TOPOLOGICAL CONSISTENCY VIA KERNEL ESTIMATION

By OMER BOBROWSKI*, §, SAYAN MUKHERJEE†, §, and JONATHAN E. TAYLOR‡, ¶

Duke University§, Stanford University¶

We introduce a consistent estimator for the homology (an algebraic structure representing connected components and cycles) of super-level sets of both density and regression functions. Our method is based on kernel estimation. We apply this procedure to two problems: 1) inferring the homology structure of manifolds from noisy observations, 2) inferring the persistent homology (a multi-scale extension of homology) of either density or regression functions. We prove consistency for both of these problems. In addition to the theoretical results we demonstrate these methods on simulated data for binary regression and clustering applications.

1. Introduction. Clustering has been a classic area of research in statistics. Given a set of observations generated by some unknown process in \( \mathbb{R}^d \) (e.g. a probability density function), clustering can be loosely described as inferring the number of components underlying this process (cf. [28, 32, 42, 45]). One standard formulation of the clustering problem is to consider the support of the density, or the super level sets of the density, as a geometric object \( A \subset \mathbb{R}^d \). Clustering then reduces to questions about the connected components of \( A \). Proving consistency for clustering methods has been a topic of interest and research in statistics [6, 27, 38, 46].

A statistical perspective of the recent efforts in topological data analysis (TDA) [13, 22] has been to compute topological invariants from observations generated by a random process. The idea is that these topological summaries are useful for statistical inference. From a topological perspective, clustering can be viewed as a question about the homology of subsets \( A \subset \mathbb{R}^d \) of the...
ambient space. Briefly, homology is a set of algebraic structures, denoted by \( \{H_0(A), H_1(A), \ldots \} \), where \( H_0(A) \) contains information about the connected components of \( A \), and for \( k > 0 \), \( H_k(A) \) contains information about ‘cycles’, or ‘holes’ of different dimensions (see Section 2 for more details). From the perspective of algebraic topology, the clustering problem is thus equivalent to recovering \( H_0(A) \). The idea of characterizing points or subsets of \( \mathbb{R}^d \) by their homology was developed in a series of papers in the late 1990’s [39, 40]. Asymptotic and non-asymptotic analysis of consistency and convergence of topological summaries as the number of observations increase has been examined for a variety of geometric objects using a variety of statistical and probabilistic tools [1, 2, 3, 4, 5, 8, 9, 12, 16, 30, 31, 36, 37].

The objective of this paper is to provide a consistent method for recovering the homology of super level sets of functions \( f: \mathbb{R}^d \rightarrow \mathbb{R} \), defined as

\[
D_L \triangleq \{ x \in \mathbb{R}^d : f(x) \geq L \},
\]

where \( f \) will be either a probability density function generating the data, or a regression function.

We will provide conditions under which one can recover the homology with high confidence. The first step in recovering the homology will be to use a kernel estimator \( \hat{f} \) to approximate the set \( D_L \). We will see in Section 3 that estimates of the homology of \( D_L \) using \( \hat{f} \) are very noisy. We will show that a better idea which results in much more robust estimates is to approximate two “close” sets: \( D_{L+\epsilon} \subset D_{L-\epsilon} \) (for a properly chosen \( \epsilon > 0 \)), and study the map in homology between them. While the homology at a single level is very noisy, the map between them can be robustly estimated and preserves the homology information of \( D_L \). The map serves as a filtering mechanism for the homological noise. While we will be using a kernel-based estimator for \( f \), it is interesting to note that the conditions for homology recovery need not imply consistency of the kernel estimator \( \hat{f} \) itself.

We present two direct applications to recovering the homology of super level sets. The first application is inferring the homology of a manifold from noisy samples. This problem was previously studied in [3, 37]. In this paper we address this inference problem for a broader class of noise models, requiring fewer assumptions than previous methods and analysis. The second application is estimating the persistent homology of the function \( f \). Persistent homology (described in Section 2) is a multi-scale topological summary, where instead of considering the homology of a single level \( D_L \), we look at the entire sequence of super level sets as \( L \) decreases from \( \infty \) to \( -\infty \) and track the changes occurring in homology. The logic behind this computation is that features that exist throughout a long range of levels are considered
to be significant, representing real features of the data, whereas the others can be considered as noise.

The paper is structured as follows. In Section 2 we provide a short introduction to the topological definitions and concepts that will be developed in this paper. Specifically, we provide a brief introduction to homology and persistent homology. In Section 3 we state our main results and delay the proofs to the appendix. In Section 4 we provide a procedure to compute the estimate of the homology of super level sets. We provide some intuition about the estimator as well as our homology inference results using simulation studies in Section 5. We close with a discussion.

2. Topological preliminaries. In this section we introduce the basic ideas of homology and persistent homology we use in this paper. To help fix ideas we present a particular example of persistent homology that is related to agglomerative hierarchical clustering.

2.1. Homology. We wish to introduce the concept of homology here in an intuitive rather than a rigorous way. For a comprehensive introduction to homology, see [29, 35]. Let $X$ be a topological space. The homology of $X$ is a set of abelian groups $\{H_k(X)\}_{k=0}^{\infty}$, called homology groups. In this paper we consider homology with coefficients in a field $F$, in this case $H_k(X)$ is actually a vector space. The zeroth homology group $H_0(X)$ is generated by elements that represent connected components of $X$. For example, if $X$ has three connected components, then $H_0(X) \cong F \oplus F \oplus F$ (here $\cong$ denotes group isomorphism), and each of the three generators of this group corresponds to a different connected component of $X$. For $k \geq 1$, the $k$-th homology group $H_k(X)$ is generated by elements representing $k$-dimensional “holes” or “cycles” in $X$. An intuitive way to think about a $k$-dimensional hole is as the result of taking the boundary of a $(k+1)$-dimensional body. For example, if $X$ a circle then $H_1(X) \cong F$, if $X$ is a two dimensional sphere then $H_2(X) \cong F$, and in general if $X$ is a $n$-dimensional sphere, then

$$H_k(X) \cong \begin{cases} F & k = 0, n \\ \{0\} & \text{otherwise.} \end{cases}$$

Another interesting example is the 2-dimensional torus denoted by $T$ (see Figure 1). The torus has a single connected component, therefore $H_0(T) \cong F$, and a single 2-dimensional hole (the void inside the surface) implying that $H_2(T) \cong F$. As for 1-cycles (or closed loops) the torus has two distinct features (see Figure 1) and therefore $H_1(T) \cong F \oplus F$. 


The ranks of the homology groups (the number of generators) are called the Betti numbers, and are denoted by $\beta_k(X) \triangleq \text{rank}(H_k(X))$. In the following, in cases where we wish to refer to all the homology groups simultaneously, we use the notation $H_*(X)$.

In addition to providing a summary for a single space, homology can also characterize the topological behavior of functions. Let $f : X \to Y$ be a map between two topological spaces, then homology theory provides a way to define the ‘induced map’ $f_* : H_*(X) \to H_*(Y)$ mapping between the homology groups of the two spaces.

Another term we will use is homotopy equivalence (cf. [29, 35]). Loosely speaking, two topological spaces $X, Y$ are homotopy equivalent if we can continuously transform one into the other. We denote this property by $X \simeq Y$. If $X \simeq Y$ then they have the same homology, i.e. $H_*(X) \cong H_*(Y)$.

2.2. Persistent homology. Let $\mathcal{X} = \{X_t\}_{t=a}^{b}$ be a filtration of topological spaces, such that $X_{t_1} \subset X_{t_2}$ if $t_1 < t_2$. As the parameter $t$ increases, the homology of the spaces $X_t$ may change (e.g. components are added and merged, cycles are formed and filled up). The persistent homology of $\mathcal{X}$, denoted by $\text{PH}_*(\mathcal{X})$, keeps track of this process. Briefly, $\text{PH}_*(\mathcal{X})$ contains the information about the homology of the individual spaces $\{X_t\}$ as well as the mappings between the homology of $X_{t_1}$ and $X_{t_2}$ for every $t_1 < t_2$. The birth time of an element in $\text{PH}_*(\mathcal{X})$ can be thought of as the value of $t$ where this element appears for the first time. The death time is the value of $t$ where an element vanishes, or merges with another existing element. We refer the reader to [22, 23, 26, 47] for more details and formal definitions. Another perspective of persistence homology is as a summary statistic of point cloud data that is robust to certain invariances, this perspective has been developed in [7, 11, 33, 44]. A useful way to describe persistent homology is via the notion of barcodes.
A barcode for the persistent homology of a filtration $\mathcal{X}$ is a collection of graphs, one for each order of homology group. A bar in the $k$-th graph, starting at $b$ and ending at $d$ ($b \leq d$) indicates the existence of a generator of $H_k(X_t)$ (or a $k$-cycle) whose birth and death times are $b, d$ respectively. In Figure 2 we present an example for a barcode generated in the following way. We take $n = 50$ samples $P_1, \ldots, P_n \in \mathbb{R}^2$ sampled from a uniform distribution on an annulus. We then define $X_r = \bigcup_i B_r(X_i)$ to be the union of balls around the samples. Increasing $r$ makes the space $X_r$ grow. In this process connected components merge, and cycles are formed and then filled up. In Figure 2(a) we present a few snapshots of the space $X_r$ for different values of $r$ where different feature show. The barcode in Figure 2(b) presents a summary of all the homology features in this process. We can see that there are two bars that are significantly longer than the others (one in $H_0$ and one in $H_1$) indicating that the underlying space has a single connected component, and a single cycle (as the annulus does).

There are many choices of filtrations (sequences of nested spaces). In this paper the filtrations we work with are the super level sets of functions. Specifically, let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and let $D_L \triangleq \{ x \in \mathbb{R}^d : f(x) \geq L \}$ be a super level set of $f$. As the level $L$ is decreased from $\infty$ to $-\infty$ the sets $D_L$ grow, and in this process components and cycles are created and destroyed. We denote by $\text{PH}^* f$ the persistent homology for this process.

To show later that we can recover the persistent homology structure, we will need a notion of distance between the persistent homology of two different filtrations. If $\mathcal{X}$ is a filtration, the persistence diagram of $\mathcal{X}$, denoted by $\text{Dgm}(\mathcal{X})$ is the set of all pairs $(b,d)$ of birth-death times of features in $\text{PH}(\mathcal{X})$. The bottleneck distance between the persistent homology of the filtrations $\mathcal{X}, \mathcal{Y}$ is defined as

$$d_B(\text{PH}(\mathcal{X}), \text{PH}(\mathcal{Y})) = \inf_{\gamma} \sup_{p \in \text{Dgm}(\mathcal{X})} \| p - \gamma(p) \|_\infty.$$ 

The set $\Gamma$ consists all the bijections $\gamma : \text{Dgm}(\mathcal{X}) \cup \text{Diag} \rightarrow \text{Dgm}(\mathcal{Y}) \cup \text{Diag}$, where $\text{Diag} = \{(x,x) : x \in \mathbb{R}\} \subset \mathbb{R}^2$ is the diagonal line, and $\| \cdot \|_\infty$ is the sup-norm in $\mathbb{R}^2$. In other words, we are looking for a matching between the points in $\text{Dgm}(\mathcal{X})$ and $\text{Dgm}(\mathcal{Y})$ that requires the minimal translations of birth and death times. We add the diagonal to each diagram for two reasons. Firstly, we want to be able to consider diagrams with different number of features, and secondly, we want to allow deleting points from a diagram rather than forcing them to match.

To conclude this section, we note that the zeroth persistent homology, $\text{PH}_0$, is closely related to hierarchical clustering as the following example
Fig 2. (a) $X_r$ is a union of balls of radius $r$ around a random set of $n = 50$ points, generated from a uniform distribution on an annulus in $\mathbb{R}^2$. We present five snapshots of this filtration. (b) The persistent homology of the filtration $\{X_r\}_{r=0}^\infty$. The $x$ axis is the radius of the balls, and the bars represent the homology features that are born and died. For $H_0$ we observe that at radius zero the number of components is exactly $n$ and as the radius increases components merge (or die). The cycles show up later in this process. There are two bars that are significantly longer than the others (one in $H_0$ and one in $H_1$). These correspond to the true feature of the annulus.

will illustrate. Let $\mathcal{P} \subset \mathbb{R}^d$ be a finite set of points in Euclidean space. We define the distance function from the set $d_{\mathcal{P}} : \mathbb{R}^d \to \mathbb{R}$ as

$$d_{\mathcal{P}}(x) = \min_{p \in \mathcal{P}} \|x - p\|.$$ 

In this case computing the 0-th persistent homology for the sub level set filtration of $d_{\mathcal{P}}$ is very simple. We start at level 0 with just the finite set $\mathcal{P}$, and as we increase the level we merge connected components according to the distances between points in $\mathcal{P}$. The bottom of Figure 3 is the barcode generated by such a process, the top figure is the dendrogram generated by the same set of points. One can observe that the end points of the bars in the barcode are the nodes in the dendrogram.
FIG 3. Persistent homology and hierarchical clustering. The figure on top is the dendrogram generated by a set of 10 random points in the interval [0, 1]. The bottom figure is the barcode generated by the 0-persistent homology for the sub-level sets of the distance function from the same set of points. The x-axis represents function values (distance, in our case). In this example all the connected components are created at distance zero, and only differ by their death point (when two components merge). Note, that one of the components (the top bar) lives forever. The death points in the barcode correspond to nodes in the dendrogram, we marked the bars with different colors matching the relevant part of the dendrogram.

