

## RESEARCH ARTICLE

WILEY

# Random Čech complexes on manifolds with boundary

Henry-Louis de Kergorlay<sup>1</sup>  | Ulrike Tillmann<sup>2</sup>  | Oliver Vipond<sup>2</sup> <sup>1</sup>School of Mathematics, University of Edinburgh, Edinburgh, UK<sup>2</sup>Mathematical Institute, University of Oxford, Oxford, UK**Correspondence**

Oliver Vipond, Mathematical Institute, University of Oxford, Oxford, UK.

Email: [olivervipond@gmail.com](mailto:olivervipond@gmail.com)**Abstract**

Let  $M$  be a compact, unit volume, Riemannian manifold with boundary. We study the homology of a random Čech-complex generated by a homogeneous Poisson process in  $M$ . Our main results are two asymptotic threshold formulas, an upper threshold above which the Čech complex recovers the  $k$ th homology of  $M$  with high probability, and a lower threshold below which it almost certainly does not. These thresholds share the same leading term. This extends work of Bobrowski–Weinberger and Bobrowski–Oliveira who establish similar formulas when  $M$  has no boundary. The cases with and without boundary differ: the corresponding common leading terms for the upper and lower thresholds differ being  $\log(n)$  when  $M$  is closed and  $(2 - 2/d) \log(n)$  when  $M$  has boundary; here  $n$  is the expected number of sample points. Our analysis identifies a special type of homological cycle occurring close to the boundary.

**KEYWORDS**

homology, random geometric complexes, stochastic topology

## 1 | INTRODUCTION

Randomly generated simplicial complexes and their topology have recently attracted a lot of attention. The survey article by Bobrowski and Kahle [2] collects together results and provides a wealth of open problems in this field. Here we focus on understanding the homology of random geometric

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Random Structures and Algorithms* published by Wiley Periodicals LLC.

complexes. While this topic was first studied in [25] (see also [17]) our results build directly on [4,5].

Much of the current interest in the topology of random simplicial complexes is due to applications to topological data analysis, where random complexes can serve as null models when interpreting the topology of complexes on data sets. In the context of manifold learning, a nonlinear dimension reduction technique, one is interested in recovering the structure of low dimensional manifolds embedded in high dimensional space. Studying topological properties of the underlying manifold, such as homology, can inform the choices of hyperparameters in this dimension reduction technique [23,28]. In the specific context of persistent homology, one of the main tools of topological data analysis, one adopts a multiscale approach to the study of the homology of randomly sampled data from a manifold. Understanding the conditions for which the topology of a complex built from a random sample coincides with the homology of the underlying manifold informs which bars in the multiscale barcode invariant relate to inherent features of the underlying manifold. For surveys of persistent homology see [9,22,30] and for work in stochastic persistent homology see [3].

Various flavors of random simplicial complexes are present in the literature, which can include or exclude geometric considerations. Random simplicial complexes that extend the notion of a random graph in the sense of Erdős–Rényi to higher dimensional complexes are not constrained by an underlying geometry [15]. In contrast, we shall work with the random simplicial complex realized by the Čech complex associated to a Poisson point process on a Riemannian manifold and ask the question when the topology of the simplicial complex approximates that of the manifold.

The question of recovering the topology of a space from a finite sample has been studied in [7,21,29] and [4,5] in differing contexts. In [21] the authors consider submanifolds in Euclidean space and use the metric of the ambient Euclidean space when building the Čech complex. They provide explicit conditions to recover the homology of the manifold with high confidence. Naturally these explicit conditions are dependent on the curvature and nearness to self-intersection of the embedded manifold. This theory has been extended to manifolds with boundary in [29] and more general compact subset in [7]. In contrast, as in the context of [4,5], we base the construction of the Čech complex on an intrinsic metric of the manifold independent of any embedding. Our work studies asymptotic properties and the phase transition for which one can recover the homology with high probability when increasing the size of point sample and decreasing the radius over which the associated complex is constructed. Specifically, we will give an answer to the problem posed in the survey article [2] about extending the homological connectivity theory established for closed Riemannian manifolds in [4], to Riemannian manifolds with boundary.

A principal advantage of studying asymptotics is that results rely on fewer assumptions on the underlying manifold from which the point process is sampled. Our main result like that of [4] thus only has dependence on the dimension of the underlying manifold and the homological dimension one wishes to recover. Our argument follows a similar framework to the argument presented in [4,5]. We will need to develop completely new arguments to take into account the effect of the boundary.

Our main result is stated with respect to the term  $\Lambda := \omega_d n r^d$ , the expected number of points of a uniform Poisson process of intensity  $n$  lying in a  $d$ -dimensional radius  $r$  ball; here  $\omega_d$  denotes the volume of the unit ball in  $\mathbb{R}^d$ . Depending on the asymptotic behavior of the  $\Lambda$ , the associated Čech complex at scale  $r$  built on the point process exhibits different behaviors.

There are three distinct regimes of behavior for  $\Lambda$  as  $n \rightarrow \infty, r \rightarrow 0$ . In the subcritical regime ( $\Lambda \rightarrow 0$ ) the connectivity of the Čech complex is very sparse and mostly disconnected, with the number of connected components growing at the same rate as the number of points. In the critical regime ( $\Lambda \rightarrow \lambda \in (0, \infty)$ ) the Čech complex is sufficiently connected to exhibit nontrivial homology.

However, the number of connected components still grows linearly with the number of points. In the supercritical regime ( $\Lambda \rightarrow \infty$ ) for sufficiently large  $\Lambda$  the Čech complex is connected, and for even larger  $\Lambda$  the point cloud covers the underlying manifold with high probability.

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

The nontrivial boundary has a significant impact on the homological connectivity thresholds. As far as we know, our study of manifolds with boundary has uncovered a previously unobserved phenomenon occurring close to the boundary. Our analysis shows that close to the boundary a large number of spurious  $k$ -cycles appear which are not homological cycles inherent to the  $k$ th homology of the underlying manifold. This phenomenon determines that the homological connectivity thresholds for manifolds with boundary are larger than those for a closed manifold. We attain the following result:

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

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

Note that when simplified the coefficient of the second-order term in the lower threshold,  $2(\frac{1}{d} - 3)$ , is independent of the homological dimension  $k$ . We state the coefficient of the second-order term of the lower threshold of our theorem in unsimplified form to make an easy comparison to the lower thresholds established by Bobrowski and Weinberger in [5] and by Bobrowski and Oliveira in [4], who studied the case when  $M$  is a  $d$ -dimensional torus and when  $M$  is an arbitrary compact closed Riemannian manifold respectively. With the same setup as our theorem but on a closed Riemannian manifold the corresponding thresholds for  $1 \leq k \leq d - 1$  are computed to be:

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

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

Finally, we note that our assumption that  $\Lambda \rightarrow \infty$  implicitly implies that  $w(n)$  goes to infinity no faster than  $O(\log n)$ .

## 1.1 | Outline

Our argument follows the same structure as the arguments presented in [4,5]. There are several key ideas in this framework. The first essential idea is to bound the number of homological cycles of a

complex by counting the critical points of an associated Morse function. This simplifies the task of counting global phenomena of homological cycles to the purely local considerations which determine critical points of a Morse function.

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

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

In Section 2 we collate results from the theory of point processes, define our asymptotic notation and also collect Riemannian volume estimates which we use in order to produce bounds in the later sections.

Section 3 provides a brief introduction to classical Morse theory. Since the distance function induced by a point process is not necessarily smooth we also provide the necessary results from [12] which describe Morse theory for a wider class of functions to which the distance function belongs, the so-called min-type functions.

In Section 4 we derive a coverage result for Riemannian manifolds with boundary. This coverage result is required to establish the upper homological connectivity thresholds.

We introduce in Section 5 the change of variables formulae we require in later sections to compute bounds on the number of critical points. These formulae are also used in Appendix C in order to bound the variance of the number of critical points.

Section 6 is dedicated to computing an upper bound for the expected number of critical points using the tools and results provided in previous sections. This upper bound on the number of critical points is used to produce an upper threshold for the homological connectivity in terms of  $\Lambda$ .

In Section 7 we identify a special class of critical point which induce erroneous homological cycles, (homological cycles which are not inherent to the underlying manifold). We produce a lower bound for the expected number of this type of critical points. This lower bound is used to produce a lower threshold for the homological connectivity in terms of  $\Lambda$ .

In Section 8 we calculate homological connectivity thresholds given the computed bounds on the number of critical points. Finally in Section 9 we compare and contrast the homological connectivity thresholds for Riemannian manifolds with boundary we have calculated to those in [4] for closed manifolds. We indicate how the geometric difference in the two situations informs the difference in the thresholds.

For completeness we include an Appendix containing proofs which are similar in structure to their analogues in the case that the manifold is closed. In Appendix C we bound the variance of the number of occurrences of the special type of homological cycles established in Section 7. We show that the variance is small and so with high probability the number of erroneous cycles behaves like the expected number of erroneous cycles.

## 2 | BACKGROUND

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

## 2.1 | Homology and complexes

Homology is a measure of complexity of a topological space. It is an algebraic invariant that has proved a powerful tool in the study of geometry. A thorough introduction can be found in [13] far more comprehensive than the brief overview of basic homology theory we present for the unfamiliar reader.

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

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

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

**Definition 2.1** (Čech complex). Let  $\mathcal{P}$  be a collection of points in a metric space  $(M, \rho)$ . Define a one parameter family  $\mathcal{C}(\mathcal{P}, r)$  of simplicial complexes on vertex set  $\mathcal{P}$  associated to this collection of points as follows: For  $r \in [0, \infty)$ ,

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

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

**Lemma 2.2** (Nerve Lemma). Let  $\mathcal{U} = \{B_r(p) : p \in \mathcal{P}\}$  be an open cover of the metric space  $(M, \rho)$ . Suppose that the finite intersections of sets in  $\mathcal{U}$  are empty or contractible, then the Čech complex  $\mathcal{C}(\mathcal{P}, r)$  is homotopy equivalent to  $M$ .

