mathfrak{b}_k(M, \rho_c) \sqcup \mathfrak{b}_k(L, \rho_c) &= \mathfrak{b}_k(\partial M, \rho_c);\\
\mathfrak{d}_k(M, \rho_c) \sqcup \mathfrak{d}_k(L, \rho_c) &= \mathfrak{d}_k(\partial M, \rho_c).
\end{aligned}$$

Remark 3.12. Corollary 3.11 is written in a relatively compact form, but there is still a bit more information we can unpack here by combining the information from Proposition 3.9. For instance, the point $(0, \mathrm{Ord})$ is contained in precisely one of the sets $\mathfrak{b}_0(M, \rho_c)$ and $\mathfrak{b}_0(L, \rho_c)$. That is, if $c \in \mathrm{int}(M)$, then $(0, \mathrm{Ord}) \in \mathfrak{b}_0(M, \rho_c)$, and vice versa. This observation is particularly useful when we state Theorem 3.13. Furthermore, from the proof of Proposition 3.9, we know that if $c \notin S$, then $\mathfrak{b}_n(S, \rho_c) = \mathfrak{d}_n(S, \rho_c) = \emptyset$, for $S = M$ or L.

Proof. Observe that if $c \in M$, then $c \notin L$ and vice versa. Since every critical point of $\rho_c^{\partial M}$ is either $(+)$-critical or $(-)$-critical, taking the union of the relevant relations in Proposition 3.9 and Corollary 3.10 gives the claim. □

Corollary 3.11 compares the birth and death parameters over the subsets $\partial M, M$ and L. In fact, a stronger result (Theorem 3.13), which says that the decomposition of persistence parings is (mostly) consistent, holds. The proof for the result is almost identical to the proof of Theorem 4.17 in [29] for $0 < k < n-1$. The only modification required is to change the height function to radial functions, but we do not need to use any assumption about radial functions. However, extra consideration needs to be included for $k = 0, n$ and $n - 1$ since we have seen in Proposition 3.9 and in the proof of Corollary 3.11 that the birth and death parameters for the three sets do not match up as nicely in these dimensions as those for the other dimensions. We further note that since when $c \in M$, then $(0, \mathrm{Ord})$ is a birth parameter in $\mathfrak{b}_0(\mathrm{XRPH}_0(M, \rho_c))$, but $R, \mathrm{Rel})$ will be a death parameter in $\mathfrak{b}_0(\mathrm{XRPH}_0(L, \rho_c))$. This means when $c \in M$, the extra birth and death parameters in dimension 0 do not correspond to a single component in M or L. Hence, the matching of interval decomposition becomes more difficult. However, the centre points we choose in Section 2.2.3 are all in the non-vascular regions, which correspond to the case $c \notin M$ in our setup.

Hence, from now on, we assume that $c \notin M$.

Theorem 3.13. *We have the following equalities of direct sums:*

$$\begin{aligned} XRPH_0(\partial M, \rho_c) \oplus \mathcal{I}_{[(0,Ord),(R,Rel))} &= XRPH_0(M, \rho_c) \oplus XRPH_0(L, \rho_c), \\ XRPH_k(\partial M, \rho_c) &= XRPH_k(M, \rho_c) \oplus XRPH_k(L, \rho_c), \end{aligned}$$

for all $0 < k < n - 1$, *and*

$$\begin{aligned} XRPH_n(L, \rho_c) &= \mathcal{I}_{[(r,Rel),(0,Rel))}, \\ XRPH_n(M, \rho_c) &= 0. \end{aligned}$$

Remark 3.14. Note that in the statement of Theorem 3.13, the extended persistence pairings for dimension $n-1$ is not included. This is because regardless of position of c, the effect of mismatch in the death parameters is not caused solely by what happens at $\partial B(c, R)$. However, it is still possible to recover the interval decomposition of M from that of ∂M in dimension $n-1$ in certain cases. We will discuss this more after the proof of Theorem 3.13.

Proof. For $0 < k < n-1$, since $\partial M \subseteq M$ and $\partial M \subseteq L$, there is an induced morphism on persistence modules:

$$\phi_k : \mathrm{XRPH}_k(\partial M, \rho_c) \to \mathrm{XRPH}_k(M, \rho_c) \oplus \mathrm{XRPH}_k(L, \rho_c).$$

Moreover, LES (3.1) and (3.8) above shows that $\phi_{(t,\mathrm{Ord})}$ and $\phi_{(t,\mathrm{Rel})}$ are both isomorphisms. Hence, ϕ_k is injective.

Induced matching theorem in [3] implies that there exists an induced injective map $\bar{\phi}_k$ from ϕ_k on the set of intervals in the interval decomposition of

$$\mathrm{XRPH}_k(\partial M, \rho_c)$$

to the set of intervals in the interval decomposition of

$$\mathrm{XRPH}_k(M, \rho_c) \oplus \mathrm{XRPH}_k(L, \rho_c).$$

Furthermore, if $[b.d)$ is an interval in $\mathrm{XRPH}_k(\partial M, \rho_c)$, then we have

$$\bar{\phi}_k([b, d)) = [b', d)$$

with $b' \leq d$.

Now we know from Corollary 3.11 that

$$\mathfrak{b}_k(\partial M, \rho_c) = \mathfrak{b}_k(M, \rho_c) \sqcup \mathfrak{b}_k(L, \rho_c).$$

So the two persistence modules have the same number of intervals, and thus the induced map $\bar{\phi}$ is a bijection. Since $\bar{\phi}_k([b, d)) = [b', d)$ with $b' \leq d$, we have that $\bar{\phi}_k$ is actually the identity map, i.e. $\bar{\phi}_k([b, d)) = [b, d)$. This means the interval decomposition of $\mathrm{XRPH}_k(\partial M, \rho_c)$ agrees with that of $\mathrm{XRPH}_k(M, \rho_c) \oplus \mathrm{XRPH}_k(L, \rho_c)$, and hence we have shown that

$$\mathrm{XRPH}_k(\partial M, \rho_c) = \mathrm{XRPH}_k(M, \rho_c) \oplus \mathrm{XRPH}_k(L, \rho_c),$$

as required.