3. Statistical model and main results. Given a function $f : \mathbb{R}^d \to \mathbb{R}$ the objects we analyze in this paper are the super-level sets of $f$

$$D_L \triangleq \{ x \in \mathbb{R}^d : f(x) \geq L \}. \quad (3.1)$$

Note that for any $L_1 < L_2$ we have $D_{L_2} \subset D_{L_1}$. Our main assumption on $f$ is ‘tameness’ as defined in [19].

**Definition 3.1 ([19]).** Let $f : \mathbb{R}^d \to \mathbb{R}$, and $D_L$ as defined in (3.1).

1. We say that $L$ is a **homological critical value** of $f$ if there exists $k$ such that for any small enough $\epsilon > 0$ the map $H_k(D_{L+\epsilon}) \to H_k(D_{L-\epsilon})$ induced by inclusion is not an isomorphism.
2. A function $f$ is called **tame** if it has a finite number of homological critical values, and $\dim(H_k(D_L))$ is finite for all $L$ and $k$.

Our main goal in this paper is to present a consistent method for recovering the homology of a given super level set $D_L$. We will examine the super
level sets of two classical quantities of interest in statistics:

1. Density functions – Given Data $= \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x)$, where $f$ is a probability density function, our objective is to recover the super-level sets of $f$.

2. Regression functions – Given Data $= \{(X_1, Y_1), \ldots, (X_n, Y_n)\} \overset{iid}{\sim} f_{X,Y}(x, y)$, where $f_{X,Y}(x, y)$ is a joint probability density function, our objective is to recover the super level sets of the regression function $f(x) \triangleq \mathbb{E}\{Y \mid X = x\}$.

A common method to recover the homology of an unknown space $S$ from a set random samples $\mathcal{X} \subset S$ is to compute the homology of a union of balls

$$U(\mathcal{X}, r) \triangleq \bigcup_{X \in \mathcal{X}} B_r(X), \quad (3.2)$$

for some choice of radius $r$ (cf. [9, 36]). We can use this idea to estimate the homology of the set $D_L$ using the following procedure (P1):

1. Use the entire data set to construct an estimator $\hat{f}$.
2. Using the estimator $\hat{f}$, define

$$\mathcal{X}^L = \{X_i : \hat{f}(X_i) \geq L\},$$

as the set of data points lying in the $L$-th super level set of $\hat{f}$.
3. Consider $U(\mathcal{X}^L, r)$ as an estimate of $D_L$, and the homology $H_*(U(\mathcal{X}^L, r))$ as an estimate of $H_*(D_L)$.

We will use kernel estimators for $\hat{f}$ in both the regression and density estimation case. A key difficulty in the above procedure is that the estimator $\hat{f}$ may introduce errors in the filtering step 2 of the above procedure. Since homology is a discrete descriptor, even small errors in the filtering step can introduce large errors in the homology estimates. For example, even a single point incorrectly included in the super level set assignment can form an extra connected component, and increase the zeroth Betti number by one. One of the main challenges we will address in this paper is providing an estimator that is robust to this sensitivity to noise.

Given a kernel function $K : \mathbb{R}^d \rightarrow \mathbb{R}$ we construct our estimators as follows. In the density estimation case we define

$$\hat{f}_n(x) \triangleq \frac{1}{(n - 1) \times C_K r^d} \sum_{i=1}^{n} K_r(x - X_i),$$

$1$ All the results in this paper also hold for $\hat{f}_n(x) = \frac{1}{n \times K_{x\in\mathbb{R}^d}} \sum_{i=1}^{n} K_r(x - X_i)$. 

where $X_1, \ldots, X_n$ are the observed data, $K_r(x) = K(x/r)$, and $C_K$ is a normalizing constant defined below. In the regression setting we use the Nadaraya-Watson estimator \cite{41}

$$\hat{f}_n(x) \triangleq \frac{\sum_{i=1}^n Y_i K_r(x - X_i)}{\sum_{i=1}^n K_r(x - X_i)},$$

where $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ are the observed data.

The kernel functions $K(x)$ we consider satisfy the following conditions (C1):

1. The support of the kernel function is contained within the unit ball of radius 1, i.e. $\text{supp}(K) \subset B_1(0)$;
2. The kernel function has a maximum at the origin, with $K(0) = 1$, and $\forall x : K(x) \in [0, 1]$;
3. The kernel function is smooth within the unit ball, and

$$\int_{\mathbb{R}^d} K(\xi) d\xi = C_K, \quad \text{for } C_K \in (0, 1).$$

To prove consistency of our estimates of the homology of the super level sets of $f$ we will require weak regularity conditions on the density or regression function. For density estimation we require the density function to be tame and bounded, and we define

$$f_{\max} \triangleq \sup_{x \in \mathbb{R}^d} f_X(x).$$

For the regression case we require in addition the following set of conditions (C2):

1. The marginal density of $X$ has compact support, i.e. $\text{supp}(f_X)$ is compact;
2. The marginal density of $X$ is bounded away from zero within its support, i.e. $f_{\min} \triangleq \inf_{x \in \text{supp}(f)} f_X(x) > 0$;
3. The response variables are almost surely bounded, i.e. $|Y_i| \leq Y_{\max}$ almost surely for some non-random value $Y_{\max} > 0$;
4. The conditional variance of the response variables is almost surely bounded, i.e. $\text{Var}(Y_i \mid X_i) \leq V_{\max}$ almost surely for some non-random value $V_{\max} > 0$.

Next, recall step 2 in the procedure (P1), and define

$$X_{n}^L \triangleq \left\{ X_i : \hat{f}_n(X_i) \geq L; \ 1 \leq i \leq n \right\}.$$
We will need to assign some probabilistic quantification of the accuracy of the assignments $\mathcal{X}_n^L$ with respect to $D_L$. We will do this by first defining two sets: the set $\hat{D}_{L,r}$ corresponds to “inflating” $D_L$ by a radius $r$ and $\check{D}_{L,r}$ corresponds to “deflating” $D_L$ by a radius $r$. To define these sets, we first define the tube of radius $r$ around the boundary of $D_L$

$$\partial D_{L,r} = \bigcup_{x \in \partial D_L} B_r(x), \quad \partial D_L \text{ is the boundary of } D_L.$$ 

We then define $\hat{D}_{L,r}$ and $\check{D}_{L,r}$ as follows

$$\hat{D}_{L,r} = D_L \cup \partial D_{L,r}, \quad \check{D}_{L,r} = D_L \setminus \partial D_{L,r}.$$ 

Using these definitions the following Lemma provides a bound on the false positive and false negative error of the set $\mathcal{X}_n^L$ with respect to $D_L$.

**Lemma 3.2.** Assume that constraint (C1) on the kernel function holds and either condition (C2) holds for the regression case or in the density estimation case the density is bounded and tame. For every $L > 0$, and $\epsilon \in (0, L)$, if $r \to 0$ and $nr^d \to \infty$, then there exists a constant $C_\epsilon^*$ such that for $n$ large enough we have

$$P\left( \exists X_i \notin \hat{D}_{L-\epsilon,r} : \hat{f}_n(X_i) \geq L \right) \leq ne^{-C_\epsilon^*nr^d},$$

and

$$P\left( \exists X_i \in \check{D}_{L+\epsilon,r} : \hat{f}_n(X_i) \leq L \right) \leq ne^{-C_\epsilon^*nr^d}.$$

Equation (3.3) bounds the probability of a false-positive error, and equation (3.4) bounds the probability of a false-negative error. The value of $C_\epsilon^*$ is different for density estimation versus regression. For density estimation we use

$$C_\epsilon^* = \frac{3\epsilon^2K\omega_d}{24f_{\text{max}} + 4\epsilon},$$

and for the regression model we use

$$C_\epsilon^* = \frac{3\epsilon^2f_{\text{min}}^2K\omega_d}{24(V_{\text{max}} + \epsilon^2)f_{\text{max}} + 8\epsilon f_{\text{min}}(V_{\text{max}} + \epsilon)},$$

(see appendix for more details).

The bounds in Lemma 3.2 will allow us to recover the homology of $D_L$. Given $\mathcal{X}_n^L$ we want to examine the union of balls

$$U_n(L, r) \triangleq U(\mathcal{X}_n^L, r).$$
We would like to prove that with a high probability the above empirical set is sandwiched by two sets which should be ‘close’ to $D_L$. The following Lemma states the precise result.

**Lemma 3.3.** For every $L > 0$, and $\epsilon \in (0, L)$, if $r \to 0$ and $nr^d \to \infty$, then for large enough $n$ we have

$$\mathbb{P}\left( \tilde{D}_{L+\epsilon,2r} \subset U_n(L,r) \subset \tilde{D}_{L-\epsilon,2r} \right) \geq 1 - 3ne^{-C^*nr^d},$$

In other words, the union of balls $U_n(L,r)$ is sandwiched between the two approximations of $D_L$. Note that the radius $r$ is the same $r$ as used for the bandwidth of the the kernel functions. This connection is crucial for the above approximation to hold.

We now use Lemma 3.3 to state a procedure to recover the homology of $D_L$ that is consistent. First, note that for any $\epsilon \in (0, L)$, we have that $U_n(L + \epsilon, r) \subset U_n(L - \epsilon, r)$. The inclusion map

$$\iota : U_n(L + \epsilon, r) \hookrightarrow U_n(L - \epsilon, r)$$

induces a map in homology

$$\iota_* : H_*(U_n(L + \epsilon, r)) \to H_*(U_n(L - \epsilon, r)).$$

We use this map to define

$$\hat{H}_*(L; \epsilon ; n) \triangleq \text{Im}(\iota_*).$$

We will use $\hat{H}_*(L; \epsilon ; n)$ as an estimator for $H_*(D_L)$. The intuition behind using this inclusion map is as follows. Using Lemma 3.3 we could show that with a high probability we have

$$\mathbb{P}\left( \tilde{D}_{L+\epsilon,2r} \subset U_n(L,r) \subset \tilde{D}_{L-\epsilon,2r} \right) \geq 1 - 3ne^{-C^*nr^d},$$

where $\hookrightarrow$ represents inclusion.

Assuming that $H_*(D_{L+2\epsilon}) \cong H_*(D_L) \cong H_*(D_{L-2\epsilon})$, then all the cycles in $H_*(D_L)$ must persist throughout this entire sequence (3.5) and in particular they should be present in $\hat{H}_*(L, \epsilon ; n)$. In contrast, any cycles in $U_n(L + \epsilon, r)$ that do not belong to $D_L$ must be terminated as we move from $U_n(L + \epsilon, r)$ to $U_n(L - \epsilon, r)$ via $D_L$, and therefore should not be in $\hat{H}_*(L, \epsilon ; n)$. To prove that the inclusion sequence in (3.5) holds, we require the following regularity condition on $L$. 

**TOPOLOGICAL CONSISTENCY**
Definition 3.4. Given a level $L > 0$ and $\epsilon \in (0, L/2)$, we say that $L$ is $\epsilon$-regular if
\[
\partial D_{L+2\epsilon} \cap \partial D_{L+\frac{3}{2}\epsilon} = \partial D_{L+\frac{1}{2}\epsilon} \cap \partial D_{L} = \partial D_{L} \cap \partial D_{L-\frac{1}{2}\epsilon} = \partial D_{L-\frac{3}{2}\epsilon} \cap \partial D_{L-2\epsilon} = \emptyset.
\]

This regularity condition basically guarantees sufficient ‘separation’ between the super level sets. In particular, if $f$ is continuous in $f^{-1}([L-2\epsilon, L+2\epsilon])$, then $L$ is $\epsilon$-regular. We will assume that the levels we are studying are always $\epsilon$-regular. We now state the main result in this paper.

Theorem 3.5. Let $L > 0$ and $\epsilon \in (0, L/2)$ be such that the function $f(x)$ has no critical values in the range $[L-2\epsilon, L+2\epsilon]$. If $r \to 0$, and $nr^d \to \infty$, then for $n$ large enough we have
\[
P(\hat{H}_s(L, \epsilon; n) \cong H_s(D_L)) \geq 1 - 6ne^{-C^*/2nr^d},
\]
In particular, if $nr^d \geq D \log n$ with $D > (C^*/2)^{-1}$, then
\[
\lim_{n \to \infty} P(\hat{H}_s(L, \epsilon; n) \cong H_s(D_L)) = 1.
\]

In other words, Theorem 3.5 states that if we want to recover the homology of the super level set $D_L$ we need to compute the image of the homology map as we move from $U_n(L+\epsilon, r)$ to the slightly larger complex $U_n(L-\epsilon, r)$. We shall describe an algorithm to compute this homology in Section 4.

In the following sections we describe two applications for the estimator we proposed, addressing problems that are of significant interest in the fields of topological data analysis and machine learning.

3.1. An application to manifold learning. Let $\mathcal{M}$ be a smooth $m$-dimensional, closed manifold (compact and without a boundary), embedded in $\mathbb{R}^d$. Given a random sample $X_n = \{X_1, \ldots, X_n\} \subset \mathbb{R}^d$ we wish to recover the homology of $\mathcal{M}$. The case where the observations are drawn directly from the manifold (i.e. $X_n \subset \mathcal{M}$), has been extensively studied (see [9, 36]). In [9] the following asymptotic result was presented.

Theorem 3.6 (Theorem 4.9 in [9]). If $nr^d \geq C \log n$, and $C > (\omega_df_{\min})^{-1}$, then:
\[
\lim_{n \to \infty} P(H_*(U(X_n, r)) \cong H_*(\mathcal{M})) = 1,
\]
where $\omega_m$ is the volume of a $m$-dimensional unit ball, and $f_{\min} = \inf_{x \in \mathcal{M}} f(x) > 0$. 