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

## 2.2 | Poisson point processes

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

**Definition 2.3** (General Poisson point process [1]). Let  $(M, \mathcal{F}, \mu)$  be a measure space with  $M$  a compact metric space,  $\mathcal{F}$  a collection of measurable sets, and measure  $\mu$  which is finite on compact sets and with no atoms. The Poisson process on  $M$  of intensity measure  $\mu$  is a point process on  $M$  such that:

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

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

A salient feature of Poisson point processes is the following independence result, Theorem 2.4. We shall use this result in our calculations to compute bounds on the number of critical points of the distance function from the point process which correspond to simplices of the associated Čech complex.

**Theorem 2.4** (Palm Theory [24]). *Let  $(X, \rho)$  be a metric space,  $f : X \rightarrow \mathbb{R}$  a probability density and  $P_n$  a Poisson process on  $X$  with intensity  $\lambda_n = nf$ . If  $h(\mathcal{Y}, \mathcal{X})$  is a measurable function for all finite subsets  $\mathcal{Y}, \mathcal{X} \subset X$ , with  $\mathcal{Y} \subset \mathcal{X}$  and  $|\mathcal{Y}| = k + 1$  then:*

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

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

### 2.3 | Asymptotic notation

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

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

### 2.4 | Riemannian volumes

A major portion of our later proofs require us to bound various Riemannian volumes by their Euclidean counterparts in order to control the asymptotic behavior of the Betti numbers. In this section we shall provide these approximations of Riemannian volumes. Let us denote a smooth Riemannian manifold as the pair  $(M, g)$  and consider the case when  $M$  is compact, of dimension  $d$ , and the metric  $g$  is smooth. An introduction to Riemannian geometry can be found for example in [16].

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

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

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

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

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

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

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

The Ricci curvature tensor at a point  $p$  is given by  $\text{Ric}_{ij} = -\sum_k R_{ikkj}$ , and we can calculate that:

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

Our first volume approximations apply to balls within a manifold with boundary which are wholly contained within the manifold, that is the centers lie far enough from the boundary. The proofs of these approximations can be found in the Appendix of [4].

**Lemma 2.5** ([4]). *Let  $s(p) = \sum_i \text{Ric}_{ii}$  denote the scalar curvature at  $p$  and define  $s_{\min}(\nu) = \inf_{p \in M} \frac{s(p)}{6(d+2)} - \nu$ ,  $s_{\max}(\nu) = \sup_{p \in M} \frac{s(p)}{6(d+2)} + \nu$ . Then for all  $\nu > 0$  there is a continuous choice of  $r_\nu > 0$  such that for all  $r \leq r_\nu$*

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

**Lemma 2.6** ([4]). *Let  $(M, g)$  be a compact Riemannian manifold of dimension  $d$ . Then there is some  $\nu > 0$  and  $r_\nu > 0$  such that for all  $r < r_\nu$  and any two point with  $\text{dist}(p_1, p_2) < 2r$  we have:*

$$(B_{(1-\nu)r}^E(p_1) \cup B_{(1-\nu)r}^E(p_2)) \subset (B_r(p_1) \cup B_r(p_2)) \subset (B_{(1+\nu)r}^E(p_1) \cup B_{(1+\nu)r}^E(p_2)).$$

Let us now consider how we must adapt these approximations for balls intersecting the boundary. We may consider our manifold to be embedded in  $\mathbb{R}^D$ , for some  $D$ . Indeed, Nash's Imbedding Theorem [20] guarantees existence of an isometric embedding.

**Definition 2.7** (Reach of a manifold). Let the medial axis  $\text{Med}(M)$  of a manifold  $M$  embedded in  $\mathbb{R}^D$ , be the set of points in  $\mathbb{R}^D$  which do not have a unique nearest element in  $M$ . The reach  $\tau_M$  of a manifold is defined to be  $\tau_M = \inf_{p \in \text{Med}(M)} \text{dist}(M, p)$ .

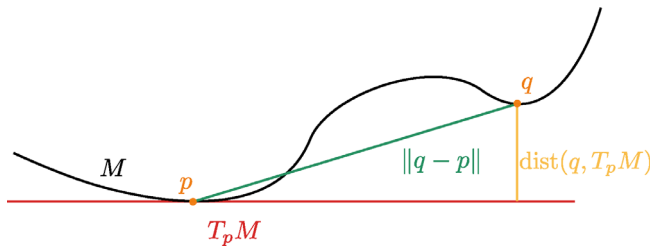
**Theorem 2.8** (Theorem 4.18 [10]). *Let  $M$  be a submanifold of  $\mathbb{R}^D$ ,  $\|\cdot\|$  denote the standard Euclidean norm on  $\mathbb{R}^D$ , and  $\tau_M$  the reach of  $M$ . Then the reach is realized as:*

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

See Figure 1.

The manifolds we consider are smooth and compact and so in particular the reach of our manifolds is nonzero. We shall frequently consider the double manifold  $DM = M \cup_{\partial M} M'$  where  $M'$  is a copy of  $M$ , and the two copies of  $M$  are glued along their common boundary. This gives us a canonical compact





**FIGURE 1** We sketch the distances involved in Theorem 2.8 for which we can calculate the reach of a manifold embedded in Euclidean space

closed manifold in which  $M$  is embedded. The following result provides an estimate for the volume of a ball centered near the boundary.

**Proposition 2.9.** *Let  $B_r^{DM}(c)$  be a ball of radius  $r$  centered at  $c \in M$  in the double manifold  $DM$ , and let  $\delta$  denote the distance from the center  $c$  to the boundary. Let  $B_r^M(c) = B_r^{DM}(c) \cap M$  denote the portion of this ball contained in  $M$ . Then the volume of  $B_r^M(c)$  can be expressed as:*

$$\text{Vol}(B_r^M(c)) = \left( \frac{1}{2} + \alpha \right) \text{Vol}(B_r^{DM}(c)),$$

where  $\alpha = O\left(\frac{\delta}{r} + r\right)$ .

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

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

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

$$\tilde{\alpha} = O\left(\frac{\delta}{r}\right).$$

Let us now estimate the error volume bounded between the boundary and  $T_p M$ . For all  $q \in M$  Theorem 2.8 implies that  $\text{dist}(q, T_p M) \leq \|q - p\|^2 / \tau_M$ . Hence we can bound the error volume between  $T_p M$  and a parallel hyperplane at distance  $O(r^2)$  from  $T_p M$ . Thus the error volume is bounded by  $\text{Vol}(B_r^{(d-1)}) \cdot r^2 = O(r^{d+1})$  (where  $B_r^{(d-1)}$  denotes the radius  $r$ ,  $(d-1)$ -dimensional ball). We attain the desired result that:

$$\text{Vol}(B_r^M(c)) = \left( \frac{1}{2} + \alpha \right) \text{Vol}(B_r^{DM}(c)),$$

where  $\alpha = O\left(\frac{\delta}{r} + r\right)$ . ■



We can combine this result with Lemma 2.6 to similarly estimate the volume of intersecting balls lying close to the boundary.

If  $M$  is a smooth manifold with nonempty boundary then we call a neighborhood of the boundary  $\partial M$  a *collar neighborhood* if it is the image of a smooth embedding  $\partial M \times [0, 1) \rightarrow M$  and the embedding restricts to the identification  $\partial M \times \{0\} \rightarrow \partial M$ . A standard result (sometimes called the Collar Neighborhood Theorem [16]) guarantees any smooth manifold with nonempty boundary has a collar neighborhood. For small  $r$  the  $r$ -neighborhood of the boundary denoted  $\partial M_r$ , is a collar neighborhood. It is straight forward to show that the volume of this collar neighborhood  $\text{Vol}(\partial M_r) \sim r \text{Vol}(\partial M)$ .

### 3 | MORSE THEORY

#### 3.1 | Morse theory on manifolds with boundary

We start with a terse summary of Morse theory for manifolds with boundary. The upshot of this section is that given a suitable smooth function on the manifold we define the Morse complex in the same manner as in [19]. This complex will compute relative homology  $H_*(M, \partial M)$  if the function attains a minimum on the boundary, and will compute absolute homology  $H_*(M)$  if the function attains a maximum on the boundary.

**Definition 3.1** (Non-degenerate critical point [19]). Let  $M$  be a smooth  $d$ -manifold with boundary and  $f : M \rightarrow \mathbb{R}$  a smooth function. A point  $p \in M$  is critical if  $\nabla f(p) = 0$ . A critical point is non-degenerate if the Hessian  $H(p)$  is non-singular.

**Definition 3.2** (Smooth manifold triad [19]). We say  $(M; V_0, V_1)$  is a smooth manifold triad if  $M$  is a smooth manifold with boundary, and the boundary  $\partial M$  is the disjoint union of the open and closed submanifolds  $V_0, V_1$ .

**Definition 3.3** (Morse function [19]). A smooth function  $f : (M; V_0, V_1) \rightarrow [a, b]$  is a Morse function on a smooth manifold triad if  $f^{-1}(a) = V_0, f^{-1}(b) = V_1$  and all the critical points of  $f$  lie in the interior of  $M$  ( $\text{int}(M) = M \setminus \partial M$ ).

In this setting we still have the classical Morse Lemma [27] for closed manifolds that asserts the existence of a coordinate system about each critical point for which the Morse function has a diagonal quadratic form. The index of a critical point is again given by the dimension of the negative eigenspace of the Hessian matrix  $H(p)$ .

**Theorem 3.4** (Theorem 7.4 [19]). *There is a chain complex  $C_*$ , where  $C_k$  is the free abelian group on the set of index  $k$  critical points of  $f$ , and moreover  $H_k(C_*) \cong H_k(M, V_0)$ .*

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

#### 3.2 | Morse theory for min-type functions

While regular Morse theory is concerned with smooth functions, the distance function associated to a point cloud on a Riemannian manifold is generally not smooth. However, the square of such a distance function is a min-type function for which a version of Morse theory has been developed [12].