Now for $k = 0$, we need to also consider the homology class of $\partial B(c, R)$. Again, from Corollary 3.11, we have

$$\mathfrak{b}_0(\partial M, \rho_c) \sqcup \{(0, \mathrm{Ord})\} = \mathfrak{b}_0(M, \rho_c) \sqcup \mathfrak{b}_0(L, \rho_c),$$

denoted by $\mathfrak{b}$ and

$$\mathfrak{d}_0(\partial M, \rho_c) \sqcup \{(R, \mathrm{Rel})\} = \mathfrak{d}_0(M, \rho_c) \sqcup \mathfrak{d}_0(L, \rho_c),$$

which we denote by $\mathfrak{d}$. So we can define a bijection $\hat{\phi}_0 : \mathfrak{b} \to \mathfrak{b}$ by $\hat{\phi}_0(b) = b'$ if there exists a $d \in \mathfrak{d}$ such that

$$\mathcal{I}_{[b,d)} \in \mathrm{XRPH}_0(\partial M, \rho_c) \oplus \mathcal{I}_{[(0,\mathrm{Ord}),(R,\mathrm{Rel}))},$$

and

$$\mathcal{I}_{[b',d)} \in \mathrm{XRPH}_0(M, \rho_c) \oplus \mathrm{XRPH}_0(L, \rho_c).$$

Since $c \notin M$, we have $c \in L$ and hence, $[(0, \mathrm{Ord}), (R, \mathrm{Rel}))$ is an interval in the interval decomposition of $\mathrm{XRPH}_0(L, \rho_c)$. This corresponds to the connected component of L containing $\partial B(c, R)$. Hence, we have $\hat{\phi}((0, \mathrm{Ord})) = (0, \mathrm{Ord})$.

Again by checking the ordinary and relative Mayer Vietories sequences (3.1) and (3.8), we know that both maps

$$H_0((\partial M)_s) \to H_0(M_s) \oplus H_0(L_s),$$

and

$$H_0(\partial M, (\partial M)^s) \to H_0(M, M^s) \oplus H_0(L, L^s)$$

are induced by inclusion and are injective for all $0 \leq s \leq R$. Hence, the morphism

$$\phi_0 : \mathrm{XRPH}_0(\partial M, \rho_c) \to \mathrm{XRPH}_0(M, \rho_c) \oplus \mathrm{XRPH}_0(L, \rho_c)$$

is injective. Again, the map ϕ_0 induces an injective map $\bar{\phi}_0$ that matches the interval $[b, d)$ in the interval decomposition of

$$\mathrm{XRPH}_0(\partial M, \rho_c)$$

to the interval $[b', d)$ in the interval decomposition of

$$\mathrm{XRPH}_0(M, \rho_c) \oplus \mathrm{XRPH}_0(L, \rho_c)$$

such that $b' \leq b$. Combined with the above, we have that $\hat{\phi}_0 : \mathfrak{b} \to \mathfrak{b}$ with $\hat{\phi}_0(b) \leq b$ for all $b \in \mathfrak{b}_0(\mathrm{XRPH}_0(\partial M, \rho_c))$. Since $\mathfrak{b} \backslash \mathfrak{b}_0(\mathrm{XRPH}_0(\partial M, \rho_c)) = \{(0, \mathrm{Ord})\}$,

combined with the previous argument, we have shown that $\hat{\phi}_0(b) \leq b$ for all $b \in \mathfrak{b}$. It follows that $\hat{\phi}_0$ is the identity map since $\mathfrak{b}$ is a finite set. Hence, the interval decomposition of $\mathrm{XRPH}_0(\partial M, \rho_c)$ agrees with that of $\mathrm{XRPH}_0(M, \rho_c) \oplus \mathrm{XRPH}_k(L, \rho_c)$, which implies that

$$\mathrm{XRPH}_0(\partial M, \rho_c) \oplus \mathcal{I}_{[(0,\mathrm{Ord}),(R,\mathrm{Rel}))} = \mathrm{XRPH}_0(M, \rho_c) \oplus \mathrm{XRPH}_0(L, \rho_c),$$

as required.

Finally, for $k = n$, from Proposition 3.11, we have

$$\{(r, \mathrm{Rel})\} = \mathfrak{b}_n(M, \rho_c) \sqcup \mathfrak{b}_n(L, \rho_c),$$

and

$$\{(0, \mathrm{Rel})\} = \mathfrak{d}_n(M, \rho_c) \sqcup \mathfrak{d}_n(L, \rho_c).$$

Since $c \notin M$, we observe that $[(r, \mathrm{Rel}), (0, \mathrm{Rel}))$ is an interval in the interval decomposition of $\mathrm{XRPH}_n(L, \rho_c)$. Furthermore, we have $(r, \mathrm{Rel}) \notin \mathfrak{b}_n(M, \rho_c)$ and $(0, \mathrm{Rel}) \notin \mathfrak{d}_n(M, \rho_c)$. Hence, combining them, we get

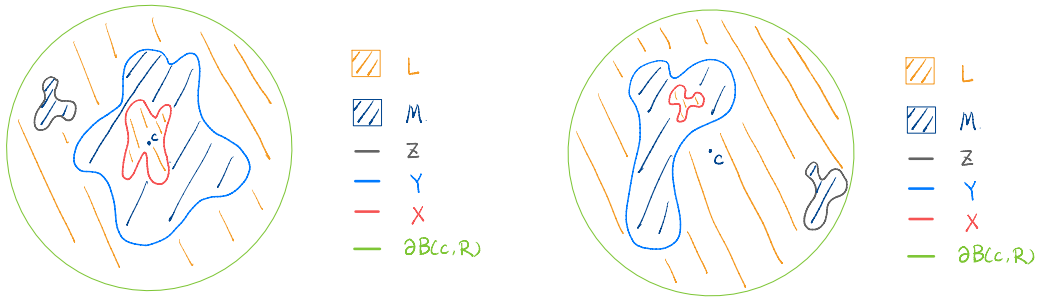
$$\mathcal{I}_{[(r,\mathrm{Rel}),(0,\mathrm{Rel}))} = \mathrm{XRPH}_n(L, \rho_c),$$

as required. □

As mentioned above, even when $c \notin M$, the extended persistence pairings in dimension $n-1$ for ∂M do not match up as nicely to those for M and L as in other dimensions. In particular, there are different cases that we need to consider. We illustrate this with two scenarios.

Example 3.15. Consider the two scenarios in Figure 3.1. In both cases, we have M being a 2-manifold with boundary embedded in $\mathbb{R}^2$ with a 1-dimensional hole and two connected components. Moreover, these are both examples of $c \notin M$ and ρ_c^M being a Morse function. However, in example (a), the centre c is inside the finite component of $\mathbb{R}^2 \backslash X$, which we will call the *cavity* of M. This is the main difference between the two examples. Now we consider the 1-dimensional extended persistence pairings of M, L and $\partial M = X \sqcup Y$ for example (a) and (b). Define

$$r_X \coloneqq \min_{p \in X} \rho_c(p), \hat{r}_X \coloneqq \max_{p \in X} \rho_c(p),$$
$$r_Y \coloneqq \min_{p \in Y} \rho_c(p), \hat{r}_Y \coloneqq \max_{p \in Y} \rho_c(p),$$
$$r_Z \coloneqq \max_{p \in Z} \rho_c(p), \hat{r}_Z \coloneqq \max_{p \in Z} \rho_c(p).$$



(a) (b)

Figure 3.1: Two scenarios of $c \notin M$, where M is a manifold with boundary (X, Y and Z are the boundary components) embedded in $\mathbb{R}^2$. In Figure (a), the centre c is the cavity of M; while in Figure (b), the centre c is not.