In this section we wish to extend this result to the case where noise is present. The term ‘noise’ in this context refers to the fact that the samples do not necessarily lie on $\mathcal{M}$, but rather in its vicinity. As an example consider the observations $X_1, \ldots, X_n$ defined as

\begin{equation}
X_i = Y_i + Z_i, \quad \text{where } Y_i \sim \rho(\mathcal{M}), \text{ and } Z_i \sim N(0, \sigma^2 I_d),
\end{equation}

where $Y_i$ is drawn from a distribution $\rho$ that is supported on a manifold $\mathcal{M}$, and $Z_i$ is drawn from the normal distribution in the ambient space $\mathbb{R}^d$. For this model the methods used to prove consistency of the estimator in [9, 36] no longer apply since the outliers produced by the noise create their own topology, and interfere with our ability to recover $H_*(\mathcal{M})$.

The seminal work in [37] studies the following special case. Let $Y_i \sim \rho(\mathcal{M})$, for each $i \in \{1, \ldots, n\}$ let $N_i$ be the normal space to $\mathcal{M}$ at $Y_i$, and let $Z_i \sim N(0, \sigma^2 I_{d-m})$ be a multivariate normal variable in the normal space $N_i$. Our observations are then taken to be $X_i = Y_i + Z_i$. Under explicit assumptions on $\sigma$ and $\mathcal{M}$, they show that the homology of $\mathcal{M}$ can be recovered from $X_n$ with a high probability. The work in [3] extends this idea to a few other noise models. The results and proofs in [3, 37] are tied to specific noise models and rely on the parameters of the noise model and the geometry of $\mathcal{M}$. We wish to use the result in Theorem 3.5 to study the same homology inference problem for a large class of distributions, and with as few assumptions as possible.

We start by defining a general class of density functions on $\mathbb{R}^d$, from which it would be possible to extract the homology of $\mathcal{M}$.

**Definition 3.7.** Let $f : \mathbb{R}^d \to \mathbb{R}_+$ be a probability density function. We say that $f$ represents a noisy version of $\mathcal{M}$, if there exist $0 < A < B < \infty$ such that:

1. For every $L \in [A, B]$ we have $D_L \simeq \mathcal{M}$,
2. For every $L > B$, we have $D_L \simeq \mathcal{M}'$, where $\mathcal{M}' \subset \mathcal{M}$ is a locally contractible proper subset of $\mathcal{M}$,

where ‘$\simeq$’ stands for homotopy equivalence (see Section 2).

In other words, we consider density functions $f$ for which there is a range where the super-level sets are ‘similar’ to $\mathcal{M}$. For levels higher than this range, the super-level sets are ‘similar’ to nice subsets of $\mathcal{M}$. For example, the distribution in (3.6) satisfies this conditions for small enough $\sigma$. By ‘locally contractible’ we refer to the property that every point $x$ has a neighborhood $N_x$ that is homotopy equivalent to a single point. For example, if $\mathcal{M}'$ is a
compact manifold with boundary, then it is locally contractible. We need this requirement to rule out the appearance of highly twisted topological spaces. In Figure 4 we present a sequence of super level sets for a density function that represents a noisy version of the torus. This density was generated by taking a uniform distribution on the latitude angle, a wrapped normal distribution on the longitude angle, and adding independent Gaussian noise. Note that the super level sets are 3-dimensional whereas the torus is 2-dimensional. Nevertheless, we can see that there is a whole range of levels where they are topologically equivalent.

In this figure we demonstrate a sequence of super level sets for a density function $f$ that is a noisy version of the 2-dimensional torus. The horizontal axis represents the function levels in a decreasing order. For very high values ($L > B$) we see that the super level sets look like a subset of the torus. Note that they are not real subsets, since these are 3-dimensional shapes, whereas the torus is 2-dimensional. Inside the range $(A, B)$ the super level sets look exactly like the torus (where $\beta_0 = \beta_2 = 1$, and $\beta_1 = 2$). For low levels the topology changes again, but we no longer require any assumptions.

If we know a-priori the values of $A$ and $B$, then the recovery method would be simple. Given a sample $X_n = \{X_1, \ldots, X_n\} \sim f$, we choose $L$ and $\epsilon$ such that $[L - 2\epsilon, L + 2\epsilon] \subset (A, B)$, and compute $H_\ast(L, \epsilon; n)$. Theorem 3.5 guarantees that with high probability $\hat{H}_\ast(L, \epsilon; n) \cong H_\ast(D_L) \cong H_\ast(M)$.

However, in real problems we are not given $A, B$ so the real challenge is to recover $M$ without knowing the stable range. To show that the procedure described below is consistent, we require the following assumptions to hold.

(i) $M$ is connected and orientable;
(ii) $B - A > 8\epsilon$;
(iii) There exists a known value $L_{\text{max}}$, such that $f \leq L_{\text{max}}$.

The following procedure will be used to estimate the homology of $M$ from the noisy samples $X_n$. In this procedure, we will use the estimated Betti numbers defined as $\hat{\beta}_k(L, \epsilon; n) \triangleq \text{rank}(\hat{H}_\ast(L, \epsilon; n))$. The following procedure (P2) gives us an estimate of the homology of the manifold from the the noisy samples $X_n$. 

\begin{align*}
\beta_0 = 0 & \quad \beta_2 = 0 \\
\beta_0 = 1 & \quad \beta_2 = 1 \\
\beta_0 = 1 & \quad \beta_2 = 0
\end{align*}
1. Set a fine enough grid on the level set values

\[ L_{\text{max}} = L_1 > L_2 > \cdots > L_M = 0, \]

such that \( L_i - L_{i+1} = 2\epsilon \) for all \( 1 \leq i \leq M - 1 \).
2. Compute \( \hat{H}_s(L_i, \epsilon; n) \) for all \( i = 1, \ldots, M \).
3. Define

\[ i^\star \triangleq 1 + \min \left\{ i \in \{1, \ldots, M\} : \hat{\beta}_m(L_i, \epsilon; n) = 1 \right\}, \]

to be the point where we expect to observe the homology of \( \mathcal{M} \) for the first time.
4. Our estimator for the homology of \( \mathcal{M} \) will then be \( \hat{H}_s(L^\star_i, \epsilon; n) \).

Given this procedure, the following Theorem states that we can estimate the homology for a manifold from noisy observations.

**Theorem 3.8.** Let \( \mathcal{M} \) be a \( m \)-dimensional closed, connected, orientable manifold embedded in \( \mathbb{R}^d \). Let \( X_1, \ldots, X_n \) be data points sampled from a density function \( f \) satisfying the conditions in Definition 3.7. Applying procedure (P2) we have

\[ \lim_{n \to \infty} \Pr \left( \hat{H}_s(L^\star_i, \epsilon; n) = H_s(\mathcal{M}) \right) = 1. \]

We state here the main ideas used in proving the above, while the detailed proof is given in the appendix. We use Poincaré duality, a fundamental idea in algebraic topology. Poincaré duality relates homology groups to co-homology groups of closed orientable \( m \)-dimensional manifolds, stating that \( H_k(\mathcal{M}) \cong H^{m-k}(\mathcal{M}) \), where \( H^{m-k}(\mathcal{M}) \) is the co-homology of \( \mathcal{M} \) (cf. \([29, 35]\)). An important consequence of Poincaré duality is that \( \beta_k(\mathcal{M}) = \beta_{m-k}(\mathcal{M}) \) for every \( k = 0, \ldots, m \), and in particular \( \beta_0(\mathcal{M}) = \beta_m(\mathcal{M}) \). Our assumption that \( \mathcal{M} \) is connected implies that \( \beta_0(\mathcal{M}) = 1 \), and from Poincaré duality we conclude that \( \beta_m(\mathcal{M}) = 1 \) as well. In contrast, if \( \mathcal{M}' \subset \mathcal{M} \) is a proper locally contractible subset of \( \mathcal{M} \) then using a different type of duality one can show that \( \beta_m(\mathcal{M}') = 0 \) (see Proposition 3.46 in \([29]\)). Our assumptions on \( A, B \) then imply that if \( L_i > B \) we have \( \beta_m(D_{L_i}) = 0 \), while if \( L_i \in (A, B) \) then \( \beta_m(D_{L_i}) = 1 \). Therefore, the first \( L_i \) for which the \( m \)-th Betti number switches from 0 to 1 necessarily lies in \( (A, B) \), and we can use this \( L_i \) to recover the homology of \( \mathcal{M} \). In practice, we defined \( i^\star \) to be the second level at which we have \( \hat{\beta}_m(L_i, \epsilon; n) = 1 \). This is a precautionary measure which we discuss in the proof.
3.2. An application to clustering and persistent homology. A common topological summary used in TDA is persistent homology (see Section 2). If the function \( f \) was known we would be interested in computing the persistent homology of \( f \), \( \text{PH}_*(f) \), as a summary of \( f \). This summary contains information about the creation and destruction of connected components and cycles of the super level sets. By its definition, \( \text{PH}_*(f) \) is computed from the continuous filtration \( D = \{ D_L \}_{L \in \mathbb{R}} \) as \( L \) decreases from \( \infty \) to \( -\infty \).

Note that the persistent homology \( \text{PH}_*(f) \) contains much more information than just the homology at each level \( D_L \). It also contains information about mappings between different levels, and hence enables us to track the evolution of cycles. We showed that the estimator \( \hat{H}_*(L, \epsilon, n) \) can recover the homology of \( D_L \) for every \( L \), but in order to recover the persistent homology we also need to make sure that the mappings between different levels are recovered as well. The error measure we use is the commonly used ‘bottleneck distance’ (see Section 2). To estimate \( \text{PH}_*(f) \) we will consider the following discrete filtration

\[
\hat{D}_\epsilon = \{ U_n(L_i, r) \}_{i \in \mathbb{Z}},
\]

where \( L_i = f_{\text{max}} - 2i\epsilon \). Denoting the persistent homology of \( \hat{D}_\epsilon \) by \( \hat{\text{PH}}_*(f) \), and using the methods presented in this section we prove the following.

**Theorem 3.9.** If \( r \to 0 \) and \( nr^d \to \infty \), then

\[
P \left( d_B \left( \hat{\text{PH}}_*(f), \text{PH}_*(f) \right) \leq 5\epsilon \right) \geq 1 - 3Mn e^{-C_{\epsilon/2}nr^d},
\]

where \( M \geq \lceil f_{\text{max}}/2\epsilon \rceil \). In particular, if \( nr^d \geq D \log n \) with \( D > (C_{\epsilon/2})^{-1} \), we have

\[
\lim_{n \to \infty} P \left( d_B \left( \hat{\text{PH}}_*(f), \text{PH}_*(f) \right) \leq 5\epsilon \right) = 1.
\]

In other words, we state that the estimator \( \hat{\text{PH}}_*(f) \) is ‘consistent’ up to a given precision - \( 5\epsilon \). Note that we should not expect to have no error at all, since our estimator is discrete by nature (having an inherent step size \( \epsilon \)) while the filtration we wish to study is continuous. However, we can take \( \epsilon \) as small as we want achieving a higher precision. The smaller value of \( \epsilon \) we choose the smaller \( C_{\epsilon/2} \) will be, and the convergence of \( \hat{\text{PH}}_*(f) \) to \( \text{PH}_*(f) \) will be slower.

To prove this theorem (see appendix), we invoke Lemma 3.3 \( M \) times in order to form a sequence of inclusions alternating between super level sets \( D_L \) and union of balls \( U_n(L, r) \). This alternating sequence is called ‘interleaving’ and the work in [15] provides means to bound the distance between the persistent homology computed for these two types of filtrations. In Section
we provide several examples for the estimation of persistent homology using $\hat{PH}_s(f)$.

As we discuss in Section 4, Theorem 3.9 can be adjusted to use the filtration of Rips complexes $\{R_n(L_i, r)\}_{i \in \mathbb{Z}}$ instead of $\{U_n(L_i, r)\}_{i \in \mathbb{Z}}$. The work in [14, 17] studies a different method to recover the persistent homology of $f$ using Rips complexes. In order to recover $PH_s(f)$, [14] considers the maps $\iota^L_* : H_s(R_n(L, r)) \to H_s(R_n(L, 2r))$ induced by inclusion for all values of $L$ and for a fixed $r$. The persistence module for the family of images $\{\text{Im}(\iota^L_*)\}^L_{\mathbb{Z}}$ is then used as an approximation for $PH_s(f)$. In a way, one can think of the transition $R_n(L, r) \to R_n(L, 2r)$ as playing the same role as the transition $R_n(L + \epsilon, r) \to R_n(L - \epsilon, r)$ we study in this paper, “filtering” the noisy homology. Changing the radius rather than the level, allows one to avoid the level discretization that our method relies on, which leads to a more accurate approximation. On the other hand, this method requires further assumptions on the model parameters, and computing the estimator is more complicated. It remains future work to study whether these two methods could be combined into a more powerful and robust one.

In a different line of previous work [12, 18, 24] persistent homology is recovered by constructing a kernel-based estimator $\hat{f}$ for the function at hand and then computing the persistent homology of the estimator $PH(\hat{f})$. The validity of these methods is established by using the stability theorem [19] stating that $d_B(PH_s(f), PH_s(\hat{f})) \leq ||f - \hat{f}||_\infty$. There are two significant advantages to the estimator we suggest here. Firstly, we do not require assumptions about the global sup-norm convergence of the estimator. Secondly, computing the estimator $PH(\hat{f})$ in practice involves discretizing the space, and this may have a significant effect on the ability to recover small features in the data (see for example, the clustering examples in Section 5). The estimator we propose does not require such a discretization.