**Definition 3.5** (Min-type function [12]). Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a germ of a continuous function at  $p \in \mathbb{R}^d$ . Then  $f$  is a germ of a min-type function at  $p$  if there exist germs of smooth functions  $\alpha_i$  at  $p$  such that locally around  $p$  we have that  $f = \min_{i=1}^m \alpha_i$ . A function on a  $d$ -manifold  $f : M \rightarrow \mathbb{R}$  is min-type if for all  $p \in M$  the germ at  $p$  is a germ of a min-type function.

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

**Definition 3.6** (Non-degenerate regular point [12]). Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a germ of a min-type function. The point  $p \in \mathbb{R}^d$  is a non-degenerate regular (NDR) point if  $f$  has a minimal representation at  $p$ ,  $f = \min_{i=1}^m \alpha_i$ , such that:

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

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

**Definition 3.7** (Non-degenerate critical point [12]). Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a germ of a min-type function. The point  $p \in \mathbb{R}^d$  is a non-degenerate critical (NDC) point if  $f$  has a minimal representation at  $p$ ,  $f = \min_{i=1}^m \alpha_i$ , such that:

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

**Definition 3.8** (Morse min-type function). We say that a min-type function  $f : M \rightarrow \mathbb{R}$  is a Morse min-type function if every point is either an NDR or an NDC point.

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

**Definition 3.9** (Index [12]). Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a Morse min-type function with NDC point  $p$  and associated minimal representation  $f = \min_{i=1}^m \alpha_i$ . The index of  $p$  is defined to be  $\text{Ind}_p(f) = (m - 1) + \text{Ind}_p(f|_{\mathcal{G}_f})$

One of the main results from [12] relates Morse min-type functions to smooth Morse functions. In essence the following theorem says that a Morse min-type function can be  $\varepsilon$ -approximated by a classical smooth Morse function with the same number of critical points of the same index, and these points are arbitrarily close to the original critical points.

**Theorem 3.10** (Morse min-type approximation theorem [12]). Let  $f$  be a Morse min-type function on a compact closed Riemannian manifold  $(M, g)$  with critical points  $y_1, \dots, y_m$  and corresponding indices  $q_1, \dots, q_m$ . For any  $\varepsilon > 0$  there is a smooth Morse function  $f_\varepsilon$  satisfying the following properties:

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

### 3.3 | Morse theory for the distance function on a compact manifold with boundary

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

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

**Lemma 3.11** ([4]). *Let  $(M, g)$  be a compact Riemannian manifold with boundary, then there exists a positive real number  $r_{\text{mt}} > 0$  such that for every finite sample of points  $\mathcal{P} \subset M \setminus \partial M$  the distance function  $\rho_{\mathcal{P}}^2$  is a Morse min-type function on the  $r_{\text{mt}}$  neighborhood of these points  $B_{r_{\text{mt}}}(\mathcal{P})$ .*

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

**Proposition 3.12.** *Let  $(M, g)$  be a compact Riemannian manifold with boundary and let  $\mathcal{P}$  be a finite subset of  $M$  such that  $M \subset B_r(\mathcal{P})$  for some  $r < \frac{r_{\text{mt}}}{2}$ . Then the critical points of the Morse min-type distance function  $\rho_{\mathcal{P}}^2$  are in one to one index preserving correspondence with the critical points of a Morse function on the smooth manifold triad  $(M; \emptyset, \partial M)$ .*

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

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

Then the distance function  $\rho_{\mathcal{P} \cup \mathcal{P}'}^2$  is of Morse min-type on all of  $DM$  and hence, by Theorem 3.10, has an  $\varepsilon$ -approximation by a smooth Morse function  $f$  on  $DM$ .

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

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

**Corollary 3.13.** *Let  $\mathcal{P}_n$  be a Poisson process on  $M$  a compact Riemannian manifold with boundary. Then with high probability the distance function  $\rho_{\mathcal{P}_n}^2$  induces a Morse function on the smooth manifold triad  $(M; \emptyset, \partial M)$ . That is to say  $\mathbb{P}(\{\rho_{\mathcal{P}_n}^2 \text{ induces a Morse function}\}) \rightarrow 1$  as  $n \rightarrow \infty$ .*

*Proof.* Observe that for a Poisson process  $\mathcal{P}_n$  on  $M$  and fixed  $r < \frac{r_{\text{mt}}}{2}$ ,  $\mathbb{P}(M \subset B_r(\mathcal{P}_n)) \rightarrow 1$  as  $n \rightarrow \infty$ . Hence using Proposition 3.12 the point process  $\mathcal{P}_n$  induces a Morse min-type distance function  $\rho_{\mathcal{P}_n}^2$  whose critical points recover the homology of  $M$  with high probability. ■

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

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

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

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

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

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

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

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

**Lemma 3.14** ([4]). *There is a positive  $r_{\text{max}} < r_{\text{mt}}$  such that if  $\mathcal{Y}$  is generic with  $E_{r_{\text{max}}}(\mathcal{Y}) \neq \emptyset$ , then the set  $\mathcal{Y}$  has a unique point  $c(\mathcal{Y}) \in M$  such that for all  $p \in \mathcal{Y}$*

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

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

## 4 | ASYMPTOTIC COVERAGE

In forming the upper threshold we need to understand the asymptotic coverage of a Riemannian manifold with boundary. We would like to understand the conditions under which  $B_r(\mathcal{P})$  covers  $M$  w.h.p.

The paper Random Coverings [11] gives a comprehensive treatment of the asymptotic behavior of the number of randomly chosen radius  $r$  balls required to cover a compact *closed* Riemannian manifold

$m$  times. The result is translated into a sharp coverage threshold in [5], but again this result applies only to *closed* Riemannian manifolds. We use the coverage result of [5] to derive conditions (not necessarily sharp) for asymptotic coverage of a compact Riemannian manifold with boundary. A sharp coverage result for manifolds with boundary found in [6] has come to our attention. Nevertheless, we include our coverage result which is sufficiently strong for our purposes and arises from a natural geometric argument.

**Theorem 4.1** (Sharp coverage threshold [5]). *Let  $M$  be a compact, closed, finite volume Riemannian manifold. Let  $\mathcal{P}_n$  be a uniform Poisson process on  $M$  of intensity  $n$ . Let  $B_r(\mathcal{P}_n)$  denote the  $r$  neighborhood of the points  $\mathcal{P}_n$  and  $w(n) \rightarrow \infty$ , then we yield:*

$$\lim_{n \rightarrow \infty} \mathbb{P}(M \subset B_r(\mathcal{P}_n)) = \begin{cases} 1 & \Lambda = \log n + (d-1) \log \log n + w(n), \\ 0 & \Lambda = \log n + (d-1) \log \log n - w(n). \end{cases}$$

Our coverage result below is not proposed as a sharp threshold but merely a threshold that suits our needs in subsequent proofs.

**Corollary 4.2** (Coverage of compact manifold with boundary). *Let  $M$  be a compact Riemannian manifold with boundary and let  $\mathcal{Q}_n$  be a uniform Poisson process on  $M$  with intensity  $n$ . If  $\Lambda = \log n + (d-1) \log \log n + w(n)$  with  $w(n) \rightarrow \infty$ , then the  $2r$  neighborhood of the Poisson process covers  $M$  w.h.p:*

$$\lim_{n \rightarrow \infty} \mathbb{P}(M \subset B_{2r}(\mathcal{Q}_n)) = 1.$$

*Proof.* Our proof will be an application of the sharp threshold developed in [5]. Let us form the double manifold  $DM = M \cup_{\partial M} M'$ , which has double the volume of  $M$ . Now consider a uniform Poisson process  $\mathcal{P}_n$  on  $DM$  with intensity  $n$ . This restricts to a uniform Poisson process  $\mathcal{Q}_n = \mathcal{P}_n \cap M$  on  $M$  with intensity  $n$ .

Then for  $\Lambda = \log n + (d-1) \log \log n + w(n)$  Theorem 4.1 implies that

$$\lim_{n \rightarrow \infty} \mathbb{P}(DM \subset B_r(\mathcal{P}_{Cn})) = 1.$$

Let us define the following notation:  $M_r = M \setminus (\partial M \times [0, r)) = \{x \in M : \min_{y \in \partial M} \rho_M(x, y) \leq r\}$ , and the  $r$ -neighborhoods:

$$B_r^{(M)}(\mathcal{P}) = \{x \in M : \min_{p \in \mathcal{P}} \rho_M(x, p) < r\} \text{ and } B_r^{(DM)}(\mathcal{P}) = \{x \in DM : \min_{p \in \mathcal{P}} \rho_{DM}(x, p) < r\}.$$

For any point  $x \in M_r$  the  $r$ -neighborhood of  $x \in DM$  coincides with the  $r$ -neighborhood of  $x \in M$ ,  $B_r^{(M)}(\{x\}) = B_r^{(DM)}(\{x\})$ . For points in the collar  $\partial M \times [0, r)$  we cannot make the same statement, however the result follows immediately from the observation that:

$$DM \subset B_r^{(DM)}(\mathcal{P}_n) \Rightarrow M_r \subset B_r^{(DM)}(\mathcal{P}_n \cap M) \Rightarrow M_r \subset B_r^{(M)}(\mathcal{Q}_n) \Rightarrow M \subset B_{2r}^{(M)}(\mathcal{Q}_n).$$

The last implication follows since for any  $x \in M$  there is a  $y \in M_r$  and  $q \in \mathcal{Q}_n$  such that  $\rho_M(x, y) < r$  and  $y \in B_r^{(M)}(q)$ . Hence  $x \in B_{2r}^{(M)}(q)$ . ■

## 5 | BLASCHKE-PETKANTSCHIN FORMULAE

A key component of our later arguments and the arguments found in [4] is an integral formula which facilitates calculating bounds on the expected number of critical points of a distance function associated to a Poisson process on a Riemannian manifold. In this section, we explain this change of variables formula and make appropriate adaptations to the case for Riemannian manifolds with boundary.

### 5.1 | The Blaschke-Petkantschin formula in the Euclidean case

We first recall a derivation of the classical Blaschke-Petkantschin formula in the Euclidean case. Our derivation and Proposition 5.1, roughly follow Sections 2 and 3 of [18].