Note that in different examples, these values are different. We use the same notation for easier comparison of the extended persistence pairings.

In example (a), we have the following extended persistence modules:

$$\begin{aligned}
\mathrm{XRPH}_1(\partial M, \rho_c) &= \mathcal{I}_{[(\hat{r}_X,\mathrm{Ord}),(r_X,\mathrm{Rel}))} \oplus \mathcal{I}_{[(\hat{r}_Y,\mathrm{Ord}),(r_Y,\mathrm{Rel}))} \oplus \mathcal{I}_{[(\hat{r}_Z,\mathrm{Ord}),(r_Z,\mathrm{Rel}))},\\
\mathrm{XRPH}_1(M, \rho_c) &= \mathcal{I}_{[(\hat{r}_X,\mathrm{Ord}),(r_Y,\mathrm{Rel}))},\\
\mathrm{XRPH}_1(L, \rho_c) &= \mathcal{I}_{[(\hat{r}_Z,\mathrm{Ord}),(R,\mathrm{Rel}))} \oplus \mathcal{I}_{[(\hat{r}_Y,\mathrm{Ord}),(r_Z,\mathrm{Rel}))}.
\end{aligned}$$

Now in example (b), we have the following extended persistence modules:

$$\begin{aligned}
\mathrm{XRPH}_1(\partial M, \rho_c) &= \mathcal{I}_{[(\hat{r}_X,\mathrm{Ord}),(r_X,\mathrm{Rel}))} \oplus \mathcal{I}_{[(\hat{r}_Y,\mathrm{Ord}),(r_Y,\mathrm{Rel}))} \oplus \mathcal{I}_{[(\hat{r}_Z,\mathrm{Ord}),(r_Z,\mathrm{Rel}))},\\
\mathrm{XRPH}_1(M, \rho_c) &= \mathcal{I}_{[(\hat{r}_X,\mathrm{Ord}),(r_X,\mathrm{Rel}))},\\
\mathrm{XRPH}_1(L, \rho_c) &= \mathcal{I}_{[(\hat{r}_Z,\mathrm{Ord}),(R,\mathrm{Rel}))} \oplus \mathcal{I}_{[(\hat{r}_Y,\mathrm{Ord}),(r_Z,\mathrm{Rel}))}.
\end{aligned}$$

Note that in example (a), we have $r_X = \min_{p\in\partial M} \rho_c(p)$, and in example (b), we have $r_Y = \min_{p\in\partial M} \rho_c(p)$ instead. In both cases, those values do not appear in the birth or death parameters of M and L for dimension 1, which agrees with our result in Corollary 3.11.

In example (a), we can see that neither $\mathrm{XRPH}_1(M, \rho_c)$ nor $\mathrm{XRPH}_1(L, \rho_c)$ match with any submodule of $\mathrm{XRPH}_1(\partial M, \rho_c)$. But we can see that $\mathrm{XRPH}_1(M, \rho_c)$ match up with a submodule of $\mathrm{XRPH}_1(\partial M, \rho_c)$, and we will show that this is the case in general. At the same time, we note that $\mathrm{XRPH}_1(L, \rho_c)$ does not match with the remaining submodule of $\mathrm{XRPH}_1(\partial M, \rho_c)$ completely. While there is still some patterns when it comes to describe $\mathrm{XRPH}_1(L, \rho_c)$ using interval modules in $\mathrm{XRPH}_1(\partial M, \rho_c)$, it becomes more complicated as the number of connected

components in M gets larger. More importantly, we do not actually care about $\mathrm{XRPH}_1(L, \rho_c)$, it is the recovery of $\mathrm{XRPH}_1(M, \rho_c)$ from $\mathrm{XRPH}_1(\partial M, \rho_c)$ that we are interested in.

Since in our retinal images, the centre of optical disc is always going to be outside of any "loop" in the 2D projection of the retinal vessels. So we only need to consider the scenarios similar to the one in case (b) in Example 3.15.

From this point onwards, assume further that the centre c is not contained in any cavity of M.

Before we describe and prove the statement about recovering $\mathrm{XRPH}_{n-1}(M, \rho_c)$ from $\mathrm{XRPH}_{n-1}(\partial M, \rho_c)$, we differentiate between interior and exterior boundary components that we briefly discussed in Section 1.4.

Definition 3.16 (Interior and exterior boundary compoent)**.** Let $A \subseteq \mathbb{R}^n$ be a compact n-manifold with boundary $\partial A = X$. Let $\widetilde{X}$ be a connected component of X and $\widetilde{A}$ be the connected component of A that contains $\widetilde{X}$. Then $\widetilde{X}$ is an *interior boundary component* if $\widetilde{A} \backslash \widetilde{X}$ is contained in the unbounded connected component of $\mathbb{R}^n \backslash \widetilde{X}$. We call $\widetilde{X}$ an *exterior boundary component* if $\widetilde{M} \backslash \widetilde{X}$ is contained in the bounded component of $\mathbb{R}^n \backslash \widetilde{X}$.

Theorem 3.17. *Let* $\hat{\mathfrak{b}} = \{(\rho_c^{\partial M}(p), \mathit{Ord}) \in \mathfrak{b}_{n-1}^{ord}(\partial M, \rho_c^{\partial} M) : \mathit{sgn}(\rho_c^M, p) = +1\} \sqcup \{(\rho_c^{\partial M}(p), \mathit{Rel}) \in \mathfrak{b}_{n-1}^{rel}(\partial M, \rho_c^{\partial} M) : \mathit{sgn}(\rho_c^M, p) = -1\}$. *Consider the submodule* $\widetilde{XRPH}_{n-1}(\partial M, \rho_c) \coloneqq \sum_{i=1}^m \mathcal{I}_{I_i}$ *of* $XRPH_{n-1}(\partial M, \rho_c)$, *where* $\mathfrak{b}(I_i) \in \hat{\mathfrak{b}}$ *for all* i. *Then we have*

$$\widetilde{XRPH}_{n-1}(\partial M, \rho_c) = XRPH_{n-1}(M, \rho_c).$$

Proof. Suppose M is a disjoint union of connected components M_i. Then we have

$$\mathrm{XRPH}_{n-1}(M, \rho_c) = \mathrm{XRPH}_{n-1}(\bigsqcup M_i, \rho_c) = \oplus_i \mathrm{XRPH}_{n-1}(M_i, \rho_c).$$

So it is sufficient to prove the theorem in the case where M is connected.