4. Computing the homology estimator. The estimator we propose in Section 3 requires the computation of the image between the homology groups of $U_n(L + \epsilon, r)$ and $U_n(L - \epsilon, r)$. As a review for a more statistical audience, we provide the fundamentals tools required to compute this estimator. In general, algorithms for computing homology of unions of balls require two steps. The first step is to obtain a combinatorial representation of the geometric object that is either equivalent in homology or approximately equivalent in homology to the original geometric object. This step is outlined in subsection 4.1. The combinatorial representation reduces the homology computation to a linear algebra problem. The second step is to apply a set of linear transformations to this combinatorial representation to
compute the image of the homology groups under the inclusion map between two complexes. This step is outlined in subsection 4.2.

4.1. The Čech and Vietoris-Rips complex. Let $\mathcal{X} = \{x_1, \ldots, x_n\}$ be a set of points in $\mathbb{R}^d$, and suppose that we wish to compute the homology of $U(\mathcal{X}, r)$ for some $r > 0$. The Čech complex is a simplicial complex that allows us to convert the homology computation problem into linear algebra. The Rips complex can be thought of as an approximation to the Čech complex. This approximation offers computational advantages over the Čech complex but suffers from not sharing the same direct relation to the homology of $U(\mathcal{X}, r)$ as the Čech complex. We first provide the definitions for these complexes.

**Definition 4.1** (Čech complex). Let $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ be a collection of points in $\mathbb{R}^d$, and let $r > 0$. The Čech complex $C(\mathcal{X}, r)$ is constructed as follows:

1. The 0-simplices (vertices) are the points in $\mathcal{X}$.
2. A $k$-simplex $[x_{i_0}, \ldots, x_{i_k}]$ is in $C(\mathcal{X}, r)$ if $\bigcap_{j=0}^{k} B_r(x_{i_j}) \neq \emptyset$.

**Definition 4.2** (Rips complex). Let $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ be a collection of points in $\mathbb{R}^d$, and let $r > 0$. The Rips complex $R(\mathcal{X}, r)$ is constructed as follows:

1. The 0-simplices (vertices) are the points in $\mathcal{X}$.
2. A $k$-simplex $[x_{i_0}, \ldots, x_{i_k}]$ is in $R(\mathcal{X}, r)$ if $\|x_{i_j} - x_{i_l}\| \leq 2r$ for all $0 \leq j, l \leq k$.

Figure 5 depicts a simple example of a Čech and Rips complex in $\mathbb{R}^2$. The figure also highlights the contrast between the two complexes. The main difference is that the Rips complex is constructed simply from pairwise intersection information while the Čech complex requires high-order information. This difference is realized in Figure 5 in the far left triangle in either complex. In the Rips complex the left triangle is filled in to be a face, since all three pairwise intersections occur. In the Čech complex higher-order interactions are also computed, in this case one observes that the three pairwise intersections do not overlap resulting in three edges rather than a filled in face. The main advantage of the Rips complex is computational – all we need in order to construct the Rips complex is to compute the pairwise distances between all the points, rather than to check for all possible orders of intersections of balls as we would have to for the Čech complex.
The Rips complex can be considered as an approximation to the Čech complex. It is clear from the definitions that $C(\mathcal{X}, r) \subset R(\mathcal{X}, r)$. In addition, it is shown in [20] that $R(\mathcal{X}, r) \subset C(\mathcal{X}, \sqrt{2}r)$. Combining these two statements we have that

$$R(\mathcal{X}, r) \subset C(\mathcal{X}, \sqrt{2}r) \subset R(\mathcal{X}, \sqrt{2}r).$$

Fig 5. On the left - the Čech complex $C(\mathcal{X}, r)$, on the right - the Rips complex $R(\mathcal{X}, r)$ with the same set of vertices and the same radius. We see that the three left-most balls do not have a common intersection and therefore do not generate a 2-dimensional face in the Čech complex. However, since all the pairwise intersections occur, the Rips complex does include the corresponding face.

An important result in algebraic topology called the ‘Nerve Lemma’ (cf. [10]) states that the Čech complex $C(\mathcal{X}, r)$ is homotopy equivalent to the neighborhood set $U(\mathcal{X}, r)$. In particular it follows $H_*(C(\mathcal{X}, r)) \cong H_*(U(\mathcal{X}, r))$. As a consequence, any statement made about the homology of $U(\mathcal{X}, r)$ applies to $C(\mathcal{X}, r)$ and vice versa.

Denote the Čech complex generated by the filtered point set $\mathcal{X}_n^L$ as $C_n(L, r) \triangleq C(\mathcal{X}_n^L, r)$. We can then define

$$i_\ast : H_*(C_n(L + \epsilon, r)) \rightarrow H_*(C_n(L - \epsilon, r))$$

to be the map induced by the inclusion map between the simplicial complexes. Defining

$$\hat{H}_n^C(L, \epsilon; n) \triangleq \text{Im}(i_\ast),$$

then by the Nerve Lemma, since $U_n(L \pm \epsilon, r)$ and $C_n(L \pm \epsilon, r)$ are completely equivalent structures, Theorem 3.5 holds without changes for $\hat{H}_n^C(L, \epsilon; n)$.

Next, we denote the Rips complex constructed form the filtered samples as $R_n(L, r) \triangleq R(\mathcal{X}_n^L, r)$ and define the following inclusion map for any $\epsilon \in (0, L/2)$

$$i : R_n(L + \epsilon, r) \hookrightarrow R_n(L - \epsilon, r).$$

This inclusion induces a map in homology

$$i_\ast : H_*(R_n(L + \epsilon, r)) \rightarrow H_*(R_n(L - \epsilon, r)).$$
and we denote
\[ \hat{H}^R_\ast(L, \epsilon; n) \triangleq \text{Im}(i_\ast). \]

Note that the Nerve Lemma applies only to the Čech complex and not the Rips. Nevertheless, the following theorem states that we can compute the homology of \( D_L \) using the Rips complex as well. The importance of providing a consistent estimator for \( H_\ast(D_L) \) that uses the Rips complex is due to its computational efficiency.

**Theorem 4.3.** Let \( L > 0 \) and \( \epsilon \in (0, L/2) \) be such that the function \( f(x) \) has no critical values in the range \([L - 2\epsilon, L + 2\epsilon]\). If \( nr^d \to \infty \), then for \( n \) large enough we have
\[
\mathbb{P}\left( \hat{H}^R_\ast(L, \epsilon; n) \cong H_\ast(D_L) \right) \geq 1 - 6ne^{-C\epsilon/nr^d},
\]
In particular, if \( nr^d \geq D \log n \) with \( D > (C\epsilon/2)^{-1} \), then
\[
\lim_{n \to \infty} \mathbb{P}\left( \hat{H}^R_\ast(L, \epsilon; n) \cong H_\ast(D_L) \right) = 1.
\]

In the next subsection we provide an algorithm for computing the image of the inclusion map using either the Čech or Rips complex.

**4.2. Computing the homology of the image.** Our estimator for \( H_k(D_L) \) requires the computation of the image of the map between the homology of two nested simplicial complexes \( \Delta(1) \subset \Delta(2) \) (either Čech or Rips). This map is denoted by \( i_k : H_k(\Delta(1)) \to H_k(\Delta(2)) \). In this section we present an algebraic algorithm to compute the rank of this image, namely the estimated Betti number \( \beta_k \). Note that there are several efficient algorithms to compute persistent homology that can also be used here (see [23, 34, 43]). We present a relatively simple algorithm, in the interest of clarity for a statistical audience, for the case where \( F \) is a field of characteristic zero (e.g. \( \mathbb{R}, \mathbb{Q} \)). For a fixed homology degree \( 0 \leq k \leq d \) the algorithm would consist of two steps:

1. Finding a basis for the kernel of a squared matrix defined later as \( L_k^{(1)} \);
2. Computing the rank of two matrices, defined later as \( \hat{\partial}_k^{(2)} \) and \( \hat{\partial}_k^{(2)} \), and then we will have that
\[
\text{rank}(\text{Im}(i_k)) = \text{rank}(\hat{\partial}_k^{(2)}) - \text{rank}(\partial_k^{(2)}).
\]

In the following, we provide more details about homology computation for simplicial complexes, and in particular the definitions of the matrices \( L_k^{(1)} \), \( \partial_k^{(2)} \), and \( \hat{\partial}_k^{(2)} \) mentioned above.
4.2.1. **Computing the homology of a simplicial complex.** Let $\Delta$ be a simplicial complex, let $\Delta_k$ be the set of $k$-simplexes in $\Delta$, and let $n_k = |\Delta_k|$, so we can write

$$\Delta_k = \{\sigma_1, \sigma_2, \ldots, \sigma_{n_k}\}.$$ 

We assume that every $k$-simplex $\sigma_i \in \Delta_k$ is attached with a unique orientation (an ordering on its set of vertices), denoted by $\sigma_i = [x^i_0, \ldots, x^i_k]$. Defining $C_k \triangleq \mathbb{F}^{n_k}$, we wish to map the simplexes of $\Delta_k$ into a basis of $C_k$ in a way that preserves orientation information. To do that we define the map $T_k : \Delta_k \to C_k$ in the following way. For every simplex $\sigma_i$ we define $T(\sigma_i) = e_i$, where $e_i$ consists of one at the $i$-th entry, and zero elsewhere. For every permutation $\pi$ on $0, \ldots, k$ we then define

$$T_k([x^i_{\pi(0)}, \ldots, x^i_{\pi(k)}]) = \text{sign}(\pi)e_i,$$

where $\text{sign}(\pi) = (-1)^{P(\pi)}$, and $P(\pi)$ is the parity of the permutation $\pi$. The vector space $C_k$ is usually referred to as the ‘space of $k$-chains’ of $\Delta$.

Next, using the map $T_k$, we define the matrix $\partial_k$ to be a $n_k - 1 \times n_k$ matrix where the $i$-th column is given by

$$(\partial_k)_i = \sum_{\sigma \text{ is a face of } \sigma_i} T_{k-1}(\sigma).$$

In other words, the nonzero entries in the $i$-th column correspond to the $(k-1)$-dimensional faces of $\sigma_i \in \Delta_k$. The matrix $\partial_k$ can be thought of as a linear transformation from $C_k$ to $C_{k-1}$ and is referred to as ‘the boundary operator’. The $k$-th homology of $\Delta$ is then defined to be the quotient space given by

$$(4.1) \quad H_k(\Delta) \triangleq \ker(\partial_k)/\text{Im}(\partial_{k+1}).$$

One way to find a basis for $H_k(\Delta)$ is via the combinatorial Laplacian, defined as the following $n_k \times n_k$ matrix

$$L_k \triangleq \partial_{k+1}\partial_k^T + \partial_k^T \partial_k.$$ 

Note that $L_0$ is the well known graph Laplacian. If $\mathbb{F}$ is a field with characteristic zero (e.g. $\mathbb{R}, \mathbb{Q}$) then it is shown in [25] that the kernel of $L_k$ is isomorphic to $H_k(\Delta)$ and in particular, the Betti numbers of $\Delta$ are given by $\beta_k(\Delta) = \dim(\ker(L_k))$. 
4.2.2. The homology of the map. Our goal is not only to compute the homology of $\Delta^{(1)}$ and $\Delta^{(2)}$ separately, but rather to compute the image of the map $\iota_k : H_k(\Delta^{(1)}) \rightarrow H_k(\Delta^{(2)})$. For $j = 1, 2$ let $\Delta_k^{(j)}$ be the set of $k$-simplexes in $\Delta^{(j)}$, and let $n_k^{(j)} = |\Delta_k^{(j)}|$. Since $\Delta^{(1)} \subset \Delta^{(2)}$ we can list the simplexes in the following way:

$$\Delta_k^{(1)} = \{ \sigma_1, \sigma_2, \ldots, \sigma_{n_k^{(1)}} \},$$
$$\Delta_k^{(2)} = \{ \sigma_1, \sigma_2, \ldots, \sigma_{n_k^{(1)}}, \sigma_{n_k^{(1)}+1}, \ldots, \sigma_{n_k^{(2)}} \}.$$  

Using this ordering on the simplexes, we define the boundary operators $\partial_k^{(j)}$ and the combinatorial Laplacians $L_k^{(j)}$ for each of the complexes. It is then easy to see that

$$\partial_k^{(2)} = \begin{pmatrix} \partial_k^{(1)} & \cdots \\ 0 & \ddots \end{pmatrix}.$$  

Now, if $\{v_1, \ldots, v_m\} \subset C_k^{(1)}$ is a basis for $\ker(L_k^{(1)})$ then it represents a basis for $H_k(\Delta^{(1)})$, such that $\beta_k(\Delta^{(1)}) = m$. Let $\hat{v}_i \in C_k^{(2)}$ be a zero padded version of $v_i \in C_k^{(1)}$. From from (4.1) we know that $v_i \in \ker(\partial_k^{(1)})$, and thus from (4.2) it is clear that $\hat{v}_i \in \ker(\partial_k^{(2)})$ as well. This implies that the vectors in $\{\hat{v}_1, \ldots, \hat{v}_m\}$ are candidates to form a basis for $\text{Im}(\iota_k)$. Note, however, that while $\hat{v}_i \in \ker(\partial_k^{(2)})$, it is possible that some linear combinations of $\hat{v}_1, \ldots, \hat{v}_m$ are in $\text{Im}(\partial_{k+1}^{(2)})$, which means that they are considered as trivial in $H_k(\Delta^{(2)})$. This means that $\{\hat{v}_1, \ldots, \hat{v}_m\}$ might be larger than a basis for $\text{Im}(\iota_k)$, and we need to reduce this set. This can be done by solving several sets of linear equation, which we avoid describing here. However, the rank of $\text{Im}(\iota_k)$ can be computed easily by

$$\text{rank}(\text{Im}(\iota_k)) = \text{rank}(\partial_{k+1}^{(2)}) - \text{rank}(\partial_{k+1}^{(2)}) - \text{rank}(\partial_{k+1}^{(2)}),$$

where

$$\partial_{k+1}^{(2)} = (\partial_{k+1}^{(2)}, \hat{v}_1, \ldots, \hat{v}_m)$$

is a $n_k^{(2)} \times (n_{k+1}^{(2)} + m)$ matrix we get by concatenating the boundary matrix $\partial_{k+1}^{(2)}$ with the column vectors $\hat{v}_i$. In other words, we measure how many vectors from the set $\{\hat{v}_1, \ldots, \hat{v}_m\}$ can be added to the set of columns vectors of $\partial_{k+1}^{(2)}$ without generating linear dependency.
5. Results on simulated data. In this section we illustrate how we can use the methods in Section 3 for data analysis using some simulated examples. The examples we chose relate to classical problems in statistics: classification, non-parametric regression, and clustering. We use these examples to demonstrate the novelty and strength of the methods proposed in this paper.