Let  $\mathbb{R}^d$  be a  $d$ -dimensional Euclidean space, and let  $(e_i)_{i=1}^d$  be an orthonormal moving frame in  $\mathbb{R}^d$ , where for an infinitesimal rotation of the frame,

$$e_i \cdot de_i = 0, \forall i \in [d];$$

and set

$$\omega_{ij} := e_i \cdot de_j = -\omega_{ji}, \forall i, j \in [d].$$

Given  $r$  points  $\{x_i : i \in [r]\} \subset \mathbb{R}^d$ , Miles derives heuristically the associated volume form

$$\bigwedge_{j=1}^r dV(x_j) = \bigwedge_{i=1}^d \bigwedge_{j=1}^r e_i \cdot dx_j.$$

Furthermore, given the Grassmannian manifold  $\text{Gr}(r, d)$  with invariant measure  $d\mu_{r,d}(V)$ , Miles also derives

$$d\mu_{r,d} = \bigwedge_{i=1}^r \bigwedge_{j=r+1}^d \omega_{ij}.$$

Using the above, the Blaschke-Petkantschin formula expresses the Euclidean volume form  $dV(x_i^d)$  on  $\{x_i : i \in [r]\}$  in terms of the volume element associated to the  $r$ -plane containing  $\{x_i : i \in [r]\}$ , denoted by  $dV(x_i^r)$ .

**Proposition 5.1** (Blaschke-Petkantschin formula Euclidean case [18]). *Let  $\{x_i : i \in [r]\}$  be a linearly independent set of vectors spanning  $V = \text{Span}(\{e_i : i \in [r]\}) \in \text{Gr}(r, d)$ . For each  $j \in [r]$ , let  $(\lambda_{jk})_{k \in [r]} \in \mathbb{R}^r$  be such that*

$$x_j = \sum_{k=1}^r \lambda_{jk} e_k, \text{ and let } Y := |\det(\lambda_{jk})| > 0.$$

Then

$$\bigwedge_{i=1}^r dV(x_i^d) = Y^{d-r} d\mu_{r,d} \bigwedge_{i=1}^r dV(x_i^r).$$

*Proof.* Given  $j \in [r]$  and  $i \in \{r+1, \dots, n\}$ , we have

$$\begin{aligned} dx_j &= \sum_{k=1}^r (d\lambda_{jk}e_k + \lambda_{jk}de_k), \\ e_i \cdot dx_j &= \sum_{k=1}^r \lambda_{jk}\omega_{ik}, \\ \bigwedge_{j=1}^r e_i \cdot dx_j &= |\det(\lambda_{jk})| \bigwedge_{k=1}^r \omega_{ik}, \\ \bigwedge_{i=r+1}^d \bigwedge_{j=1}^r e_i \cdot dx_j &= \Upsilon^{d-r} \bigwedge_{i=r+1}^d \bigwedge_{k=1}^r \omega_{ik}, \\ \bigwedge_{i=r+1}^d \bigwedge_{j=1}^r e_i \cdot dx_j &= \Upsilon^{d-r} d\mu_{r,d}, \end{aligned}$$

and multiplying on both sides above by  $\bigwedge_{i=1}^r \bigwedge_{j=1}^r e_i \cdot dx_j = \bigwedge_{j=1}^r dV(x_j^r)$ , we obtain the desired result. ■

## 5.2 | Blaschke-Petkantschin formula for Riemannian manifolds

Following [4], we obtain a Riemannian generalization of the Blaschke-Petkantschin formula. The formula is valid for functions with support close to the diagonal of  $M^{k+1}$ . It enables us to reparametrize a  $(k+1)$ -tuple of points near the diagonal of  $M^{k+1}$  into local coordinates about their center. We shall use bold face notation for  $(k+1)$ -tuples  $\mathbf{y} = (y_0, \dots, y_k) \in M^{k+1}$  and denote the center and radius of the corresponding collection of points  $\mathcal{Y} = \{y_0, \dots, y_k\}$  as  $c(\mathbf{y})$  and  $\rho(\mathbf{y})$  respectively (compare Lemma 3.14). The change of variables has the following form<sup>1</sup>:

$$\begin{aligned} M^{k+1} &\leftrightarrow M \times \mathbb{R} \times \text{Gr}(k, d) \times (\mathbb{S}_1^{k-1})^{k+1} \\ \mathbf{y} &\leftrightarrow (c(\mathbf{y}), u, V, \mathbf{w}) \end{aligned}$$

A  $(k+1)$ -tuple of points in  $\mathbf{y} \in M^{k+1}$  is reparametrized by the center of this tuple  $c(\mathbf{y})$ , the distance  $u$  of the points from their center, the  $k$ -plane  $V$  in which the pre-image of the points lie in the tangent space at the center, and the  $k+1$  points  $\mathbf{w}$  of the  $(k-1)$ -sphere of  $V$  upon which they lie once the radius is normalized to 1. This change of variables enables us to compute bounds for the number of critical points, since it separates variables and decomposes the integral over  $M^{k+1}$  into components which can be estimated independently.

Suppose that  $M \subset \mathbb{R}^d$  is a closed Riemannian manifold and let  $\mathbf{y} = (y_i)_{i=1}^{k+1} \in M^{k+1}$ , with center  $c = c(\mathbf{y})$ , with radius  $\rho(\mathbf{y}) \leq r$ , and local normal coordinates  $(x^1, \dots, x^d)$ . For sufficiently small  $r$ , we can write for all  $i \in [k+1]$

$$y_i = \exp_c(v_i),$$

<sup>1</sup>This formula is a local formula in the sense that the RHS should really be thought of as a double fiber bundle: the reference  $d$ -plane of  $\text{Gr}(k, d)$  is the tangent plane  $T_{c(\mathbf{y})}M$  and the spheres are copies of the unit sphere of  $V \in \text{Gr}(k, d)$ .



with

$$v_i = \sum_{j=1}^d x^j(y_i) \left( \frac{\partial}{\partial x^i} \right)_c.$$

Let  $\mathbf{1}_r(\mathbf{y}) = \mathbf{1}\{E_{r_{\max}}(y_0, \dots, y_k) \neq \emptyset \text{ and } \rho(\mathbf{y}) \leq r\}$ . Note that this indicator function has support near to the diagonal of  $M^{k+1}$  and each tuple in the support of this function has a unique center. It is shown in [4], that  $\{v_i : i \in [k+1]\}$  have linear dependency and span a  $k$ -dimensional subspace  $V \subset T_{c(\mathbf{y})}M$  when  $c(\mathbf{y})$  is a critical point. Write  $\mathbb{S}_u(V)$  for the sphere of radius  $u$  in  $V$ . We yield the following change of variable formula.

**Lemma 5.2** ([4]). *Let  $M$  be a compact closed Riemannian manifold with  $M' \subset M$  a submanifold with or without boundary. Let  $r_{\max}$  be as in Lemma 3.14, and  $r < r_{\max}$ . Then there exists an invariant measure  $d\mu_{k,d}(V)$  on  $\text{Gr}(k, T_c M) = \text{Gr}(k, d)$ , such that for every  $f \in C^\infty(M^{k+1}; \mathbb{R})$*

$$\begin{aligned} & \int_{M^{k+1}} f(\mathbf{y}) \mathbf{1}_r(\mathbf{y}) \mathbf{1}\{c(\mathbf{y}) \in M'\} |\text{dvol}_g(\mathbf{y})| \\ &= \int_{M'} |\text{dvol}_g(c)| \int_0^r du u^{d-k-1} \int_{\text{Gr}(k, T_c M)} d\mu_{k,d}(V) \\ & \quad \times \left( \int_{\mathbb{S}_1(V)^{k+1}} \Upsilon_1^{d-k}(w) f(\exp_c(uw)) \prod_{i=1}^{k+1} \sqrt{|\det(g_{\exp_c(uw_i)})|} |\text{dvol}_{\mathbb{S}_1(V)}(w_i)| \right). \end{aligned}$$

*Proof.* First note that if  $M'$  has positive codimension then both expressions are zero, so assume  $M'$  has zero codimension. If  $\mathbf{y} \in M^{k+1}$  is such that  $E_{r_{\max}}(\mathbf{y}) \neq \emptyset$ , then the induced center  $c(\mathbf{y})$  is uniquely defined (Lemma 3.14); hence:

$$\{\mathbf{y} \in M^{k+1} | c(\mathbf{y}) \in M' \text{ and } E_{r_{\max}}(\mathbf{y}) \neq \emptyset\} = \bigcup_{c \in M'} \mathcal{Y}(c),$$

where  $\mathcal{Y}(c) := \{\mathbf{y} \in M^{k+1} | E_{r_{\max}}(\mathbf{y}) \neq \emptyset \text{ and } c(\mathbf{y}) = c\}$ , and this union is disjoint (by uniqueness of the center). Thus

$$\int_{M^{k+1}} f(\mathbf{y}) \mathbf{1}_r(\mathbf{y}) \mathbf{1}\{c(\mathbf{y}) \in M'\} |\text{dvol}_g(\mathbf{y})| = \int_{M'} |\text{dvol}_g(c)| \int_{\mathcal{Y}(c)} f(\mathbf{y}) \mathbf{1}\{\rho(\mathbf{y}) \leq r\} |\text{dvol}_g(\mathbf{y})|.$$

Now fix  $c \in M'$  with local normal coordinates  $(x^1, \dots, x^d)$ ; for  $\mathbf{y} \in \mathcal{Y}(c)$  with  $\rho(\mathbf{y}) \leq r < r_{\max}$  (this last condition ensures that  $y_i$  can be written as  $\exp_c(v_i)$ ,  $v_i \in T_c M$  and  $|v_i| = u \leq r$ , for all  $i \in [k+1]$ ), we find:

$$\begin{aligned} |\text{dvol}_g(\mathbf{y})| &= |\wedge_{i=1}^{k+1} \text{dvol}_g(y_i)| \\ &= |\wedge_{i=1}^{k+1} \sqrt{|\det(g_{y_i})|} dx^1(y_i) \wedge \dots \wedge dx^d(y_i)| \\ &= \prod_{i=1}^{k+1} \sqrt{|\det(g_{y_i})|} |\wedge_{i=1}^{k+1} dx^1(y_i) \wedge \dots \wedge dx^d(y_i)| \\ &= \prod_{i=1}^{k+1} \sqrt{|\det(g_{y_i})|} |\wedge_{i=1}^{k+1} \text{dvol}_{g_{\mathbb{R}^d}}(v_i)| \\ &= \prod_{i=1}^{k+1} \sqrt{|\det(g_{y_i})|} |\text{dvol}_{g_{\mathbb{R}^d}}(\mathbf{v})|. \end{aligned}$$