Recall that we have defined $r \coloneqq \min_{p \in \partial M} \rho_c(p)$ and $\hat{r} \coloneqq \max_{p \in \partial M} \rho_c(p)$. From Corollary 3.11, we know that

$$\mathfrak{b}_{n-1}(\partial M, \rho_c) = \mathfrak{b}_{n-1}(M, \rho_c) \sqcup \mathfrak{b}_{n-1}(L, \rho_c),$$

which we will denote by $\mathfrak{b}$ and

$$\big(\mathfrak{d}_{n-1}(\partial M, \rho_c) \sqcup \{(R, \mathrm{Rel})\}\big) \backslash \{(r, \mathrm{Rel})\} = \mathfrak{d}_{n-1}(M, \rho_c) \sqcup \mathfrak{d}_{n-1}(L, \rho_c),$$

which we will denote by $\mathfrak{d}$.

So we can define a bijection $\hat{\phi}_{n-1} : \mathfrak{d} \to \mathfrak{d}$ by $\hat{\phi}_{n-1}(d) = d'$ if there exists a $b \in \mathfrak{d}$ such that

$$\mathcal{I}_{[b,d)} \in \mathrm{XRPH}_{n-1}(M, \rho_c) \oplus \mathrm{XRPH}_{n-1}(L, \rho_c),$$

and

$$\mathcal{I}_{[b,d')} \in \mathrm{XRPH}_{n-1}(\partial M, \rho_c).$$

Since M is connected, we observe that $[(\hat{r}, \mathrm{Ord}), (R, \mathrm{Rel}))$ is an interval in the interval decomposition of $\mathrm{XRPH}_{n-1}(L, \rho_c)$ that corresponds to the exterior boundary component of M. Moreover, note that $[(\hat{r}, \mathrm{Ord}), (r, \mathrm{Rel})$ is an interval in the interval decomposition of $\mathrm{XRPH}_{n-1}(\partial M, \rho_c)$ that also corresponds to the exterior boundary component of M. Hence, we have $\hat{\phi}_{n-1}(R, \mathrm{Rel}) = (r, \mathrm{Rel})$. Note that $(r, \mathrm{Rel}) \geq (R, \mathrm{Rel})$ since $r < R$ by construction.

From LES (3.1) and (3.8), we have that the maps

$$H_{n-1}((\partial M)_s) \to H_{n-1}(M_s) \oplus H_{n-1}(L_s),$$

and

$$H_{n-1}(\partial M, (\partial M)^s) \to H_{n-1}(M, M^s) \oplus H_{n-1}(L, L^s)$$

are surjective. So the induced morphism by inclusion

$$\phi_{n-1} : \mathrm{XRPH}_{n-1}(\partial M, \rho_c) \to \mathrm{XRPH}_{n-1}(M, \rho_c) \oplus \mathrm{XRPH}_{n-1}(L, \rho_c)$$

is surjective. The induced matching theorem in [3] then tells us that there exists an induced injective map $\overline{\phi_{n-1}}$ from ϕ_{n-1} on the set of intervals in the interval decomposition of

$$\mathrm{XRPH}_{n-1}(M, \rho_c) \oplus \mathrm{XRPH}_{n-1}(L, \rho_c)$$

to the set of intervals in the interval decomposition of

$$\mathrm{XRPH}_{n-1}(\partial M, \rho_c).$$

Moreover, if $[b, d)$ is an interval in $\mathrm{XRPH}_{n-1}(M, \rho_c) \oplus \mathrm{XRPH}_{n-1}(L, \rho_c)$, then we have

$$\overline{\phi}_{n-1}([b, d)) = [b, d')$$

with $d' \geq d$.

Combined with the above, we have that $\hat{\phi}_{n-1} : \mathfrak{d} \to \mathfrak{d}$ with $\hat{\phi}_{n-1}(d) \geq d$ for all $d \in \mathfrak{d}(\mathrm{XRPH}_{n-1}(M, \rho_c) \oplus \mathrm{XRPH}_{n-1}(L, \rho_c)) = \mathfrak{d}$. So It follows that $\hat{\phi}_{n-1}|_{\mathfrak{d} \setminus \{(R, \mathrm{Rel})\}}$ is the identity map since $\mathfrak{d}$ has finite elements.

But since $\mathcal{I}_{[(\hat{r},\mathrm{Ord}),(R,\mathrm{Rel}))} \in \mathrm{XRPH}_{n-1}(L,\rho_c)$, we have shown that the interval decomposition of $\mathrm{XRPH}_{n-1}(M,\rho_c)$ agrees with that of $\widetilde{\mathrm{XRPH}}_{n-1}(\partial M,\rho_c)$ as defined in the statement. This implies that

$$\widetilde{\mathrm{XRPH}}_{n-1}(\partial M,\rho_c) = \mathrm{XRPH}_{n-1}(M,\rho_c),$$

as required. □

Combining Theorem 3.13, Theorem 3.17 and Corollary 3.11 allows us to express the radial extended persistence pairings of ρ_c^M in terms of a submodule of that of $\rho_c^{\partial M}$. We summarise the result as Theorem 3.18.

Theorem 3.18. *Suppose the k-dimensional radial extended persistence homology of $\rho_c^{\partial M} : X \to \mathbb{R}$ be*

$$XRPH_k(\partial M,\rho_c) = \bigoplus_{[b_i,d_i)\in S_{\partial M}} \mathcal{I}_{[b_i,d_i)}.$$

Let $J_M \subseteq S_{\partial M}$ be the subset of intervals $[b_i,d_i)$ such that either $b_i = (\rho_c(p), Ord)$ with $p \in Crit(\rho_c^M,(k,+1))$, or $b_i = (\rho_c(p), Rel)$ with $p \in Crit(\rho_c^M,(n-k-1,-1))\}$. Then we have

$$XRPH_k(M,\rho_c) = \bigoplus_{[b_i,d_i)\in J_M^k} \mathcal{I}_{[b_i,d_i)},$$

for all $k \geq 0$.

Finally, in the radial extended persistence diagram, we would like to classify each point into one of the ordinary, relative and essential classes. At the same time, if we can describe the essential classes, then the remaining classes can be decided accordingly by comparing the values for the birth and death parameters. Here, we can further describe the essential classes for dimension 0 and $n-1$ using the symmetry theorem if M is an n-manifold with boundary as its boundary is an $n-1$-manifold. Hence, if $n = 2$, we are able to fully decompose a radial extended persistence module of a manifold with boundary in all dimensions since there will be no persistence module for dimension 2 or higher. This is the basis of our algorithm used to obtain the results in Section 2.2.3.