Binary regression. We illustrate how we can recover the homology of a classification function. The marginal density of the explanatory variables is uniform in the unit square $X \sim U\left([\frac{-1}{2}, \frac{1}{2}]^2\right)$. We then set the conditional probability of the response $Y$ as

\begin{equation}
    P(Y = 1 \mid X = x) = f(x) \triangleq C(1 + \sin(4\pi \|x\|^2))e^{-100(\|x\| - 1/4)^2},
\end{equation}

where $C$ is a normalization factor guaranteeing that $f(x)$ is indeed a conditional probability. The graph of this conditional probability is given in Figure 6.

We generate i.i.d. samples $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ from the joint distribution and our objective is to recover the topology of the super level set $D_L$ for $L = 0.5$ which is used as the binary classifier in this case, and has the shape of an annulus. We use the Rips construction presented in Theorem 4.3, with $\epsilon = 0.2$. This gives us two complexes: $S_1 = R(0.3, r)$ and $S_2 = R(0.7, r)$. Figure 7 shows the sets of disks used to create the two Rips complexes. The light blue disks are the ones corresponding to $S_1$ and the orange ones corresponds to $S_2$. Computing the Betti numbers yields:
Indeed, while the homology of each of the complexes $S_1, S_2$ is extremely noisy, the image of the map between them looks exactly like an annulus.

|   | $S_1$ | $S_2$ | $S_1 \hookrightarrow S_2$ |
|---|---|---|---|
| $\beta_0$ | 34 | 53 | 1 |
| $\beta_1$ | 23 | 49 | 1 |

**Figure 7.** Computing the homology of a super level set for a regression function. We generated $\{(X_i, Y_i)\}_{i=1}^{50000}$ i.i.d. samples from the marginal and conditional distributions given in equation (5.1). For $L = 0.5$ and $\epsilon = 0.2$ we present the following: (a) the set $U_n(L + \epsilon, r)$, (b) the set $U_n(L - \epsilon, r)$, (c) the two sets combined. Note that both individual sets in (a) and (b) contain many connected components and cycles. However, in (c) we observe that most of these homological features do not survive the transition. All the extra connected components in (a) are merged into the large component in (b). Similarly, all the extra cycles in (a) are filled up in (b).

**Kernel regression.** In this example we consider a regression function on the unit square $f : [-1, 1]^2 \rightarrow \mathbb{R}$ with an additive noise of the form

$$Y_i = f(X_i) + \xi_i.$$ 

Our objective will be to recover the barcode or persistent homology of the above function from noisy observations.

The regression function $f$ was generated from a random mixture of Gaussians, and its graph is presented in Figure 8(a). The independent variables $X_i$ are generated from a uniform distribution in the box $[-1, 1]^2$. The noise $\xi_i$ is independent of $X_i$, and generated by a normal distribution with $\sigma = 0.2$ truncated at $5\sigma$ (we require in (C2) for the response variables to be bounded). The (true) barcode of the function $f$ is presented in Figure 9(a). To estimate this barcode we used $\hat{\text{PH}}(f)$ with $\epsilon = 0.001$ and $n = 5000$ samples, the result is presented in Figure 9(b).

**Spectral clustering.** Spectral clustering uses spectral graph theory to cluster observations. We revisit a simulated example from the spectral clustering literature to illustrate how well we can recover the number of clusters
A regression function in $\mathbb{R}^2$. (a) The graph of the function in the box $[-1,1]^2$. (b) The level sets of the function. It is easy to spot five peaks and three valleys in this image, which in persistent homology correspond to five features in $\text{PH}_0$ and three in $\text{PH}_1$. (c) Generating $\{(X_i,Y_i)\}_{i=1}^{5,000}$ i.i.d. samples from the model presented in (5.2).

The barcode of the persistent homology of the regression function $f$ presented in Figure 8. (b) The estimated persistent homology $\hat{\text{PH}}_\epsilon(f)$ is very close to the true barcode. For visualization purposes we left bars with length less than 0.05 out of the figure. In both the true and the estimated barcodes we observe five significant features in $H_0$ and three in $H_1$, corresponding to the five peaks and three valleys in the graph of the function $f$.

and cluster features using our level sets approach. We generate $n = 10000$ samples from three concentric circles (of radii 1, 2, 3) and added multivariate Gaussian noise with $\sigma = 0.2$. The result is presented in Figure 10(a). The topological features we wish to recover here are the three connected components and the three cycles (spectral clustering would find the three connected components). Figure 10(b) displays $\hat{\text{PH}}_0(f)$. Here we see that there are indeed three dominating features (bars that persist over a long period of time). The rest of the features are generated by the fluctuations in
the estimated density function. Similarly, in Figure 10(c) we observe three dominating features as well, representing the three cycles in the data.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig10.png}
\caption{(a) A sample set generated from three concentric circles. (b) The barcode for $\hat{\text{PH}}_{\epsilon}(f)$, where we indeed observe three dominating components. (c) The barcode for $\text{PH}_{\epsilon}(f)$, where we indeed observe three dominating cycles.}
\end{figure}

Hierarchical clustering. This example will be used to show how using our method we can capture features of a density function with hierarchical structure. Consider a probability density $f$ on $\mathbb{R}^2$ that consists of two concentrated densities that are far apart and centered at $(\pm 0.25, 0)$, see Figure 11(a). Once we zoom into the two densities we realize there is a finer structure in this problem. The density around $(0.25, 0)$ is a mixture of four Gaussians that are very near each other, see Figure 11(b). The density around $(-0.25, 0)$ is one density that looks like a volcano crater, see Figure 11(c). The result of this finer structure is that when we examine the persistence homology of $f$ we expect to see: (1) five dominating features in $\text{PH}_0$ - the four bumps on the right, and the entire volcano on the left, (2) two dominating features in $\text{PH}_1$ - one coming from the cycle along the rim of the volcano, and another one from the cycle that surrounds the four bumps, (3) fluctuations on the rim will introduce features in $\text{PH}_0(f)$ but these will have low persistence. We will show how we can accurately capture the homology of this hierarchical structure.

The barcode in Figure 12(a) displays $\text{PH}_*(f)$ (the true persistent homology of the function). Looking at the barcode of $\text{PH}_0(f)$, we see two dominant features, with death time close to zero. These two features correspond to the two clusters represented by the peaks seen in Figure 11(a). The other three dominant features correspond to the three additional peaks we have in Figure 11(b). The rest of the bars (as well as other shorter bars we kept out of the figure for visualization purposes) correspond to the fluctuation along
the rim of the crater in Figure 11(c). In $\text{PH}_1(f)$ we see exactly two features corresponding to the two cycles described above.

We can compare the true barcode to the barcode generated by our estimator for $\text{PH}_*(f)$ using $\hat{\text{PH}}_*(f)$. The parameters we used in the estimator are $r = 0.001, \epsilon = 3.5$. The barcode for $\hat{\text{PH}}_*(f)$ is presented in Figure 12(b). The global picture is very consistent with that of the true function. As expected our estimates have extra variation in the endpoints of the bars.

![Figure 11](image)

**Figure 11.** A hierarchical density function (a) The density function at a coarse level, consisting of two sharp peaks (b) Zooming in on the density around $(0.25, 0)$ we observe that this sharp peak actually consists of four adjacent peaks. (c) Zooming in on the density around $(-0.25, 0)$ we observe that the peak has a crater-like structure with small fluctuation around the rim. (d)-(f) Samples generated by the $f$.

In Fasy et. al. [24] an alternate approach is developed to estimate $\text{PH}_*(f)$. Their idea is to use a kernel density estimation to obtain an estimate $\hat{f}_n$ of the density $f$. Then they compute the persistent homology of $\hat{f}_n$, denoted by $\text{PH}_*(\hat{f}_n)$. They are able to provide a theoretical bound on the bottleneck distance between $\text{PH}_*(f)$ and $\text{PH}_*(\hat{f}_n)$. This result is similar in spirit to Theorem 3.9 in our paper. The main difference in their method versus our method is that they focus on getting a good estimate of the function
Fig 12. Estimating the persistent homology of the density function $f$ presented in Figure 11. (a) The ‘true’ barcode for the function $f$, i.e. $\text{PH}_*(f)$. (b) The barcode computed from the estimator $\hat{\text{PH}}_*(f)$. (c) The barcode computed for the kernel density estimator $-\text{PH}_*(\hat{f})$. Note that the estimator $\hat{\text{PH}}_*(f)$ gives a result that is very similar to the true barcode. In both cases there are five significant features in $H_0$ and two significant features in $H_1$. The barcode for $\text{PH}_*(\hat{f})$ only recover the coarse features, namely the two clusters, but completely ignores the finer structures. We note that for visualization purposes we filtered out the very small bars before drawing the barcodes here.

values or ensuring $\hat{f}_n \approx f(x)$ everywhere, whereas we compute $\hat{\text{PH}}_*(f)$ by approximating the super-level sets directly.

In the case of a density function with hierarchical structure these two approaches often have different empirical performance. In particular we argue that the estimator $\hat{\text{PH}}_*(f)$ is favorable to $\text{PH}_*(\hat{f})$. The crux of the argument in favor of computing $\hat{\text{PH}}_*(f)$ is that in evaluating the fit of $\hat{f}$ there is a resolution parameter of how fine in $\mathbb{R}^2$ one measures $f$, which we denote as $\Delta$ (in addition to the bandwidth parameter of the kernel - $r$). The problem arises in that one needs to know what value of $\Delta$ is small enough to capture fine structure in $f$. This raises two issues: 1) how to adaptively estimate $\Delta$ from data and 2) taking a finer resolution parameter will result in an increase in the sample complexity of the inference problem. Our approach of directly estimating $\hat{\text{PH}}_*(f)$ avoids these difficulties, since we only work with the original set of sample points rather than $\hat{f}$.

6. Conclusion. In this paper we introduce a consistent estimator for the homology of super-level sets for both density and regression functions. We apply this procedure to infer the homology of a manifold from noisy observations, and infer the persistent homology of either density or regression functions. The conditions we require are weaker than previous results in this direction.

We view this work as an important step in closing the gap between topological data analysis and statistics. For topological data analysis, we provide a consistent estimator for the homology and persistent homology of spaces.
underlying random data. As future work, we will consider refinements of our analysis to obtain convergence rates and confidence intervals of the estimates. We suspect this will require more assumptions on the geometry of the underlying spaces. From a statistical perspective this work suggests that topological summaries of density and regression functions are of interest and provide insights in statistical modeling. We suspect these characteristics or topological summaries will be very useful in classification or hypothesis testing problems, when the assumptions on different decision regions can be naturally captured by coarse geometry or topology.

Acknowledgements. The authors would like to thank: Robert Adler, Paul Bendich, Ezra Miller and Andrew Nobel for many useful discussions. We also wish to thank Frédéric Chazal for very useful comments on the first revision of this paper.

REFERENCES

[1] Robert J. Adler, Omer Bobrowski, Matthew S. Borman, Eliran Subag, and Shmuel Weinberger. Persistent homology for random fields and complexes. In Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown, pages 124–143. Institute of Mathematical Statistics, 2010.

[2] Robert J. Adler, Omer Bobrowski, and Shmuel Weinberger. Crackle: The homology of noise. Discrete & Computational Geometry, 2014.

[3] Sivaraman Balakrishnan, Alessandro Rinaldo, Don Sheehy, Aarti Singh, and Larry Wasserman. Minimax rates for homology inference. arXiv preprint arXiv:1112.5627, 2011.

[4] Sivaraman Balakrishnan, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. Tight lower bounds for homology inference. arXiv:1307.7666 [cs, math, stat], July 2013.

[5] Paul Bendich, Bei Wang, and Sayan Mukherjee. Local homology transfer and stratification learning. In Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, page 1355–1370, 2012.

[6] G. Biau, L. Devroye, and G. Lugosi. On the performance of clustering in hilbert spaces. IEEE Trans. Info. Theory, 54(2):781–790, 2008.

[7] Andrew J. Blumberg, Itamar Gal, Michael A. Mandell, and Matthew Pancia. Persistent homology for metric measure spaces, and robust statistics for hypothesis testing and confidence intervals. arXiv:1206.4581, 2012.

[8] Omer Bobrowski and Robert J. Adler. Distance functions, critical points, and the topology of random Čech complexes. arXiv preprint arXiv:1107.4775, 2014.

[9] Omer Bobrowski and Sayan Mukherjee. The topology of probability distributions on manifolds. Probability Theory and Related Fields, 2014.