Note that we have the polar decomposition:

$$|\mathrm{dvol}_{g_{\mathbb{R}^d}}(\mathbf{v})| = du |\mathrm{dvol}_{S_u(\mathbb{R}^d)}(\mathbf{v})|,$$

and so by the Blaschke-Petkantschin formula, since  $\{v_i : i \in [k+1]\}$  lies in a  $k$ -dimensional subspace  $V \subset T_c M$ , we have

$$|\mathrm{dvol}_{S_u(\mathbb{R}^d)}(\mathbf{v})| = Y_u(\mathbf{v})^{d-k} d\mu_{k,d}(V) |\mathrm{dvol}_{S_u(V)}(\mathbf{v})|,$$

hence, we deduce that

$$|\mathrm{dvol}_g(\mathbf{y})| = \prod_{i=1}^{k+1} \sqrt{|\det(g_{y_i})|} du Y_u(\mathbf{v})^{d-k} d\mu_{k,d}(V) |\mathrm{dvol}_{S_u(V)}(\mathbf{v})|.$$

This shows that for  $c \in M'$

$$\begin{aligned} & \int_{\mathcal{Y}(c)} f(\mathbf{y}) \mathbf{1}\{\rho(\mathcal{Y}) \leq r\} |\mathrm{dvol}_g(\mathbf{y})| \\ &= \int_0^r du u^{d-k-1} \int_{\mathrm{Gr}(k, T_c M)} d\mu_{k,d}(V) \\ & \quad \times \left( \int_{S_1(V)^{k+1}} Y_1^{d-k}(\mathbf{w}) f(\exp_c(u\mathbf{w})) \prod_{i=1}^{k+1} \sqrt{|\det(g_{\exp_c(uw_i)})|} |\mathrm{dvol}_{S_1(V)}(w_i)| \right), \end{aligned}$$

and thus the result follows.  $\blacksquare$

### 5.3 | Blaschke-Petkantschin formula for compact Riemannian manifolds with nonempty boundary

Using the change of variable formula established in Lemma 5.2 for compact closed Riemannian manifolds we attain a formula for Riemannian manifolds with nonempty boundary. We note that our formula is altered near the boundary since we must restrict our integral in the tangent space to those points whose image under the exponential map remains in the manifold.

**Lemma 5.3.** *Suppose that  $M$  is a compact Riemannian manifold with nonempty boundary, let  $DM$  be its double manifold and let  $M' \subset M$  be a submanifold. Then we have:*

$$\begin{aligned} & \int_{M^{k+1}} f(\mathbf{y}) \mathbf{1}_r(\mathbf{y}) \mathbf{1}\{c(\mathbf{y}) \in M'\} |\mathrm{dvol}_g(\mathbf{y})| \\ &= \int_{c \in M'} |\mathrm{dvol}_g(c)| \int_0^r du u^{d-k-1} \int_{\mathrm{Gr}(k, T_c DM)} d\mu_{k,d}(V) \\ & \quad \times \int_{(S_u(V) \cap \exp_c^{-1}(M))^{k+1}} Y_u^{d-k}(\mathbf{v}) f(\exp_c(\mathbf{v})) \prod_{i=1}^{k+1} \sqrt{|\det(g_{\exp_c(v_i)})|} |\mathrm{dvol}_{S_u(V)^{k+1}}(v_i)|. \end{aligned}$$

*Proof.* We have

$$\begin{aligned} & \int_{M^{k+1}} f(\mathbf{y}) \mathbf{1}_r(\mathbf{y}) \mathbf{1}\{c(\mathbf{y}) \in M'\} |\mathrm{dvol}_g(\mathbf{y})| \\ &= \int_{DM^{k+1}} f(\mathbf{y}) \mathbf{1}\{\mathbf{y} \in M^{k+1}\} \mathbf{1}_r(\mathbf{y}) \mathbf{1}\{c(\mathbf{y}) \in M'\} |\mathrm{dvol}_g(\mathbf{y})|; \end{aligned}$$

the double manifold  $DM$  is closed, hence applying the change of variables formula in Lemma 5.2 to the function

$$\mathbf{y} \mapsto f(\mathbf{y})\mathbf{1}\{\mathbf{y} \in M^{k+1}\},$$

we find

$$\begin{aligned} & \int_{DM^{k+1}} f(\mathbf{y})\mathbf{1}\{\mathbf{y} \in M^{k+1}\}\mathbf{1}_r(\mathbf{y})\mathbf{1}\{c(\mathbf{y}) \in M'\}|\mathrm{dvol}_g(\mathbf{y})| \\ &= \int_{c \in M'} |\mathrm{dvol}_g(c)| \int_0^r du u^{d-1} \int_{\mathrm{Gr}(k, T_c DM)} d\mu_{k,d}(V) \\ & \quad \times \int_{\mathbb{S}_u(V)^{k+1}} \Upsilon_u^{d-k}(\mathbf{v}) f(\exp_c(\mathbf{v})) \mathbf{1}\{\exp_c(\mathbf{v}) \in M^{k+1}\} \prod_{i=1}^{k+1} \sqrt{|\det(g_{\exp_c(v_i)})|} |\mathrm{dvol}_{\mathbb{S}_u(V)^{k+1}}(v_i)|, \end{aligned}$$

which gives the desired result.  $\blacksquare$

## 6 | UPPER THRESHOLD

In this section, we produce an upper bound on the expected number of critical points induced by a Poisson process on our manifold. Similar to the argument presented in [4] we utilize an auxiliary radius  $r_0$  and count critical points with critical value in the range  $[r, r_0)$ . The auxiliary radius is chosen to be sufficiently large so that the  $r_0$  neighborhood of the Poisson process covers the manifold w.h.p., but is simultaneously sufficiently small so that the number of critical points with critical values in the range  $[r, r_0)$  is asymptotically zero.

As before, let  $\Lambda = \omega_d n r^d$  denote the expected number of points of a Poisson process with intensity  $n$  inside a  $d$ -dimensional ball of radius  $r$ . If we append a subscript to  $\Lambda$  this corresponds to changing the radius value in this expectation:  $\Lambda_{r_0} = \omega_d n r_0^d$ . Let  $\beta_k(r)$  denote the  $k$ th Betti number of the Čech complex associated to the submanifold  $B_r(\mathcal{P}_n)$ . We shall count critical points of the associated Morse min-type function treating points close to the boundary and far from the boundary separately.

**Proposition 6.1** (Betti number upper bound). *Let  $M$  be a compact unit volume  $d$ -dimensional manifold with boundary. If  $n \rightarrow \infty$  and  $r, r_0 \rightarrow 0$  such that  $\Lambda \rightarrow \infty$ ,  $\Lambda_{r_0} r \rightarrow 0$ ,  $\Lambda_{r_0} r_0^2 \rightarrow 0$ , and  $r_0 \geq r(\frac{\omega_d}{\kappa}(1 + |\log r|))^{1/d}$  (where  $s_{\max} := s_{\max}(v_0)$  for some  $v_0 > 0$  and  $\kappa = \omega_d 2^{-(d+1)}(1 - s_{\max} r_0^2)$ ), then for all  $1 \leq k \leq d-1$ :*

$$\mathbb{E}[\beta_k(r)] \leq \beta_k(M) + O(n\Lambda^k e^{-\Lambda} + n^{1-\frac{1}{d}} \Lambda^k e^{-\frac{1}{2}\Lambda}).$$

To prove the above let us follow [4] in attempting to bound the number of  $k$ -critical points of the distance function associated to a Poisson process on  $M$  with critical value in the range  $[r, r_0)$ . Let us denote this set of critical points by  $C_k^{\rho_M}(r, r_0)$ .

**Lemma 6.2.** *If  $n \rightarrow \infty$  and  $r, r_0 \rightarrow 0$  such that  $\Lambda \rightarrow \infty$ ,  $\Lambda_{r_0} r \rightarrow 0$ ,  $\Lambda_{r_0} r_0^2 \rightarrow 0$  and  $r = o(r_0)$  then for all  $1 \leq k \leq d$ :*

$$\mathbb{E}[C_k^{\rho_M}(r, r_0)] = O(n^{1-\frac{1}{d}} \Lambda^{k-1} e^{-\frac{1}{2}\Lambda} + n\Lambda^{k-1} e^{-\Lambda}).$$

The bounding term  $n\Lambda^{k-1}e^{-\Lambda}$  arises from counting critical points at distance more than  $r_0$  from the boundary of the manifold and the bounding term  $n^{1-\frac{1}{d}}\Lambda^{k-1}e^{-\frac{1}{2}\Lambda}$  arises from counting critical points at distance less than  $r_0$  from the boundary of the manifold.

In what follows let  $f = \mathbf{1}\{x \in M\}$  and  $\mathcal{P}_n$  denote a Poisson process of intensity  $nf$  on  $DM$ . Let us use the shorthand notation  $\partial M_{r_0}$  for an  $r_0$  neighborhood of the boundary, and  $M_{r_0} = M \setminus \partial M_{r_0}$ .