Proposition 3.19. *Let $M \subseteq \mathbb{R}^n$ be an n-manifold with boundary ∂M. Consider the radial function $\rho_c : \mathbb{R}^n \to \mathbb{R}$ with suitably chosen c satisfying the assumptions discussed above. Let $\{X_1,\cdots,X_k\}$ be the interior boundary components of ∂M and $\{Y_1,\cdots,Y_l\}$ be the exterior boundary components of ∂M. Then we have*

$$Ess_0(M,\rho_c) = \sum_{j=1}^{l} \mathcal{I}_{[(\min_{p\in Y_j}\rho_c(p),Ord),(\max_{p\in Y_j}\rho_c(p),Rel))},$$

and

$$Ess_{n-1}(M,\rho_c) = \sum_{j=1}^{k} \mathcal{I}_{[(\max_{p\in X_j}\rho_c(p),Ord),(\min_{p\in X_j}\rho_c(p),Rel))}.$$

Proof. If $M = \sqcup_{i=1}^{k} M_i$ is a disjoint union of connected components, then $\mathrm{XRPH}(M) = \oplus_{i=1}^{k}\mathrm{XRPH}(M_i)$. So we can assume without loss of generality that M is connected.

Note that ∂M is a disjoint union of $(n-1)$-manifolds. Moreover, for a connected $(n-1)$-manifold A, we have $\beta_0(A) = \beta_{n-1}(A) = 1$. Hence, exactly one of the intervals in the interval decomposition of $\mathrm{XRPH}(A,\rho_c)$ comes from the essential class. Furthermore, the interval in $\mathrm{Ess}_0(A,\rho_c)$ is born when it first sees A at $(\min_{p\in A}\rho_c(p), \mathrm{Ord})$ and dies at the largest value A takes, i.e. at $(\max_{p\in A}\rho_c(p), \mathrm{Rel})$. Hence,

$$\mathrm{Ess}_0(A,\rho_c) = \mathcal{I}_{[(\min_{p\in A}\rho_c(p),\mathrm{Ord}),(\max_{p\in A}\rho_c(p),\mathrm{Rel}))}.$$

Now by Theorem 1.47, we have that

$$\mathrm{Ess}_{n-1}(A,\rho_c) = \mathcal{I}_{[(\max_{p\in A}\rho_c(p),\mathrm{Ord}),(\min_{p\in A}\rho_c(p),\mathrm{Rel}))}.$$

Now since M is connected, we have that $\partial M = Y \sqcup \left(\sqcup_{i=1}^{k} X_i\right)$. So it follows that

$$\mathrm{Ess}_0(\partial M,\rho_c) = \mathcal{I}_{[(\min_{p\in Y}\rho_c(p),\mathrm{Ord}),(\max_{p\in Y}\rho_c(p),\mathrm{Rel}))}\oplus \left(\oplus_{i=1}^{k}\mathcal{I}_{[(\min_{p\in X_i}\rho_c(p),\mathrm{Ord}),(\max_{p\in X_i}\rho_c(p),\mathrm{Rel}))}\right),$$

and

$$\mathrm{Ess}_{n-1}(\partial M,\rho_c) = \mathcal{I}_{[(\max_{p\in Y}\rho_c(p),\mathrm{Ord}),(\min_{p\in Y}\rho_c(p),\mathrm{Rel}))}\oplus \left(\oplus_{i=1}^{k}\mathcal{I}_{[(\max_{p\in X_i}\rho_c(p),\mathrm{Ord}),(\min_{p\in X_i}\rho_c(p),\mathrm{Rel}))}\right).$$

Now using Theorem 3.18, we are able to translate these to the essential persistence modules of M.

Consider an interior boundary component X_i, then M is contained in the infinite component of $\mathbb{R}^n\backslash X_i$. Let p, q be the points on X_i that attains the minimum and maximum values of $\rho_c^{X_i}$ respectively. Then $\mathrm{sgn}(p) = -1$ and $\mathrm{sgn}(q) = +1$. At the same time, let p', q' be the points on Y that attains the minimum and maximum values of ρ_c^{Y} respectively. Then we have $\mathrm{sgn}(p') = +1$ and $\mathrm{sgn}(q') = -1$.

Hence, by checking the signs of the critical points above,we have that

$$\mathrm{Ess}_0(X,\rho_c) = \mathcal{I}_{[(\min_{p\in Y}\rho_c(p),\mathrm{Ord}),(\max_{p\in Y}\rho_c(p),\mathrm{Rel}))},$$

and

$$\mathrm{Ess}_{n-1}(M,\rho_c) = \sum_{i=1}^{k} \mathcal{I}_{[(\max_{p\in X_i}\rho_c(p),\mathrm{Ord}),(\min_{p\in X_i}\rho_c(p),\mathrm{Rel}))},$$

as required. □

Remark 3.20. Note that in Section 1.4, the main modification we made to the algorithm in [29] for height filtrations is by changing the tests for local minimum. This is because under the assumptions we have here, the conclusions are exactly the same as the ones for height filtrations. Due to time constraints, we are not able to produce nice relations between extended persistence pairings of $\mathrm{XRPH}_k(\partial M,\rho_c)$ and $\mathrm{XRPH}_k(M,\rho_c)$ for $k \geq 0$ when we consider other possible choices of the location for the centre c, e.g. if $c \in M$ or if $c \notin M$ but c is in the cavity of M. We leave it for as a potential future direction.