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

[11] Peter Bubenik. Statistical topology using persistence landscapes. arXiv:1207.6437, July, 2012.
(12) Peter Bubenik, Gunnar Carlsson, Peter T. Kim, and Zhiming Luo. Statistical topology via morse theory, persistence and nonparametric estimation. *Algebr. Methods Stat. Probab. II*, 516:75, 2010.
(13) G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.
(14) Frédéric Chazal, Leonidas J. Guibas, Steve Y. Oudot, and Primož Skraba. Scalar field analysis over point cloud data. *Discrete & Computational Geometry*, 46(4):743–775, 2011.
(15) Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, page 237–246. ACM, 2009.
(16) Frédéric Chazal, David Cohen-Steiner, and André Lieutier. A sampling theory for compact sets in euclidean space. *Discrete & Computational Geometry*, 41(3):461–479, 2009.
(17) Frédéric Chazal, Leonidas J. Guibas, Steve Y. Oudot, and Primož Skraba. Persistence-based clustering in riemannian manifolds. *Journal of the ACM (JACM)*, 60(6):41, 2013.
(18) Moo K. Chung, Peter Bubenik, and Peter T. Kim. Persistence diagrams of cortical surface data. In *Information Processing in Medical Imaging*, page 386–397, 2009.
(19) David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.
(20) Vin de Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7(339-358):24, 2007.
(21) R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2nd edition, 2002.
(22) Herbert Edelsbrunner and John Harer. Persistent homology - a survey. *Contemporary Mathematics*, 453:257–282, 2008.
(23) Herbert Edelsbrunner and John L. Harer. *Computational topology: an introduction*. AMS Bookstore, 2010.
(24) Brittany Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Statistical inference for persistent homology. *arXiv preprint arXiv:1303.7117*, 2013.
(25) Joel Friedman. Computing betti numbers via combinatorial laplacians. *Algorithmica*, 21(4):331–346, 1998.
(26) Robert Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.
(27) J. A. Hartigan. Asymptotic distributions for clustering criteria. *Annals of Statistics*, 7(1):117–131, 1978.
(28) John A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 99th edition, 1975.
(29) Allen Hatcher. *Algebraic topology*. Cambridge University Press, 2002.
(30) Matthew Kahle. Random geometric complexes. *Discrete & Computational Geometry*, 45(3):553–573, 2011.
(31) Matthew Kahle and Elizabeth Meckes. Limit the theorems for betti numbers of random simplicial complexes. *Homology, Homotopy and Applications*, 15(1):343–374, 2013.
[32] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.

[33] Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, 2011.

[34] Konstantin Mischaikow and Vidit Nanda. Morse theory for filtrations and efficient computation of persistent homology. *Discrete & Computational Geometry*, 50(2):330–353, 2013.

[35] James R. Munkres. *Elements of algebraic topology*, volume 2. Addison-Wesley Reading, 1984.

[36] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.

[37] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. A topological view of unsupervised learning from noisy data. *SIAM Journal on Computing*, 40(3):646–663, 2011.

[38] D. Pollard. Strong consistency of $k$-means clustering. *Annals of Statistics*, 9(1):135–140, 1981.

[39] V. Robbins, J.D. Meiss, and E. Bradley. Computing connectedness: An exercise in computational topology. *Nonlinearity*, 11:913–922, 1998.

[40] V. Robbins, J.D. Meiss, and E. Bradley. Computing connectedness: Disconnectedness and discreteness. *Physica D*, 139:276–300, 2000.

[41] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986.

[42] P.H.A. Sneath and R.R. Sokal. *Numerical Taxonomy. The Principles and Practice of Numerical Classification*. Freeman, 1973.

[43] Andrew Tausz, Mikael Vejdemo-Johansson, and Henry Adams. Javaplex: A research software package for persistent (co)homology. Software, 2011.

[44] Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *Discrete and Computational Geometry*, 2014.

[45] R.C. Tyron. *Cluster Analysis*. Edwards Bros., Oxford, UK, 1939.

[46] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *Annals of Statistics*, 36(2):555–586, 2008.

[47] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.
APPENDIX A: PROOFS

In this section we provide the proofs for Lemmas 3.2 and 3.3, as well as
Theorems 3.5, 3.8, 3.9, and 4.3.

One of the main probability tools we use is Bernstein’s inequality [21],
basically a law of large numbers bound. If \( Z_1, \ldots, Z_n \) are i.i.d., with \( \mathbb{E} \{ Z_i \} = 0 \), \( \text{Var}(Z_i) = \sigma^2 \) such that \( |Z_i| \leq M \) almost surely, then

\[
\mathbb{P} \left( \sum_{i=1}^{n} Z_i \geq t \right) \leq \exp \left( -\frac{\frac{1}{2}t^2}{n\sigma^2 + \frac{1}{3}Mt} \right)
\]

(A.1)

Proof of Lemma 3.2 - Density estimation. To reconstruct the su-
per level sets of the density, we will use kernel density estimator. Recall that
the kernel function \( K : \mathbb{R}^d \to \mathbb{R} \) we use satisfies the following:

- \( \text{supp}(K) \subset B_1(0) \),
- \( K(x) \in [0, 1] \), and \( K(0) = 1 \),
- \( \int K(\xi)d\xi = C_K \), for some \( C_K \in (0, 1) \).

In this case, our kernel estimator is

\[
\hat{f}_n(x) = \frac{\sum_{i=1}^{n} K_r(x - X_i)}{(n - 1)C_Kr^d},
\]

where \( K_r(x) = K(x/r) \). We start by proving (3.3). Using a simple union
bound we have

\[
\mathbb{P} \left( \exists i \in \hat{D}_{L-\epsilon,r} : \hat{f}_n(X_i) \geq L \right) \leq n\mathbb{P} \left( X_1 \in (\hat{D}_{L-\epsilon,r})^c : \hat{f}_n(X_1) \geq L \right)
\]

(A.2)

\[
= n \int_{(\hat{D}_{L-\epsilon,r})^c} f_X(x)\mathbb{P} \left( \hat{f}_n(X_1) \geq L \mid X_1 = x \right) dx.
\]

Next,

\[
\mathbb{P} \left( \hat{f}_n(X_1) \geq L \mid X_1 = x \right) \leq \mathbb{P} \left( K_r(0) + \sum_{i=2}^{n} K_r(x - X_i) \geq L(n-1)C_Kr^d \right)
\]

(A.3)

\[
= \mathbb{P} \left( \sum_{i=2}^{n} Z_i \geq (n-1)(LC_Kr^d - p_r(x)) - 1 \right)
\]

where

\[
p_r(x) \triangleq \mathbb{E} \{ K_r(x - X_i) \},
\]
and $Z_i = K_r(x - X_i) - p_r(x)$ are independent variables with $\mathbb{E} \{ Z_i \} = 0$. Note that $p_r(x) \in [0, 1]$ since $K_r(x) \in [0, 1]$. Also, since $x \in (D_{L-\epsilon,r})^c$, we have that

\begin{equation}
(A.4) \quad p_r(x) = \int_{B_r(x)} f(\xi)K_r(x - \xi)d\xi \leq (L - \epsilon)C_K r^d,
\end{equation}

and therefore from (A.3) we have,

\begin{equation}
(A.5) \quad \mathbb{P} \left( \hat{f}_n(X_1) \geq L \mid X_1 = x \right) \leq \mathbb{P} \left( \sum_{i=2}^{n} Z_i \geq (n - 1)\epsilon C_K r^d - 1 \right).
\end{equation}

We would like to apply the inequality in (A.1) for $t = (n - 1)\epsilon C_K r^d - 1$. Note that $|Z_i| \leq 1$, and also that

$$\text{Var} (Z_i) \leq \mathbb{E} \{ K_r^2(x - X_i) \} \leq f_{\text{max}} C_K r^d.$$ 

Therefore, we have

\begin{equation*}
\mathbb{P} \left( \hat{f}_n(X_1) \geq L \mid X_1 = x \right) \leq \exp \left( -\frac{1}{3} t^2 \left( \frac{1}{2} t^{-1} (n - 1) f_{\text{max}} C_K r^d + \frac{1}{3} \right) \right).
\end{equation*}

Since $nr^d \to \infty$, for $n$ large enough we have that $t > \frac{1}{2} n \epsilon C_K r^d$, and therefore we have

\begin{equation*}
\mathbb{P} \left( \hat{f}_n(X_1) \geq L \mid X_1 = x \right) \leq \exp \left( -\frac{3}{4} C_{\epsilon}^* \frac{n \epsilon C_K r^d}{24 f_{\text{max}} + 4 \epsilon} \right).
\end{equation*}

Thus, we have

\begin{equation*}
\mathbb{P} \left( \hat{f}_n(X_1) \geq L \mid X_1 = x \right) \leq e^{-C_{\epsilon}^* \frac{n \epsilon r^d}{24 f_{\text{max}} + 4 \epsilon}},
\end{equation*}

with

\begin{equation}
(A.6) \quad C_{\epsilon}^* = \frac{3 \epsilon^2 C_K}{24 f_{\text{max}} + 4 \epsilon}.
\end{equation}

Which completes the proof of (3.3)

To prove (3.4) we start the same way, and similarly to (A.5) we have,

\begin{equation*}
\mathbb{P} \left( \hat{f}_n(X_1) \leq L \mid X_1 = x \right) \leq \mathbb{P} \left( \sum_{i=2}^{n} Z_i \leq -(n - 1)\epsilon C_K r^d - 1 \right),
\end{equation*}
where we used the fact that we consider only \( x \in \hat{D}_{L+\epsilon,r} \), and therefore we have \( p_r(x) \geq (L + \epsilon)C_K r^d \). Thus, to complete the proof we should use (A.1) for the variables \((-Z_i)\) and \( t = (n - 1)\epsilon C_K r^d + 1 \). Similarly to the proof above, we then have that

\[
\P\left( \hat{f}_n(X_1) \leq L \mid X_1 = x \right) \leq e^{-C^*_n \epsilon r^d},
\]

which completes the proof.

\[\square\]

**Proof of Lemma 3.2 - Kernel regression.** Recall that in the kernel regression model, we have a set of pairs \((X_1, Y_1), \ldots, (X_n, Y_n)\), where the pairs are i.i.d., \( X_i \in \mathbb{R}^d \), \( Y_i \in \mathbb{R} \), and they have a common density function \( f_{X,Y} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R} \). Our estimation target is the conditional expectation

\[
f(x) = \mathbb{E} \{ Y \mid X = x \}.
\]

The estimator we use is given by

\[
\hat{f}_n(x) = \frac{\sum_{i=1}^n Y_i K_r(x - X_i)}{\sum_{i=1}^n K_r(x - X_i)},
\]

where the assumptions on \( K_r \) are the same as above. In addition we have the following assumptions:

- \( f_X \) has a compact support - \( \text{supp}(f) \).
- \( f_{\min} \triangleq \inf_{x \in \text{supp}(f)} f_X(x) > 0 \).
- \( |Y_i| \leq Y_{\max} \) almost surely, for some non-random value \( Y_{\max} > 0 \).
- \( \text{Var}(Y_i|X_i) \leq V_{\max} \) almost surely, for some non-random value \( V_{\max} > 0 \).

We start by proving (3.3). We use the union bound again to have

(A.7)

\[
\P\left( \exists X_i \notin \hat{D}_{L-\epsilon,r} : \hat{f}_n(X_i) \geq L \right) \\
\leq n \int_{(\hat{D}_{L-\epsilon,r})^c} \int_{\mathbb{R}} f_{X,Y}(x,y) \P\left( \hat{f}_n(X_1) \geq L \mid X_1 = x, Y_1 = y \right) dy dx.
\]

Note that writing \( \hat{f}_n(x) \geq L \) is equivalent to

\[
\sum_{i=1}^n Y_i K_r(x - X_i) \geq \sum_{i=1}^n L K_r(x - X_i).
\]
Therefore,

\[
P \left( \hat{f}_n(X_1) \geq L \mid X_1 = x, Y_1 = y \right) \\
= P \left( y + \sum_{i=2}^{n} Y_i K_r(x - X_i) \geq L + \sum_{i=2}^{n} L K_r(x - X_i) \right) \\
= P \left( \sum_{i=2}^{n} (Y_i - f(X_i)) K_r(x - X_i) \geq \sum_{i=2}^{n} (L - f(X_i)) K_r(x - X_i) + L - y \right) \\
\leq P \left( \sum_{i=2}^{n} (Y_i - f(X_i)) K_r(x - X_i) \geq \epsilon \sum_{i=2}^{n} K_r(x - X_i) + L - y \right)
\]

where the inequality is due to the facts that \( x \in (\hat{D}_L - \epsilon, r) \), and \( K_r(x - X_i) = 0 \) if \( X_i \notin B_r(x) \). Recall that \( p_r(x) = \mathbb{E} \{ K_r(x - X_i) \} \), and define

\[ Z_i \triangleq (Y_i - f(X_i)) K_r(x - X_i) - \epsilon (K_r(x - X_i) - p_r(x)) \]

Then \( \mathbb{E} \{ Z_i \} = 0 \), and we have

\[
P \left( \hat{f}_n(X_1) \geq L \mid X_1 = x, Y_1 = y \right) \leq P \left( \sum_{i=2}^{n} Z_i \geq \epsilon (n - 1) p_r(x) + L - y \right)
\]

We would like to use Bernstein’s inequality to bound this probability. First, denote

\[
Z_i^{(1)} = (Y_i - f(X_i)) K_r(x - X_i), \\
Z_i^{(2)} = \epsilon (K_r(x - X_i) - p_r(x))
\]