Let us define the following indicator functions which we use to count critical points:

- $h(\mathcal{Y}) = \mathbf{1}\{0 \in \text{Grad}(\mathcal{Y})\}$ .
- $h_{r,r_0}(\mathcal{Y}) = h(\mathcal{Y})\mathbf{1}\{r \leq \rho(\mathcal{Y}) < r_0\}$ .
- $g_{r,r_0}(\mathcal{Y}, \mathcal{P}_n) = h_{r,r_0}(\mathcal{Y})\mathbf{1}\{B(\mathcal{Y}) \cap \mathcal{P}_n = \emptyset\}$ ; where  $B(\mathcal{Y})$  is the open ball  $B_{\rho(\mathcal{Y})}(c(\mathcal{Y}))$
- $g_{r,r_0}^{M_{r_0}}(\mathcal{Y}, \mathcal{P}_n) = g_{r,r_0}(\mathcal{Y}, \mathcal{P}_n)\mathbf{1}\{c(\mathcal{Y}) \in M_{r_0}\}$ .
- $g_{r,r_0}^{\partial M_{r_0}}(\mathcal{Y}, \mathcal{P}_n) = g_{r,r_0}(\mathcal{Y}, \mathcal{P}_n)\mathbf{1}\{c(\mathcal{Y}) \in \partial M_{r_0}\}$ .

We can express the number of critical points as the following sum over generic  $\mathcal{Y} \subset \mathcal{P}_n$ :

$$|C_k^{\mathcal{P}_n}(r, r_0)| = \sum_{|\mathcal{Y}|=k+1} |g_{r,r_0}^{M_{r_0}}(\mathcal{Y}, \mathcal{P}_n)| + |g_{r,r_0}^{\partial M_{r_0}}(\mathcal{Y}, \mathcal{P}_n)|.$$

We can count the critical points in  $M_{r_0}$  and  $\partial M_{r_0}$  separately. Let us denote the critical points in  $\partial M_{r_0}$  with critical value in the interval  $[r, r_0]$  as  $C_{\partial M_{r_0}}$ , and the critical points in  $M_{r_0}$  with critical value in the interval  $[r, r_0]$  as  $C_{M_{r_0}}$ . The counting of the critical points  $C_{M_{r_0}}$  is not affected by the presence of a nontrivial boundary and so we can use the analysis of [4] to bound the term  $\mathbb{E}[|C_{M_{r_0}}|]$  by an order of  $O(n\Lambda^{k-1}e^{-\Lambda})$ .

We require new analysis to calculate an upper bound for the expected number of critical points  $\mathbb{E}[|C_{\partial M_{r_0}}|]$ . Applying Palm Theory (Theorem 2.4) to the uniform Poisson process on  $M$  with intensity  $n$  we yield:

$$\mathbb{E}[|C_{\partial M_{r_0}}|] = \frac{n^{k+1}}{(k+1)!} \mathbb{E}[g_{r,r_0}^{\partial M_{r_0}}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)].$$

Let us now condition on a given sample  $\mathcal{Y}'$  of  $k+1$  points of  $M$  with a view to then integrating over the whole double manifold. Recall  $B(\mathcal{Y}')$  denotes the ball centered at  $c(\mathcal{Y}')$  of radius  $\rho(\mathcal{Y}')$  in  $DM$ , and that the number of points in a given subset of a uniform Poisson process has Poisson distribution with parameter proportional to the volume of the given subset.

$$\mathbb{E}[\mathbf{1}\{B(\mathcal{Y}') \cap \mathcal{P}_n = \emptyset\} | \mathcal{Y}'] = \mathbb{P}(\mathcal{P}_n(B(\mathcal{Y}')) = 0 | \mathcal{Y}') = e^{-n\text{Vol}(B(\mathcal{Y}') \cap M)}.$$

$$\mathbb{E}[|C_{\partial M_{r_0}}|] = \frac{n^{k+1}}{(k+1)!} \int_{DM^{k+1}} \mathbf{1}\{c(\mathbf{y}) \in \partial M_{r_0}\} h_{r,r_0}(\mathbf{y}) e^{-n\text{Vol}(B(\mathbf{y}) \cap M)} |\text{dvol}_g(\mathbf{y})|.$$

For sufficiently small  $r_0$  we have the following lower bound for the volume of a ball with center at distance  $\delta(\mathbf{y})$  from the boundary via Proposition 2.9:

$$\text{Vol}(B(\mathbf{y}) \cap M) \geq \left(\frac{1}{2} - O(r)\right) \text{Vol}(B(\mathbf{y})) + \frac{1}{2} \omega_d \delta(\mathbf{y})^d.$$

Note that  $e^{nO(r)\text{Vol}(B(\mathbf{y}))} = O(1)$  since  $\Lambda_{r_0} r \rightarrow 0$ , and so we have an upper bound:

$$\mathbb{E}[|C_{\partial M_{r_0}}|] \leq C \frac{n^{k+1}}{(k+1)!} \int_{DM^{k+1}} \mathbf{1}\{c(\mathbf{y}) \in \partial M_{r_0}\} h_{r,r_0}(\mathbf{y}) e^{-\frac{n}{2}(\text{Vol}(B(\mathbf{y})) + \omega_d \delta(\mathbf{y})^d)} |\text{dvol}_g(\mathbf{y})|.$$

Using the Blaschke-Petkantschin formula from Section 5 with  $f(\mathbf{y}) = h_{r,r_0}(\mathbf{y})e^{-n\frac{1}{2}(\text{Vol}(B(\mathbf{y})) + \omega_d \delta(\mathbf{y})^d)}$ ,  $c = c(\mathbf{y})$ ,  $u = \rho(\mathbf{y})$ , and  $\mathbf{y} = \exp_c(u\mathbf{w})$  we attain the bound:

$$\begin{aligned} \mathbb{E}[|C_{\partial M_{r_0}}|] &\leq \frac{n^{k+1}}{(k+1)!} \int_{\partial M_{r_0}} |\text{dvol}_g(c)| \int_r^{r_0} du u^{dk-1} \int_{\text{Gr}(k, T_c M)} d\mu_{k,d}(V) \\ &\quad \times \prod_{i=1}^k \left( \int_{\mathbb{S}_1(V)} \sqrt{|\det(g_{\exp_c(uw_i)})|} |\text{dvol}_{\mathbb{S}_1(V)}(w_i)| \right) \Upsilon_1^{d-k}(\mathbf{w}) f(\exp_c(u\mathbf{w})). \end{aligned}$$

Using compactness of the Grassmannian we can bound the final integral over  $(\mathbb{S}_1(V))^k$  by some maximizing subspace. The parallelogram volume  $\Upsilon_1^{d-k}(\mathbf{w})$  is taken over unit vectors and so bounded. Since our manifold is compact and  $g$  is smooth we may bound  $\sqrt{|\det(g_{\exp_c(uw_i)})|}$  by some constant. (For details of the bounding constants see [4].) Thus we yield for some constant  $C$ :

$$\mathbb{E}[|C_{\partial M_{r_0}}|] \leq Cn^{k+1} \int_{\partial M_{r_0}} |\text{dvol}_g(c)| \int_r^{r_0} du u^{dk-1} e^{-n\frac{1}{2}(\text{Vol}(B_u(c)) + \omega_d \delta(c)^d)}.$$

Since  $\Lambda_{r_0} r_0^2 \rightarrow 0$  we may bound the term  $e^{-n\frac{1}{2}(\text{Vol}(B_u(c)) + \omega_d \delta(c)^d)}$  using the second-order Taylor expansion of exp and Lemma 2.5:

$$e^{-n\frac{1}{2}(\text{Vol}(B_u(c)) + \omega_d \delta(c)^d)} \leq e^{-n\frac{1}{2}\omega_d(u^d + \delta(c)^d)} (1 + s_{\max} n \omega_d r_0^{d+2}).$$

Let us apply a change of variables  $s = \frac{u}{r}$  and additionally separate our integral to integrate over the distance of the critical point,  $\delta = \delta(c)$ , to the boundary. We make a change of variables using the diffeomorphism  $\partial M_{r_0} \cong \partial M \times [0, r_0)$  given to us from the Collar Neighborhood Theorem. Since our manifold is compact the Jacobian term introduced will be bounded by a constant which we absorb into the constant term  $C$ .

$$\begin{aligned} \mathbb{E}[|C_{\partial M_{r_0}}|] &\leq Cn^{k+1}(1 + s_{\max} n \omega_d r_0^{d+2}) \int_{\partial M_{r_0}} e^{-n\frac{1}{2}\omega_d \delta^d} |\text{dvol}_g(c)| \int_1^{\frac{r_0}{r}} ds r^{dk} s^{dk-1} e^{-n\frac{1}{2}\omega_d r^d s^d} \\ &\leq Cn\Lambda^k(1 + s_{\max} \Lambda_{r_0} r_0^2) \int_0^{r_0} e^{-n\frac{1}{2}\omega_d \delta^d} d\delta \int_1^{\frac{r_0}{r}} ds s^{dk-1} e^{-\frac{1}{2}\Lambda s^d}. \end{aligned}$$

We may bound the first integral by a term  $O(n^{-\frac{1}{d}})$  by changing variables ( $t = \frac{1}{2}\Lambda\delta$ ) and observing that the integral takes the form of an upper incomplete gamma function:

$$\int_0^{r_0} e^{-n\frac{1}{2}\omega_d \delta^d} d\delta = Cn^{-\frac{1}{d}} \int_0^{\frac{1}{2}\Lambda r_0} t^{\frac{1}{d}-1} e^{-t} dt \leq C\Gamma\left(\frac{1}{d}\right) n^{-\frac{1}{d}}.$$

The last integral also has the form of an upper incomplete gamma function.

$$\Gamma(k, x) = \int_x^\infty t^{k-1} e^{-t} dt = (k-1)! e^{-x} \sum_{i=0}^{k-1} \frac{x^i}{i!}.$$

Let  $t = \frac{1}{2}\Lambda s^d$

$$\int_1^{\frac{r_0}{r}} \left(\frac{1}{2}\Lambda\right)^k s^{dk-1} e^{-\frac{1}{2}\Lambda s^d} ds = \frac{1}{d} \int_{\frac{1}{2}\Lambda}^{\frac{1}{2}\Lambda r_0} t^{k-1} e^{-t} dt = \Gamma\left(k, \frac{1}{2}\Lambda\right) - \Gamma\left(k, \frac{1}{2}\Lambda r_0\right).$$

Absorbing surplus constants into the term  $C$  we attain:

$$\mathbb{E}[|C_{\partial M_{r_0}}|] \leq Cn^{1-\frac{1}{d}}(1 + s_{\max}\Lambda_{r_0}r_0^2) \left( e^{-\frac{1}{2}\Lambda} \sum_{j=0}^{k-1} \frac{(\frac{1}{2}\Lambda)^j}{j!} - e^{-\frac{1}{2}\Lambda_{r_0}} \sum_{j=0}^{k-1} \frac{(\frac{1}{2}\Lambda_{r_0})^j}{j!} \right).$$

Using the assumptions that  $\Lambda_{r_0}r_0^2 \rightarrow 0$ ,  $\Lambda \rightarrow \infty$ , and  $r = o(r_0)$  yields that  $\mathbb{E}[|C_{\partial M_{r_0}}|] = O(n^{1-\frac{1}{d}}\Lambda^{k-1}e^{-\frac{1}{2}\Lambda})$ . Thus we prove Lemma 6.2.