Appendix A

Centre location

The following Matlab function is used to compute the coordinates of the centres of optic disks in the original retinal images. The coordinates are then used as the centre for radial filtration.

```
function [x,y,p] = centreLocate(p1)
% this function locates the bright spot inside retina for RIDB retinal
% images so that the coordinates can be used as the centre point for EPHT
% with radial functions.
% input, p1, initial retinal images
% output, x,y, x,y-coordinates of the bright spot
%         p, original image with the bright spot located with an 'x'

 % red channel
 p1_red = p1(:,:,1);

 % disk matrix
 imageSizeX = 150;
 imageSizeY = 150;
 [columnsInImage, rowsInImage] = meshgrid(1:imageSizeX, 1:imageSizeY);
 % Next create the circle in the image.
 centerX = 75;
 centerY = 75;
```

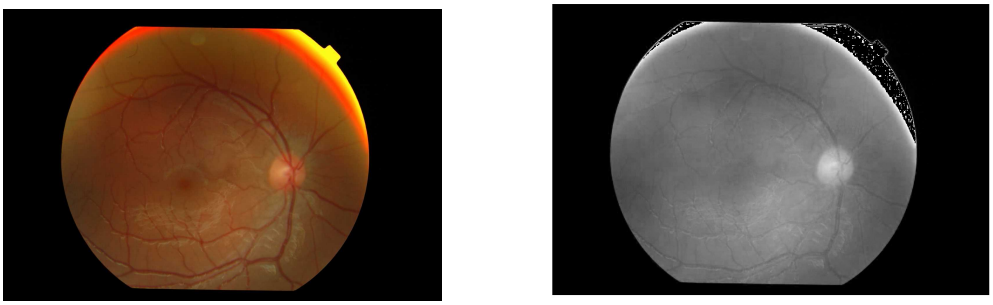
```
    radius = 50;
    circlePixels = (rowsInImage - centerY).^2 ...
        + (columnsInImage - centerX).^2 <= radius.^2;
    % imshow(circlePixels)

    % convolution
    temp = conv2(circlePixels,p1_red);
    [M1,I1]=max(max(temp));
    [M2,I2]=max(temp(:,I1));

    % x, y coordinates
    x = I1-75;
    y = I2-75;
    % display figure to check correctness
    fig = figure;
    imshow(p1)
    hold on
    plot(x,y,'xb')
    title('Original image')
    drawnow
    frame = getframe(fig);
    p = frame2im(frame);
end
```

We note that the above function works for all but one images (Figure A.1 is the problem image). However, the function works after we filter out the bright rim by setting a threshold pixel value, i.e. after we run the problem image through the following codes.

```
p1 = imread('RIDB/data/IM000002_1.JPG'); % read in problem
    image
p1_red = p1(:,:,1); % filter out red channel
p1_red(p1(:,:,1)>250)= 0; % remove the bright rim by
   setting those pixel values to 0 (black)
```



(a) Original problem image

(b) Fixed problem image after filtering

Figure A.1: The function centreLocate fails to find the correct centre coordinate for the second image sample from individual 1. This is due to the extreme exposure on the top right of the image, which leads to a higher pixel value when restricting to the red channel.

Appendix B

Computing distances between persistence diagrams.

The following R function is used to compute distances between persistence pairings obtained using Diamorse.

```
#' Compute the distance matrix between Extended Persistent Homology Transforms
#'
#' The function [computePersistentDistanceMatrix()] computes the q-Wasserstein distance
#'  between the persistent homology transforms of multiple images.
#'
#' Given a collection of objects \eqn{O_1,\dots,O_n}, compute the distance matrix \eqn{D}$ between all pairs
#'  of objects \eqn{O_i,O_j} for \eqn{i\neq j}.
#'
#' @param diagrams `list` of extended persistence diagrams of multiple objects.
#' @param q `numeric` determining the \eqn{L^q} distance to use. Must have
#'  \eqn{1\leq q < \infty}. The default value is 1.
#' @export
computePersistentDistanceMatrix <- function(diagrams,
                                            q = 1) {
  nObjects <- length(diagrams)
```

```
16
17    distance_matrix <- matrix(0, nrow = nObjects, ncol =
          nObjects)
18
19    # Distance matrix is symmetric, so only compute upper
          triangular part.
20    # Diagonal entries are zero.
21    for (i in 1:(nObjects - 1)) {
22      for (j in (i + 1):(nObjects)) {
23        object_1 <- diagrams[[i]]
24        object_2 <- diagrams[[j]]
25
26        d_ij <- alignedDistancePer(
27            object_1 = object_1,
28            object_2 = object_2,
29            q = q
30          )
31
32        d_ij <- d_ij^(1 / q)
33
34        distance_matrix[i, j] <- d_ij
35        distance_matrix[j, i] <- d_ij
36
37        print(paste("Computed distance (", i, ",", j, ") and
              (",
38                    j, ",", i, ")",
39                    sep = ""
40        ))
41      }
42    }
43
44    return(distance_matrix)
45  }
46
47
48
49  alignedDistancePer <- function(object_1, object_2, q) {
```

```
50    ds <- c("dim0", "dim1")
51
52    total <- 0
53
54    for(x in ds){
55      if(length(object_1[[x]])!=0 || length(object_2[[x]])!=
          0){
56        inf_1 <- length(object_1[[x]][is.infinite(rowSums(
            object_1[[x]])),])/2
57        inf_2 <- length(object_2[[x]][is.infinite(rowSums(
            object_2[[x]])),])/2
58        if (inf_1 != inf_2){
59          total <- Inf
60        } else{
61          object_1_noInf <- object_1[[x]][!is.infinite(
              rowSums(object_1[[x]])),]
62          object_2_noInf <- object_2[[x]][!is.infinite(
              rowSums(object_2[[x]])),]
63          total_temp <- pointDistancePer(
64            points_1 = object_1_noInf,
65            points_2 = object_2_noInf,
66            q = q
67          )
68          total <- total + total_temp + sum(abs(object_1[[x
              ]][is.infinite(rowSums(object_1[[x]])),1]-
              object_2[[x]][is.infinite(rowSums(object_2[[x
              ]])),1])^q)
69        }
70      }
71    }
72
73    return(total^(1/q))
74  }
75
76  pointDistancePer <- function(points_1, points_2, q) {
77    n_1 <- length(points_1)/2
78    n_2 <- length(points_2)/2
```

```
79
80    if (n_1 + n_2 == 0) {
81      # No points
82      return(0)
83    } else if (n_1 > 0 && n_2 == 0) {
84      # Only points_1 has points
85      if (n_1 == 1){
86        d <- distanceToDiagonal(points_1,q)
87      }else{
88        d <- sum(sapply(
89          1:n_1,
90          function(i) distanceToDiagonal(points_1[i, ], q)
91        ))
92      }
93
94      return(d)
95    } else if (n_1 == 0 && n_2 > 0) {
96      # Only points_2 has points
97      if (n_2 == 1){
98        d <- distanceToDiagonal(points_2,q)
99      }else{
100       d <- sum(sapply(
101         1:n_2,
102         function(i) distanceToDiagonal(points_2[i, ], q)
103       ))
104     }
105
106     return(d)
107   } else {
108     # Hungarian Algorithm
109     cost_matrix <- vector()
110
111     for (i in 1:n_1) {
112       if (n_2==1){
113         if (n_1==1){
114           r_1 <- distanceLq(points_1,points_2,q)
115         }else{
```

```
116             r_1 <- distanceLq(points_1[i, ],points_2,q)
117           }
118         }else{
119           if (n_1==1){
120             r_1 <- apply(points_2, 1, function(x) distanceLq
                  (points_1, x, q))
121           }else{
122             r_1 <- apply(points_2, 1, function(x) distanceLq
                  (points_1[i, ], x, q))
123           }
124         }
125
126         if (n_1 == 1){
127           dist_xy <- distanceToDiagonal(points_1, q)
128         }else{
129           dist_xy <- distanceToDiagonal(points_1[i, ], q)
130         }
131
132         r_2 <- rep(dist_xy, n_1)
133
134         cost_matrix <- rbind(cost_matrix, c(r_1, r_2))
135       }
136       if (n_2==1){
137         r_1 <- distanceToDiagonal(points_2,q)
138       }else{
139         r_1 <- apply(points_2, 1, function(x)
              distanceToDiagonal(x, q))
140       }
141       r_2 <- rep(0, n_1)
142
143       for (j in 1:n_2) {
144         cost_matrix <- rbind(cost_matrix, c(r_1, r_2))
145       }
146
147       pairing <- RcppHungarian::HungarianSolver(cost_matrix)
148       idxs <- pairing[["pairs"]]
149
```

```
150       pair_vals <- apply(
151         idxs, 1,
152         function(x) cost_matrix[x[1], x[2]]
153       )
154 
155       return(sum(pair_vals))
156     }
157 }
158 
159 distanceToDiagonal <- function(p, q) {
160   return(2 * ((abs(p[2] - p[1]) / 2)^q))
161 }
162 
163 distanceLq <- function(p1, p2, q) {
164   return(abs(p1[1] - p2[1])^q + abs(p1[2] - p2[2])^q)
165 }
```