Then it is easy to show that \( \mathbb{E} \{ Z_i^{(1)} \} = \mathbb{E} \{ Z_i^{(2)} \} = \mathbb{E} \{ Z_i^{(1)} Z_i^{(2)} \} = 0 \), which implies that \( Z_i^{(1)} \) and \( Z_i^{(2)} \) are uncorrelated, and therefore

\[
\sigma^2 = \text{Var} (Z_i) = \text{Var} (Z_i^{(1)}) + \text{Var} (Z_i^{(2)}).
\]

Also, it is easy to show that

\[
\text{Var}(Z_i^{(1)}) = \mathbb{E} \{ \text{Var} (Y_i | X_i) \ K_r^2(x - X_i) \}.
\]

Therefore, we have

- \( \text{Var}(Z_i^{(1)}) \leq V_{\text{max}} \mathbb{E} \{ K_r^2(x - X_i) \} \leq V_{\text{max}} C_K f_{\text{max}} r^d \),
- \( \text{Var}(Z_i^{(2)}) \leq \epsilon^2 \mathbb{E} \{ K_r^2(x - X_i) \} \leq \epsilon^2 C_K f_{\text{max}} r^d \).
and almost surely:
\[ |Z_i| \leq |Y_i| + |f(X_i)| + \epsilon(1 + p_r(x)) \leq 2Y_{\text{max}} + \epsilon(1 + C_K f_{\text{max}} r^d) < 2(Y_{\text{max}} + \epsilon), \]

Using Bernstein’s inequality (A.1), for \( t = \epsilon(n - 1)p_r(x) + L - y, \) we have
\[
P \left( \hat{f}_n(X_1) \geq L \mid X_1 = x, Y_1 = y \right) \leq \exp \left( -\frac{\frac{1}{2}t^2}{V_{\text{max}} + \epsilon^2 C_K f_{\text{max}} (n - 1)r^d + \frac{2}{3}(Y_{\text{max}} + \epsilon)t} \right)
\] \[= \exp \left( -\frac{1}{2t} \right) \left( \frac{1}{t^{-1}(V_{\text{max}} + \epsilon^2 C_K f_{\text{max}} (n - 1)r^d + \frac{2}{3}(Y_{\text{max}} + \epsilon))} \right). \]

Using the facts that \( p_r(x) \geq f_{\text{min}} C_K r^d, \) and \( nr^d \to \infty, \) we have that for \( n \) large enough \( t > \frac{\epsilon}{2} f_{\text{min}} C_K n r^d, \) and
\[
P \left( \hat{f}_n(X_1) \geq L \mid X_1 = x, Y_1 = y \right) \leq \exp \left( -\frac{\frac{T}{2f_{\text{min}} C_K n r^d}}{2(V_{\text{max}} + \epsilon^2 f_{\text{max}}) + \frac{2}{3}(Y_{\text{max}} + \epsilon)} \right)
\] \[= \exp \left( -\frac{3\epsilon^2 f_{\text{min}}^2 C_K n r^d}{24(V_{\text{max}} + \epsilon^2 f_{\text{max}} + 8\epsilon f_{\text{min}}(Y_{\text{max}} + \epsilon))} \right).
\]

Thus, we established that
\[
P \left( \hat{f}_n(X_1) \geq L \mid X_1 = x, Y_1 = y \right) \leq e^{-C^*_\epsilon n r^d},
\]

with
\[
(A.8) \quad C^*_\epsilon = \frac{3\epsilon^2 f_{\text{min}}^2 C_K}{24(V_{\text{max}} + \epsilon^2 f_{\text{max}} + 8\epsilon f_{\text{min}}(Y_{\text{max}} + \epsilon))}.
\]

Putting this back into (A.7) completes the proof of (3.3). The proof of (3.3) is similar, with some adjustments. We review it briefly.

\[
P \left( \exists X_i \in \hat{D}_{L+\epsilon, r} : \hat{f}_n(X_i) \leq L \right) \leq n \int_{\hat{D}_{L+\epsilon, r}} \int_{\mathbb{R}} f_{X,Y}(x,y) P \left( \hat{f}_n(X_1) \leq L \mid X_1 = x, Y_1 = y \right) dy dx.
\]

(A.9)
Now,
\[
\mathbb{P}\left( \hat{f}_n(X_1) \leq L \mid X_1 = x, Y_1 = y \right) \\
= \mathbb{P}\left( y + \sum_{i=2}^{n} Y_i K_r(x - X_i) \leq L + \sum_{i=2}^{n} L K_r(x - X_i) \right) \\
= \mathbb{P}\left( \sum_{i=2}^{n} (f(X_i) - Y_i) K_r(x - X_i) \geq \sum_{i=2}^{n} (f(X_i) - L) K_r(x - X_i) - L + y \right) \\
\leq \mathbb{P}\left( \sum_{i=2}^{n} (f(X_i) - Y_i) K_r(x - X_i) \geq \epsilon \sum_{i=2}^{n} K_r(x - X_i) - L + y \right)
\]
where the inequality is due to the facts that \( x \in \hat{D}_{L+\epsilon,r} \), and \( K_r(x - X_i) = 0 \) if \( X_i \not\in B_r(x) \). Here we define.
\[
Z_i = (f(X_i) - Y_i) K_r(x - X_i) - \epsilon (K_r(x - X_i) - p_r(x)).
\]
Then we have
\[
\mathbb{P}\left( \hat{f}_n(X_1) \leq L \mid X_1 = x, Y_1 = y \right) \leq \mathbb{P}\left( \sum_{i=2}^{n} Z_i \geq \epsilon (n - 1) p_r(x) - L + y \right).
\]
We would like to use Bernstein’s inequality to bound this probability. Similarly to the above, we have that for small enough \( \epsilon \),
\begin{itemize}
  \item \( \text{Var} (Z_i) \leq (V_{\text{max}} + \epsilon^2) C_{K f_{\text{max}}} r^d \), and
  \item \( |Z_i| \leq 2(Y_{\text{max}} + \epsilon) \), almost surely.
\end{itemize}
Using Bernstein’s inequality (A.1) for \( t = \epsilon (n - 1) p_r(x) - L + y \), in a similar way as the proof above, completes the proof for (3.4).

We now prove Lemma 3.3. As a first step we prove the following.

**Lemma A.1.** If \( nr^d \to \infty \), then
\[
\mathbb{P}\left( \bar{D}_{L+\epsilon,2r} \not\subseteq U_n(L,r) \right) \leq 2n e^{-C^*_r n r^d},
\]
where \( C^*_r \) is the same as in Lemma 3.2.

**Proof.** Note that in both cases (density estimation and kernel regression) we have that the set \( \bar{D}_{L+\epsilon,2r} \) is compact. Let \( \delta \in (0,1) \), and let \( S \) be a finite set of points satisfying that for every \( x \in \bar{D}_{L+\epsilon,2r} \) there exists \( s \in S \).
such that $\|x-s\| \leq \delta r$. Then there exists a constant $c > 0$ such that we can construct $S$ with $|S| \leq c(\delta r)^{-d}$ points. Note that if there is $x \in \hat{D}_{L+\epsilon,2r}$ that is not covered by the balls of radius $r$, it necessarily means that there is $s \in S$ that is not covered by the balls of radius $(1-\delta)r$. Therefore,

$$\mathbb{P}(\hat{D}_{L+\epsilon,2r} \not\subset U_n(L, r)) \leq \mathbb{P}(\exists s \in S : B_{(1-\delta)r}(s) \cap \mathcal{X}_n = \emptyset)$$

$$= \mathbb{P}(\exists s \in S : B_{(1-\delta)r}(s) \cap \mathcal{X}_n = \emptyset; \hat{D}_{L+\epsilon,r} \cap \mathcal{X}_n \subset \mathcal{X}_n^L)$$

$$+ \mathbb{P}(\exists s \in S : B_{(1-\delta)r}(s) \cap \mathcal{X}_n = \emptyset; \hat{D}_{L+\epsilon,r} \cap \mathcal{X}_n \not\subset \mathcal{X}_n^L)$$

$$\leq \mathbb{P}(\exists s \in S : B_{(1-\delta)r}(s) \cap \mathcal{X}_n = \emptyset) + \mathbb{P}(\hat{D}_{L+\epsilon,r} \cap \mathcal{X}_n \not\subset \mathcal{X}_n^L).$$

In other words, the event of not covering $\hat{D}_{L+\epsilon,2r}$ might occur for two different reasons. Either the original sample set (before filtering) $\mathcal{X}_n$ does not cover $\hat{D}_{L+\epsilon,2r}$ (the first term), or our filtering method got rid of too many samples (second term). The second term can be bounded using Lemma 3.2. For the first term we have

$$\mathbb{P}(\exists s \in S : B_{(1-\delta)r}(s) \cap \mathcal{X}_n = \emptyset) \leq \sum_{s \in S} \mathbb{P}(B_{(1-\delta)r}(s) \cap \mathcal{X}_n = \emptyset)$$

$$= \sum_{s \in S} (1 - F(B_{(1-\delta)r}(s)))^n$$

$$\leq \sum_{s \in S} e^{-nF(B_{(1-\delta)r}(s))},$$

where $F(A) = \int_A f_X(x)dx$. For the density estimation, $s \in \hat{D}_{L+\epsilon,2r}$ implies that

$$F(B_{(1-\delta)r}(s)) \geq (L + \epsilon)^d \omega_d r^d \geq (1-\delta)^d \omega_d.$$ 

For the kernel regression model, we have that

$$F(B_{(1-\delta)r}(s)) \geq f_{\min}(1-\delta)^d \omega_d r^d.$$

Thus, if we choose $C_1 = c\delta^{-d}$, and

$$C_2 = \begin{cases} 
L(1-\delta)^d \omega_d & \text{density estimation,} \\
 f_{\min}(1-\delta)^d \omega_d & \text{kernel regression,}
\end{cases}$$

we have that

$$\mathbb{P}(\exists s \in S : B_{(1-\delta)r}(s) \cap \mathcal{X}_n = \emptyset) \leq C_1 r^{-d} e^{-C_2 nr^d}.$$ 

From Lemma 3.2 we know that

$$\mathbb{P}(\hat{D}_{L+\epsilon,r} \cap \mathcal{X}_n \not\subset \mathcal{X}_n^L) \leq ne^{-C_3 nr^d}.$$
Note that for both models we have that $C_\epsilon^* < C_2$ (see (A.6), (A.8)), and also that $r^{-d} = o(n)$. Therefore the latter probability is necessarily the dominant one in the bound we have. This completes the proof.

**Proof of Lemma 3.3.** If $nr^d \to \infty$, then by Lemma A.1 we have

(A.10) \[ \mathbb{P} (\hat{D}_{L+2r} \not\subset U_n(L,r)) \leq 2ne^{-C_\epsilon^*nr^d}. \]

In addition, from Lemma 3.2 we have

(A.11) \[ \mathbb{P} (U_n(L,r) \not\subset \hat{D}_{L-2r}) \leq \mathbb{P} (X^L_n \cap (\hat{D}_{L-2r})^c \neq \emptyset) \leq ne^{-C_\epsilon^*nr^d}, \]

Using the union bound completes the proof.

We now prove Theorem 3.5. We start with the following purely algebraic lemma.

**Lemma A.2.** Consider the following commutative diagram of groups,

\[
\begin{array}{ccc}
G_1 & \xrightarrow{g_{12}} & G_2 \\
\downarrow{g_{13}} & & \downarrow{g_{24}} \\
G_3 & \xrightarrow{g_{34}} & G_4 \\
\downarrow{g_{35}} & & \downarrow{g_{45}} \\
G_5 & & \\
\end{array}
\]

(by ‘commutative’ we mean that all paths with the same endpoints lead to the same result), and for every $i,j$ define $G_{ij} = \text{Im}(g_{ij}) \subset G_j$.

If $G_3 \cong G_{15}$, then $G_{24} \cong G_3$. More specifically, the map $g_{34} : G_3 \to G_4$ is an isomorphism between $G_3$ and $G_{24}$.

**Proof.** First, we want to prove that the assumption $G_3 \cong G_{15}$ implies that $g_{35} : G_3 \to G_5$ is an isomorphism between $G_3$ and $G_{15}$. Let $h : G_{15} \to G_3$ be an isomorphism, and let $p : G_5 \to G_{15}$ be the projection on $G_{15}$. Consider the sequence,

\[ G_3 \xrightarrow{g_{35}} G_{35} \xrightarrow{p} G_{15} \xrightarrow{h} G_3. \]

Since $G_{15} \subset G_{35}$ we have that $p \circ g_{35}$ is surjective, and also $h$ is surjective (as an isomorphism). Thus, $h \circ p \circ g_{35}$ is surjective, and therefore an automorphism. This, combined with the fact that $g_{35}$ is onto $G_{35}$, implies that $p : G_{35} \to G_{15}$ is injective, and therefore that $G_{15} = G_{35}$ and that $p$ is the identity. In other words, we have that $g_{35}$ is an isomorphism between $G_3$ and $G_{15}$ as desired.
Next, we need to show that \( \text{Im}(g_{35}) \subset G_{24} \). Let \( \sigma_3 \in G_3 \), and \( \sigma_4 = g_{34}(\sigma_3) \in G_4 \). If \( \sigma_5 = g_{35}(\sigma_4) = g_{35}(\sigma_3) \), then necessarily \( \sigma_5 \in G_{15} \) (since we just proved that \( G_{35} = G_{15} \)). Therefore, there exists \( \sigma_1 \in G_1 \) such that \( g_{15}(\sigma_1) = \sigma_5 \). Denoting by \( \bar{\sigma}_3 = g_{13}(\sigma_1) \) we have that \( g_{35}(\bar{\sigma}_3) = \sigma_5 \) and since \( g_{35} \) is an isomorphism we have that \( \bar{\sigma}_3 = \sigma_3 \). Therefore \( \sigma_4 = g_{14}(\sigma_1) \), and this proves that \( \sigma_4 \in G_{24} \). This establishes that we can write \( g_{34} : G_3 \to G_{24} \).