**Proposition.** Having established Lemma 6.2 this proof proceeds like the proof of Proposition 6.1 in [4] mutatis mutandis. Details are provided in Appendix A.

## 7 | LOWER THRESHOLD

In this section, we shall find a lower bound for the expected Betti numbers of the Čech complex  $C(n, r)$ . We utilize the concept of a special type of critical point called a  $\Theta$ -cycle defined in [5]. Such a critical point is guaranteed to induce a nontrivial cycle in the homology of the resulting Čech complex. The lower bound shows that if the convergence of  $\Lambda \rightarrow \infty$  is sufficiently slow then, w.h.p the Čech complex will have Betti numbers larger than that of the manifold. Thus we will attain a lower threshold for  $\Lambda$ .

The paper [4] counts  $\Theta$ -cycles, a special class of critical points  $c(\mathcal{Y})$  on a general compact closed Riemannian manifold with critical values in the range  $(r_1, r]$ . The radius  $r_1$  is chosen to be sufficiently small that there are a large number of critical points with critical value in the range  $(r_1, r]$  but simultaneously sufficiently large that a  $\Theta$ -cycle with critical value  $r_1$  persists and remains a  $\Theta$ -cycle in the Čech complex at radius  $r$ .

The conditions that determine a critical point to be a  $\Theta$ -cycle are all local conditions. As such we can replicate the analysis in [4] to count the  $\Theta$ -cycles in a manifold with boundary which are sufficiently distant from the boundary. It transpires that this lower bound can be improved by counting a collection of cycles which occur close to the boundary, which we shall call  $\Theta$ -like-cycles.

Let us proceed to define the conditions given in [4] that determine a critical point to be a  $\Theta$ -cycle. Let  $\mathcal{Y} \subset \mathcal{P} \subset M$  be a generic subset of a point process on  $M$  inducing center  $c(\mathcal{Y})$  and  $\varepsilon \in (0, 1)$ . Let us denote the closed annulus in  $M$  by:

$$A_\varepsilon(c(\mathcal{Y})) = \overline{B_{\rho(\mathcal{Y})}(c(\mathcal{Y}))} \setminus B_{\varepsilon\rho(\mathcal{Y})}(c(\mathcal{Y})).$$

Intuitively a  $\Theta$ -cycle is formed at critical point  $c$  when an annulus surrounding  $c$  is covered and the critical point  $c$  introduces a  $k$  simplex which cuts across this annulus and introduces an erroneous homological cycle. If our center is close to the boundary, our annulus will be cut by the boundary and so we must modify the argument in [4] to apply to this case.

Lemma 7.1 is a slight modification of that presented in [5] and having taken into consideration that the critical point lies far from the boundary the proof follows identically.

**Lemma 7.1** ([5]). *Let  $\mathcal{Y} \subset \mathcal{P} \subset M$  with  $\mathcal{Y}$  inducing a critical point  $c(\mathcal{Y})$  of index  $k$ . Let us define:*

$$\phi(\mathcal{Y}) = \frac{1}{2\rho(\mathcal{Y})} \min_{v \in \partial\Delta(\mathcal{Y})} |v|,$$

where  $\Delta(\mathcal{Y}) = \text{Grad}(\mathcal{Y})$ . Suppose  $\rho(\mathcal{Y}) < r_{\max}$ ,  $\text{dist}(c(\mathcal{Y}), \partial M) > \rho(\mathcal{Y})$  and  $A_{\phi(\mathcal{Y})}(c(\mathcal{Y})) \subset B_{\rho(\mathcal{Y})}(\mathcal{P})$ , then the critical point  $c(\mathcal{Y})$  generates a new nontrivial cycle in  $H_k(B_{\rho(\mathcal{Y})}(\mathcal{P}))$  which we call a  $\Theta$ -cycle.

The following technical lemma is established by showing that a  $\Theta$ -cycle with critical value in the range  $(r_1, r]$  will remain a  $\Theta$ -cycle at radius  $r$ . We introduce an auxiliary radius  $r_2 > r$  and require that  $B_{r_2}(c(\mathcal{Y})) \cap \mathcal{P} = \mathcal{Y}$ . This condition ensures that the homology class introduced by a  $\Theta$ -cycle is not annihilated by the introduction of another critical point with critical value in the range  $(r_1, r]$ .

**Lemma 7.2** ([4], Lemma 7.3). *Suppose  $r_2 > r > 0$  and that  $r_1 > r \sqrt{1 - \frac{1}{c_g^2}(\frac{r_2}{r} - 1)^2}$  (where  $c_g = c_g(r)$  is a constant determined in [4], Lemma D.1 such that  $c_g(r) \rightarrow 1$  as  $r \rightarrow 0$ ). Let  $\varepsilon \in (0, 1)$  and  $\Theta_k^\varepsilon(r_1, r)$  denote the number of  $\Theta$ -cycles induced by subsets of  $(k + 1)$  points  $\mathcal{Y}$  with the properties:*

$$\rho(\mathcal{Y}) \in (r_1, r], \quad B_{r_2}(c(\mathcal{Y})) \cap \mathcal{P} = \mathcal{Y}, \quad \phi(\mathcal{Y}) \geq \varepsilon, \quad \text{dist}(c(\mathcal{Y}), \partial M) > r_2.$$

For any  $\varepsilon \in (0, 1)$  we have that  $\beta_k(r) \geq \Theta_k^\varepsilon(r_1, r)$ .

Let us use the following notation for indicator functions which track when a subset of  $(k + 1)$  elements  $\mathcal{Y} \subset \mathcal{P}$  induces an element of  $\Theta_k^\varepsilon(r_1, r)$ . Recall the definition of  $h_{r_1, r}(\mathcal{Y})$  which indicates when the collection of points induces a critical point with critical value in the range  $(r_1, r]$ .

- $h_r^\varepsilon(\mathcal{Y}) = h_{r_1, r}(\mathcal{Y}) \mathbf{1}\{\phi(\mathcal{Y}) \geq \varepsilon\} \mathbf{1}\{\text{dist}(c(\mathcal{Y}), \partial M) > r_2\}$ .
- $g_r^\varepsilon(\mathcal{Y}, \mathcal{P}) = h_r^\varepsilon(\mathcal{Y}) \mathbf{1}\{B_{r_2}(c(\mathcal{Y})) \cap (\mathcal{P} \setminus \mathcal{Y}) = \emptyset\} \mathbf{1}\{A_\varepsilon(c(\mathcal{Y})) \subset B_{\rho(\mathcal{Y})}(\mathcal{P})\}$ .

Hence we may write  $\Theta_k^\varepsilon(r_1, r)$  as the sum:

$$\Theta_k^\varepsilon(r_1, r) = \sum_{|\mathcal{Y}|=k+1} g_r^\varepsilon(\mathcal{Y}, \mathcal{P}).$$

**Lemma 7.3** ([4]). *Suppose  $\varepsilon > 0$  is sufficiently small (independent of  $r$  and  $n$ ), and  $r > 0$  such that  $\Lambda \rightarrow \infty, \Lambda r^2 \rightarrow 0$ . Then for suitably chosen  $r_1, r_2$  with  $r_2 > r > r_1 > 0$  we have that:*

$$\mathbb{E}[\Theta_k^\varepsilon(r_1, r)] = \Omega(n\Lambda^{k-2}e^{-\Lambda}).$$

The proof of the above lemma follows almost identically to the proof supplied in [4].

**Proposition 7.4** ([4]). *Suppose  $\varepsilon > 0$  is sufficiently small,  $\omega(n) \rightarrow \infty$ ,  $\gamma \in (0, 1)$  is fixed and  $\Lambda = \log n + (k - 2) \log \log n - \omega(n)$ , then for all  $1 \leq k \leq d - 1$ :*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\Theta_k^\varepsilon(r_1, r) > \gamma \mathbb{E}[\Theta_k^\varepsilon(r_1, r)]) = 1.$$

This proposition is proven using a second moment argument based on Chebyshev's inequality and the details may be found in [4].

**Proposition 7.5.** *Let  $M$  be a unit volume, compact, Riemannian manifold with boundary. Let  $\Lambda = \log n + (k - 2) \log \log n - \omega(n)$  then for all  $1 \leq k \leq d - 1$ :*

$$\lim_{n \rightarrow \infty} \mathbb{P}(H_k(C(n, r)) \cong H_k(M)) = 0.$$

This threshold is identical to that found in [4] and, with the help of the above lemmas and proposition, the proof there can be adapted by only counting  $\Theta$ -cycles sufficiently far from the boundary so that the local considerations do not detect the manifold has a boundary. We will not present this argument in detail but instead prove below our stronger lower bound.



## 7.1 | Lower threshold refined

We shall now adapt the argument of counting  $\Theta$ -cycles to establish that a large number of  $\Theta$ -like-cycles are present close to the boundary. From this we can deduce a greater lower threshold. Let  $\Theta_k^{\varepsilon, \partial M}(r_1, r)$  denote the number of  $\Theta$ -like-cycles of index  $k$  with critical value in the range  $(r_1, r]$ . This type of critical point induces an erroneous homological cycle near the boundary and will be defined thoroughly below (Definition 7.9).