Appendix C

Distance to measure transform

The following Matlab script is used to perform distance to measure transform of the segmented images using a discrete set of distances with threshold measure being 200 white pixels. The translation and rescaling are set so that the filtration will see the vessels last.

```
% disk matrix
imageSizeX = 600;
imageSizeY = 600;
[columnsInImage, rowsInImage] = meshgrid(1:imageSizeX, 1:
    imageSizeY);
% centre of the circle in the image.
centerX = 300;
centerY = 300;
radius = linspace(10,300,256);

[nrow,ncol] = size(file{1});
output_dir = 'results/dist2meas200';
for i = 1:100
    im_i = file{i};
    output_name = [output_dir '/' original_files(i).name];
    dist2meas_i = nan(nrow,ncol);
    for j = 0:255
        % create circle
        circlePixels = (rowsInImage - centerY).^2 ...
        + (columnsInImage - centerX).^2 <= radius(j+1).^2;

```

```
21          % convolution to find measure within circle
22          im_conv_i = conv2(circlePixels,im_i);
23          im_conv_i = im_conv_i(300:(end-300),300:(end-300))
                ;
24          test_i = find(im_conv_i(:)>=200*255);
25          for k = 1:length(test_i)
26              if isnan(dist2meas_i(test_i(k)))
27                  dist2meas_i(test_i(k)) = 255-j;
28              end
29          end
30      end
31
32      dist2meas_i(isnan(dist2meas_i)) = 0;
33
34      imwrite(dist2meas_i/255,output_name);
35
36  end
```

Appendix D

Threshold selection

The following R script is used to simulate threshold distance selection processes.

```
# threshold selection
nSim <- 10
RR_rec <- rep(0,nSim)
cutoff_rec <- rep(0,nSim)
testAcc <- rep(0,nSim)
for(i in 1:nSim){
  # training database
  trainBase <- rep(0,60)
  nInterval <- 1000
  cutoff_dist <- seq(0,500,length.out = nInterval)
  # record recognition rate
  RR_temp <- rep(0,nInterval)
  for(j in 1:20){
    x <- sample(1:5,3,replace = FALSE)
    x <- sort(x)
    trainBase[((j-1)*3+1):(j*3)] <- x + (j-1)*5
  }

  # remaining set
  r <- c(1:100)
  r <- r[!r %in% trainBase]
  # training set
  trainSet <- rep(0,20)
  for (j in 1:20) {
```

```
25      x <- sample(1:2,1)
26      trainSet[j] <- r[(j-1)*2+x]
27    }
28
29    # use training set to find recognition rate
30    distTrain <- matrix(0,nrow = 20, ncol = 60)
31    for (j in 1:60) {
32      distTrain[ ,j] <- full_distance[trainSet,trainBase[j]]
33    }
34
35    for (k in 1:nInterval) {
36      match_i <- distTrain <= cutoff_dist[k]
37      nTP <- 0
38      nTN <- 0
39      for (j in 1:20){
40        trueID <- j
41        pos <- c((trueID-1)*3+1,(trueID-1)*3+2,(trueID-1)*
            3+3)# actual positive
42        neg <- 1:60 # actual negative
43        neg <- neg[!neg %in% pos]
44
45        match_ij_neg <- match_i[j,neg]
46        match_ij_pos <- match_i[j,pos]
47        nTP <- nTP + length(match_ij_pos[match_i[j,pos]==
            TRUE]) # true positive: trueID and accept
48        nTN <- nTN + length(match_ij_neg[match_i[j,neg]==
            FALSE]) # true negative: not trueID and reject
49      }
50      RR_temp[k] <- (nTP+nTN)/(60*20)
51    }
52
53    # pick cutoff with max RR
54    maxPos <- which(RR_temp == max(RR_temp))
55    cutoff_rec[i] <- mean(cutoff_dist[maxPos])
56
57    # testing set
58    testSet <- c(1:100)
```

```
59    testSet <- testSet[!testSet %in% trainBase]
60    testSet <- testSet[!testSet %in% trainSet]
61
62    # distance: dataBase vs testing set
63    test_dist <- matrix(0,nrow = 20, ncol = 20)
64    for (k in 1:20) {
65      for (j in 1:20) {
66        test_dist[k,j] <- mean(full_distance[testSet[k],
            trainBase[((j-1)*3+1):(j*3)]])
67      }
68    }
69
70    match_test <- test_dist<=cutoff_rec[i]
71    nTP <- 0
72    nTN <- 0
73    for (j in 1:20){
74      trueID <- j
75      pos <- trueID # actual positive
76      neg <- 1:20 # actual negative
77      neg <- neg[!neg %in% pos]
78
79      match_ij_neg <- match_test[j,neg]
80      match_ij_pos <- match_test[j,pos]
81      nTP <- nTP + length(match_ij_pos[match_ij_pos==TRUE])
          # true positive: trueID and accept
82      nTN <- nTN + length(match_ij_neg[match_ij_neg==FALSE])
           # true negative: not trueID and reject
83    }
84    RR_rec[i] <- (nTP+nTN)/(20*20)
85  }
86
87  par(mfrow=c(1,1))
88  plot(cutoff_rec,RR_rec,main = "Scatterplot of recognition
      rate vs cutoff distance",
89       xlab = "Cutoff distance", ylab = "Recognition rate",
90       xlim = c(min(cutoff_rec)-5,max(cutoff_rec)+5),ylim=c
           (0.945,0.965),
```

```
91        pch = 19, frame = TRUE)
```

Bibliography

[1] Abdul Salam, A., Akram, U. & Ahmed Khan, S. 2020, "Retina Identification Database (RIDB)", Mendeley Data, V1, DOI: 10.17632/tjw3zwntv6.1.