Finally, we want to show that this map is an isomorphism.

To prove that \( g_{34} \) is injective, let \( \sigma_3 \in G_3 \) such that \( \sigma_3 \neq 0 \). Since \( g_{35} \) is injective we have that \( g_{35}(\sigma_3) \neq 0 \) which implies that \( g_{34}(\sigma_3) \neq 0 \).

To prove that \( g_{34} \) is surjective, let \( \sigma_4 \in G_{24} \). Then there exists \( \sigma_2 \in G_2 \) such that \( g_{24}(\sigma_2) = \sigma_4 \). Denoting \( \sigma_3 = g_{23}(\sigma_2) \) we have that \( g_{34}(\sigma_3) = \sigma_4 \). This completes the proof.

\[ \square \]

**Proof of Theorem 3.5.** Using Lemma 3.3 for \( U_n(L - \epsilon, r) \) and \( U_n(L + \epsilon, r) \) we have that for \( n \) large enough

\[
\begin{align*}
\mathbb{P}
&\left( \hat{D}_{L+\frac{3}{2}\epsilon, 2r} \subset U_n(L + \epsilon, r) \subset \hat{D}_{L+\frac{1}{2}\epsilon, 2r} \right) \geq 1 - 3n e^{-C_{\epsilon/2} 2n^d}, \\
\mathbb{P}
&\left( \hat{D}_{L-\frac{3}{2}\epsilon, 2r} \subset U_n(L - \epsilon, r) \subset \hat{D}_{L-\frac{1}{2}\epsilon, 2r} \right) \geq 1 - 3n e^{-C_{\epsilon/2} 2n^d}.
\end{align*}
\]

(A.12)

Since we assume \( L \) is \( \epsilon \)-regular, if \( r \) is small enough, we have

\[
D_{L+2\epsilon} \subset \hat{D}_{L+\frac{3}{2}\epsilon, 2r} \subset \hat{D}_{L+\frac{1}{2}\epsilon, 2r} \subset D_L \subset \hat{D}_{L-\frac{3}{2}\epsilon, 2r} \subset \hat{D}_{L-\frac{1}{2}\epsilon, 2r} \subset D_{L-2\epsilon},
\]

and from (A.12) we conclude that

\[
\mathbb{P} (D_{L+2\epsilon} \subset U_n(L + \epsilon, r) \subset D_L \subset U_n(L - \epsilon, r) \subset D_{L-2\epsilon}) \geq 1 - 6n e^{-C_{\epsilon/2} 2n^d}.
\]

Since we assume that \( f(x) \) has no critical values in \([L - 2\epsilon, L + 2\epsilon]\), we have that

\[
D_{L+2\epsilon} \simeq D_L \simeq D_{L-2\epsilon}.
\]

In particular the map \( \iota_* : H_*(D_{L+2\epsilon}) \to H_*(D_{L-2\epsilon}) \) is an isomorphism and its image is isomorphic to \( D_L \). Using Lemma A.2 completes the proof.

\[ \square \]

**Proof of Theorem 3.8.** Let \( E \) be the event that for every \( 1 \leq i \leq M \) the following inclusion holds -

\[
D_{Li+2\epsilon} \hookrightarrow U_n(L_i + \epsilon, r) \hookrightarrow D_{Li} \hookrightarrow U_n(L_i - \epsilon, r) \hookrightarrow D_{Li-2\epsilon}.
\]

(A.13)
Applying Lemma 3.3 $M + 1$ times we can show that for $n$ large enough
\[
\mathbb{P}(E) \geq 1 - 3n(M + 1)e^{-C'\epsilon^2/n^d}.
\]

From here on we will assume that (A.13) is true for all $1 \leq i \leq M$. Choosing $i^*$ as
\[
i^* \triangleq 1 + \min \left\{ i \in \{1, \ldots, M\} : \beta_m(L_i, \epsilon; n) = 1 \right\},
\]
our goal is to show that $[L_{i^*} - 2\epsilon, L_{i^*} + 2\epsilon] \subset (A, B)$, and therefore the arguments used in the proof of Theorem 3.5 guarantee that $\hat{H}_s(L_{i^*}, \epsilon; n) \cong H_s(D_{L_{i^*}}) \cong H_s(M)$.

Since $M$ is assumed to be connected, we have that $\beta_m(M) = 1$, and by Poincaré duality (cf. [29, 35]) we conclude that $\beta_m(M) = 1$. If $L_i \in (A, B)$ then from Definition 3.7 we have that $D_{L_i} \cong M$, and thus $\beta_m(D_{L_i}) = 1$ as well. On the other hand, if $L_i > B$ then $D_{L_i} \cong M'$ where $M'$ is a locally contractible proper subset of the $M$. Using Proposition 3.46 in [29] we have that $\beta_m(M') = \beta_m(D_{L_i}) = 0$.

Our requirement that $L_{i-1} - L_i = 2\epsilon$ and $B - A \geq 8\epsilon$ guarantees that there are at least four consecutive levels $L_i$ such that $L_i \in (A, B)$. Let $L_{i_1} > L_{i_2} > L_{i_3} > L_{i_4}$ be the first (highest) such levels. For $k = 2, 3$ we have that $[L_{i_k} - 2\epsilon, L_{i_k} + 2\epsilon] \subset (A, B)$, and from the proof of Theorem 3.5 and the previous paragraph we conclude that $\hat{\beta}_m(L_{i_k}, \epsilon; n) = 1$. For $i_1$ however, it is not true that $[L_{i_1} - 2\epsilon, L_{i_1} + 2\epsilon] \subset (A, B)$ and therefore, $\hat{\beta}_m(L_{i_1}, \epsilon; n)$ might be either zero or one. Finally, defining $i^*$ the way we did, $i^*$ might be either $i_2$ or $i_3$. In both cases we have $[L_{i^*} - 2\epsilon, L_{i^*} + 2\epsilon] \subset (A, B)$, and that completes the proof.

Proof of Theorem 3.9. Recall that $\mathcal{D} = \{D_L\}_{L \in \mathbb{R}}$ is the continuous filtration of the super level sets, and $\hat{\mathcal{D}} = \{U_n(L_i, r)\}_{i \in \mathbb{Z}}$ is a discrete approximation. To prove that the corresponding persistent homologies $\text{PH}_*(f), \hat{\text{PH}}_*(f)$ satisfy
\[
d_B \left( \hat{\text{PH}}_*(f), \text{PH}_*(f) \right) \leq 5\epsilon,
\]
we will use the language of $\epsilon$-interleaving introduced in [15]. The first step would be to define a discrete version of the filtration $\mathcal{D}$ given by
\[
\mathcal{D}^\epsilon \triangleq \{D_{L_i + \epsilon}\}_{i \in \mathbb{Z}},
\]
where $L_i = f_{\text{max}} - 2i\epsilon$, and denote its persistent homology by $\hat{\text{PH}}_*(f)$. Since $\mathcal{D}^\epsilon$ is a discrete approximation of the continuous filtration $\mathcal{D}$, with step size
2\epsilon, the maximum difference between \( \text{PH}_*(f) \) and \( \text{PH}_{\epsilon}^*(f) \) would be the step size, and thus we have

\[
d_B(\text{PH}_{\epsilon}^*(f), \text{PH}_*(f)) \leq 2\epsilon.
\]

To prove the theorem, it is therefore enough to show that with a high probability we have \( d_B(\hat{\text{PH}}_{\epsilon}^*(f), \text{PH}_{\epsilon}^*(f)) \leq 3\epsilon \).

Let \( E \) be the event that we have the following sequence of inclusions:

(A.14)

\[
D_{L_0+\epsilon} \;
\.

\Rightarrow

U_n(L_0, r) \;
\Rightarrow

D_{L_1+\epsilon} \;
\Rightarrow

U_n(L_1, r) \;
\Rightarrow

D_{L_2+\epsilon} \;
\Rightarrow

\cdots \Rightarrow

U_n(L_2, r)
\]

Setting \( M = \lceil f_{\text{max}}/2\epsilon \rceil \), and applying Lemma 3.3 \( M \) times we can show that

\[
P(E) \geq 1 - 3Mne^{-C_{\epsilon/2}nr_d}.
\]

Using the notation in [15] (A.14) implies that \( D^\epsilon \) and \( \hat{D}^\epsilon \) are weakly \( \epsilon \)-interleaving. Denoting the persistent homology of \( \hat{D}^\epsilon \) by \( \hat{\text{PH}}_{\epsilon}^*(f) \), using Theorem 4.3 in [15] yields

(A.15)

\[
d_B \left( \hat{\text{PH}}_{\epsilon}^*(f), \text{PH}_{\epsilon}^*(f) \right) \leq 3\epsilon.
\]

This completes the proof.

\[\square\]

**Proof of Theorem 4.3.** Consider the following sequence of simplicial complexes,

(A.16)

\[
C_n(L \pm \epsilon, r) \hookrightarrow R_n(L \pm \epsilon, r) \hookrightarrow C_n(L \pm \epsilon, \sqrt{2}r).
\]

This sequence induces the following sequence in homology

\[
H_*(C_n(L \pm \epsilon, r)) \to H_*(R_n(L \pm \epsilon, r)) \to H_*(C_n(L \pm \epsilon, \sqrt{2}r)),
\]

or equivalently,

\[
H_*(U_n(L \pm \epsilon, \sqrt{2}r)) \to H_*(R_n(L \pm \epsilon, r)) \to H_*(U_n(L \pm \epsilon, \sqrt{2}r)).
\]

From the proof of Theorem 3.5 (see (A.10),(A.11)) we have that

\[
\begin{align*}
\mathbb{P} \left( \hat{D}_{L+\frac{\epsilon}{2}, 2\epsilon} \not\subset U_n(L + \epsilon, r) \right) & \leq 2ne^{-C_{\epsilon/2}nr_d}, \\
\mathbb{P} \left( \hat{D}_{L-\frac{\epsilon}{2}, 2\epsilon} \not\subset U_n(L - \epsilon, r) \right) & \leq 2ne^{-C_{\epsilon/2}nr_d}, \\
\mathbb{P} \left( U_n(L + \epsilon, \sqrt{2}r) \not\subset \hat{D}_{L+\frac{\epsilon}{2}, 2\sqrt{2}r} \right) & \leq ne^{-C_{\epsilon/2}2d/2nr_d}, \\
\mathbb{P} \left( U_n(L - \epsilon, \sqrt{2}r) \not\subset \hat{D}_{L-\frac{\epsilon}{2}, 2\sqrt{2}r} \right) & \leq ne^{-C_{\epsilon/2}2d/2nr_d}.
\end{align*}
\]
Therefore, for $n$ large enough we have
\[
\mathbb{P}\left( \tilde{D}_{L+\frac{3}{2}\epsilon,2r} \subset U_n(L + \epsilon, r) \subset \tilde{D}_{L+\frac{3}{2}\epsilon,2\sqrt{2r}} \supset \hat{D}_{L+\frac{3}{2}\epsilon,2\sqrt{2r}} \right) \geq 1 - 3n e^{C^*_{\epsilon/2}n^{rd}},
\]
\[
\mathbb{P}\left( \tilde{D}_{L-\frac{1}{2}\epsilon,2r} \subset U_n(L - \epsilon, r) \subset \tilde{D}_{L-\frac{1}{2}\epsilon,2\sqrt{2r}} \supset \hat{D}_{L-\frac{1}{2}\epsilon,2\sqrt{2r}} \right) \geq 1 - 3n e^{C^*_{\epsilon/2}n^{rd}}.
\]
Since we assume that all the levels we study are $\epsilon$-regular and compact, if $r$ is small enough we can order them in the following way
\[
D_{L+2\epsilon} \subset \tilde{D}_{L+\frac{3}{2}\epsilon,2r} \subset \tilde{D}_{L+\frac{3}{2}\epsilon,2\sqrt{2r}} \subset D_L \subset \tilde{D}_{L-\frac{1}{2}\epsilon,2r} \subset \tilde{D}_{L-\frac{1}{2}\epsilon,2\sqrt{2r}} \subset D_{L-2\epsilon}.
\]
Combining that with (A.16), we conclude that with a high probability we have the following sequence in homology (induced by composing inclusion maps),
\[
H_*(D_{L+2\epsilon}) \rightarrow H_*(R_n(L+\epsilon, r)) \rightarrow H_*(D_L) \rightarrow H_*(R_n(L-\epsilon, r)) \rightarrow H_*(D_{L-2\epsilon}).
\]
Since $f(x)$ has no critical values in $[L - 2\epsilon, L + 2\epsilon]$, using Lemma A.2 completes the proof.

\begin{flushright}
\text{Department of Mathematics,}
\text{Duke University, Durham NC 27708}
\text{E-mail: omer@math.duke.edu}
\end{flushright}

\begin{flushright}
\text{Departments of Statistical Science,}
\text{Computer Science, and Mathematics}
\text{Duke University, Durham NC 27708}
\text{E-mail: sayan@stat.duke.edu}
\end{flushright}

\begin{flushright}
\text{Department of dept. of Statistics, Stanford University,}
\text{Stanford, CA 94305-4065}
\text{E-mail: jonathan.taylor@stanford.edu}
\end{flushright}