[2] Applied Algebraic Topology Network 2019, *Katharine Turner (12/3/19): Why should q=p in the Wasserstein distance between persistence diagrams?*, 5 December, Available at: https://www.youtube.com/watch?v=d_iqovKui6k&t=133s (Accessed: 24 August 2022).

[3] Bauer, U. & Lesnick, M. 2015, "Induced matchings and the algebraic stability of persistence barcodes", Journal of Computational Geometry, Issue 6, No. 2, pp. 162-191.

[4] Carlsson, G. 2009, "Topology and data", Bulletin of the American Mathematical Society, No. 46, pp. 255-308, DOI: https://doi.org/10.1090/S0273-0979-09-01249-X.

[5] Cohen-Steiner, D., Edelsbrunner, H. & Harer, J. 2007, "Stability of persistence diagrams", Discrete and Computational Geometry, No. 37, pp. 103-120.

[6] Cohen-Steiner, D., Edelsbrunner, H. & Harer, J. 2009, "Extending persistence using Poincaré and Lefschetz duality", Foundations of Computational Mathematics, Issue 9, No. 1, pp. 79–103.

[7] Culgin, C. 2015, *Six of the most common eye diseases*, 20/20 Onsite, accessed 10 July 2022, https://www.2020onsite.com/blog/six-of-the-most-common-eye-diseases.

[8] Dehghani, A., Ghassabi, Z., Moghddam, H.A., & Moin, M. S. 2013, "Human recognition based on retinal images and using new similarity function", J Image Video Proc 2013, No. 58, https://doi.org/10.1186/1687-5281-2013-58

[9] Delgado-Friedrichs, O. & Robins, V. 2015, Diamorse, [Software], Accessed 16 August 2022.

[10] Edelsbrunner, H., Letscher, D. & Zomorodian, A. 2002, "Topological persistence and simplification", Discrete and Computational Geometry, No. 28, pp. 511-533.

[11] Effron, D., Forcier, B.C., & Wyszynski R.E. 2016, "Chapter 3: Funduscopic findings", Knoop K.J., Stack L.B., Storrow A.B. & Thurman R(Eds.), The Atlas of Emergency Medicine, 4e. McGraw Hill.

[12] Fukuta, K., Nakagawa, T., Hayashi, Y., Hantanaka, Y., Hara, T. & Fujita, H. 2008, *Personal identification based on blood vessels of retinal fundus images*, Proceedings of SPIE - The International Society for Optical Engineering, DOI: 10.1117/12.769330.

[13] Hatcher, A. 2002, *Algebraic topology*, Cambridge University Press, Cambridge.

[14] Hofer, C. D., Kwitt, R. & Niethammer, M. 2019, "Learning representation of persistence barcodes", Journal of Machine Learning Research, No. 20, pp. 1-45.

[15] Jost, J. 1997, *Compact Riemann Surfaces*, Springer-Verlag, ISBN 3-540-53334-6.

[16] Kanari, L., Dłotko, P., Scolarmiero, M., Levi, R., Shillcock, J., Hess, K. & Markram, H. 2017, "A topological representation of branching neuronal morphologies", Neuroinform, No. 16, pp. 3-13, DOI: 10.1007/s12021-017-9341-1.

[17] Morgan, J 2022, XPHT, [Software], Accessed 2 June 2022.

[18] Morozov, D. 2008, *Homological illusions of persistence and stability*, Duke University, North Carolina.

[19] Oudot, S. Y. 2015, "Persistence theory: from quiver representations to data analysis", Mathematical Surveys and Monographs, Vol. 209, ISBN 978-1-4704-2545-6.

[20] Poonen, B. 2014, "Undecidable problems: a sampler", Interpreting Gödel: Critical essays, Cambridge University Press, pp. 211–241, DOI: https://doi.org/10.48550/arXiv.1204.0299.

[21] Rajan, S.P. 2020, "Recognition of cardiovascular diseases through retinal images using optic cup to optic disc ratio", Pattern Recognit. Image Anal. No. 30, pp. 256–263, https://doi.org/10.1134/S105466182002011X.

[22] Robins, V. 2000, *Computational topology at multiple resolutions: foundations and applications to fractals and dynamics*, University of Colorado at Boulder, Colorado.

[23] Robinson, A. & Turner, K. 2013, *Hypothesis testing for topological data analysis*, arXiv:1310.7467 [stat.AP].

[24] Ronneberger, O., Fischer, P. & Brox, T. 2015, "U-Net: convolutional networks for biomedical image segmentation", arXiv:1505.04597 [cs.CV].

[25] Silva, V. D., Morozov, D. & Vejdemo-Johansson, M. 2011, "Dualities in persistent (co)homology", arXiv:1107.5665 [math.AT].

[26] Solomon, E. & Bendich P. 2022, *A convolutional persistence transform*, arXiv:2208.02107 [math.AT].

[27] Turkeš, R., Nys, J., Verdonck, T. & Latré, S. 2021, "Noise robustness of persistent homology on greyscale images, across filtrations and signatures", PLoS ONE, pp. 1-26, https://doi.org/10.1371/journal.pone.0257215.

[28] Turner, K., Mukherjee, S. & Boyer, D. M. 2014, "Persistent homology transform for modelling shapes and surfaces", Information and Inference: A Journal of the IMA, pp. 1-38, DOI:10.1093/imaiai/drn000.

[29] Turner, K., Robins, V & Morgan, J. 2022, *The extended persistent homology transform of two-dimensional binary images*, arXiv:2208.14583 [math.AT].

[30] Waheed, Z., Akram M., Waheed, A., Khan, M., Shaukat, A. & Ishaq, M. 2016, "Personal identification using vascular and non-vascular retinal features", Computers and Electrical Engineering, No. 54, pp. 359-371.

[31] Whitney, H. 1957, *Geometric integration theory*, Princeton University Press, pp. 124–135.

[32] Zomorodian, A., Carlsson G. 2005, "Computing persistent homology", Discrete and Computational Geometry, Vol. 33, No. 2, pp. 249-274.

xt