Similarly to the previous section we count  $\Theta$ -like-cycles with critical values in the range  $(r_1, r]$  with auxiliary radius  $r_2$  used as in the previous section to ensure a  $\Theta$ -like-cycle with critical value  $r_1$  remains a  $\Theta$ -like-cycle at scale  $r$ .

**Lemma 7.6.** *Let  $M$  be a unit volume compact Riemannian manifold with boundary. Suppose  $n \rightarrow \infty$  and  $r \rightarrow 0$  such that  $\Lambda \rightarrow \infty$ ,  $\Lambda r^2 \rightarrow 0$  and suppose  $\alpha = \frac{1}{2} + O((\log n)^{-1})$ . Then for all  $1 \leq k \leq d - 1$  we have that:*

$$\mathbb{E}[|\Theta_k^{\varepsilon, \partial M}(r_1, r)|] = \Omega(n\Lambda^{k-2}e^{-\alpha\Lambda}r(\log n)^{-(k+1)}).$$

Heuristically, a  $\Theta$ -cycle is a critical point lying in the center of a surrounding annulus, and thus induces a new cycle. For a critical point close to the boundary, the whole annulus is not contained in the manifold and so we define a  $\Theta$ -like-cycle for this corresponding situation. The annuli which are cut by the boundary form a cup which for certain critical points give rise to nontrivial homological cycles. We shall identify which critical points introduce new homological cycles.

The argument to establish Lemma 7.6 in this section has the following structure. We first define a special class of critical points,  $\Theta$ -like-cycles, and establish that these critical points introduce new spurious homological cycles near the boundary (Lemma 7.8). We then wish to bound from below the expected number of  $\Theta$ -like-cycles. In order for a critical point to be a  $\Theta$ -like-cycle we require that both a partial annulus surrounding the critical point is covered and that the simplex induced by the critical point is approximately tangential to the boundary. We define the partial annulus  $A_\varepsilon^{(\varphi)}(c)$  which is constructed in order that if the points inducing the center all lie in this annulus then the simplex introduced by the critical point is approximately tangential to the boundary. Then for suitable  $\varphi$  the partial annulus is covered w.h.p. and so both conditions are met and the critical point  $c$  thus induces a  $\Theta$ -like-cycle.

**Definition 7.7** (The partial annulus). Let  $c$  be a critical point with radius  $\rho$  such that  $\text{dist}(\partial M, c) \in (0, \rho)$ . Let  $p$  be the closest point to  $c$  on  $\partial M$ . Let  $\mathbf{n} := \exp_c^{-1}p$ , which can be thought of as the vector pointing in the direction of the normal to the boundary at  $p$ . We define the partial annulus:

$$A_\varepsilon^{(\varphi)}(c) = \{x \in A_\varepsilon(c) : \exp_c^{-1}x \text{ makes angle greater than } \pi/2 - \varphi \text{ with } \mathbf{n}\}.$$

See Figure 3.

## 7.2 | $\Theta$ -Like-cycles induce homology

Let us define a function  $\psi(\mathcal{Y}, \varphi)$  for  $\Theta$ -like-cycles that plays the same role as  $\phi$  did for  $\Theta$ -cycles. Recall  $\phi(\mathcal{Y})$  was defined to be half of the distance from the center  $c(\mathcal{Y})$  to the closest point on  $\partial\Delta(\mathcal{Y}) \subset T_{c(\mathcal{Y})}M$  the boundary of the  $k$ -simplex  $\text{Conv}\{\nabla\rho_{y_i}(c)\}_{i=0}^k \subset T_{c(\mathcal{Y})}M$  divided by  $\rho(\mathcal{Y})$ . Let  $\iota_k : C_k^{\partial}(C(\mathcal{P}_n, r)) \rightarrow C_k^{\text{sing}}(B_r(\mathcal{P}_n))$  denote the natural map between the simplicial  $k$ -chains of the Čech complex and the

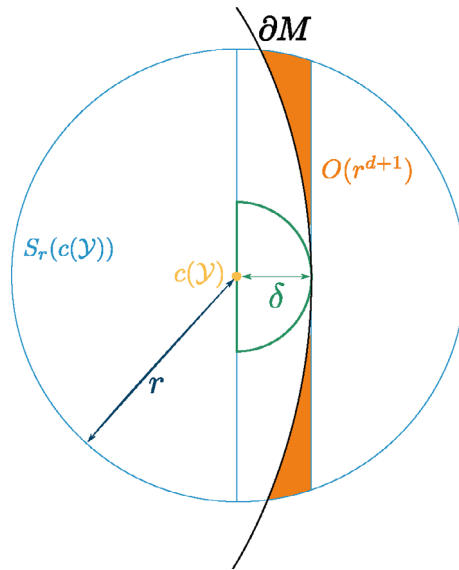


FIGURE 2 The contributions to our lower bound for  $\text{Vol}(B(\mathbf{y}) \cap M)$  are illustrated in this figure. Proposition 2.9 establishes that the shaded region has volume  $O(r^{d+1})$ . We note that for  $\delta(\mathbf{y}) < \tau_M$  ( $\tau_M$  the reach of the manifold) then the half-ball of radius  $\delta(\mathbf{y})$  contributes  $\frac{1}{2} \omega_d \delta(\mathbf{y})^d$  to the volume

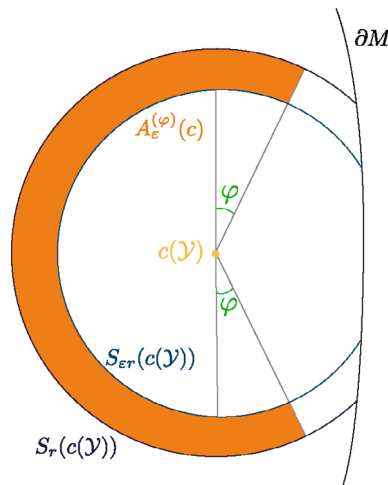


FIGURE 3 A sketch of the partial annulus  $A_\epsilon^{(\phi)}(c)$ . Our calculations in Section 7.3.1 show that if the center  $c(\mathcal{Y})$  is at distance  $\delta \sim \frac{r}{\log n}$  from the boundary and  $\phi \sim (\log n)^{-1}$  then  $A_\epsilon^{(\phi)}(c)$  is covered w.h.p.

singular  $k$ -chains of the topological space  $B_r(\mathcal{P}_n)$ . The value  $\phi(\mathcal{Y})$  is such that  $\iota_{k-1}(\partial\mathcal{Y}) \in A_\phi^{(\beta)}(c(\mathcal{Y}))$ . If also  $A_\phi(c(\mathcal{Y})) \subset B_{\rho(\mathcal{Y})}(\mathcal{P}_n)$  then  $\iota_{k-1}(\partial\mathcal{Y})$  is a boundary of a  $k$ -chain  $\gamma$  contained in  $C_k^{\text{sing}}(B_{\rho(\mathcal{Y})}(\mathcal{P}_n))$ . As a result, the  $k$ -simplex  $\mathcal{Y}$  introduces a new  $k$ -cycle  $(\iota_k(\mathcal{Y}) - \gamma)$  at radius  $\rho(\mathcal{Y})$ .

In order to prove that a  $k$ -simplex  $\mathcal{Y}$  associated to a critical point near the boundary introduces a new degree  $k$  homology class we seek to show that  $\partial\mathcal{Y}$  is the boundary of a  $k$ -chain  $\gamma$  contained in  $B_{\rho(\mathcal{Y})}(\mathcal{P}_n)$  which does not include the simplex  $\mathcal{Y}$ . In the case a critical point  $c(\mathcal{Y})$  is near the boundary, the full annulus  $A_\epsilon(c(\mathcal{Y}))$  is cut by the boundary. This has a couple of important implications. The location of the  $\mathcal{Y}$  about their center relative to the boundary is now important since the annulus is no

longer rotationally symmetric and so the probability of fulfilling the condition that the boundary of the simplex  $\mathcal{Y}$  is covered by  $B_{\rho(\mathcal{Y})}(\mathcal{P}_n)$  is dependent on the position of these points relative to the center. In short, the location of the points  $\mathcal{Y}$  about their center now impacts whether the simplex introduces a new homology class.

We define  $\psi$  to ensure that the condition  $\iota_{k-1}(\partial\mathcal{Y}) \subset A_{\psi}^{(\beta)}(c(\mathcal{Y}))$  is fulfilled and thus  $\partial\mathcal{Y}$  is a boundary of a  $k$ -chain contained in  $C_k^{\text{sing}}(A_{\psi}^{(\beta)}(c(\mathcal{Y}))) \subset C_k^{\text{sing}}(B_r(\mathcal{P}_n))$  with high probability:

$$\psi(\mathcal{Y}, \varphi) = \frac{1}{2} \sup \left( \{ \varepsilon \geq 0 \mid \partial\Delta(\mathcal{Y}) \subset \exp_c^{-1} A_{\varepsilon}^{(\varphi)}(c(\mathcal{Y})) \} \cup \{0\} \right).$$

Note that like  $\phi$ , the term  $\psi(\mathcal{Y}, \varphi)$  is not sensitive to the scale  $\rho(\mathcal{Y})$  in the sense that  $\psi(\mathcal{Y}, \varphi)$  is dependent on the distribution of the points  $\mathcal{Y}$  on the sphere centered at  $c(\mathcal{Y})$  and so has a lower bound which does not scale with the critical value  $\rho(\mathcal{Y})$ . We will sometimes abbreviate  $\psi(\mathcal{Y}, \varphi)$  to just  $\psi$ .

**Lemma 7.8** ( $\Theta$ -like-cycles induce nontrivial homological cycles). *Let  $\mathcal{Y} \subset \mathcal{P} \subset M$  be a set of points inducing an index  $k$  critical point with critical value  $\rho$ ,  $\text{dist}(c(\mathcal{Y}), \partial M) < \rho$ , and  $1 \leq k \leq d-1$ . If  $\psi = \psi(\mathcal{Y}, \varphi) > 0$  and  $A_{\psi}^{(\varphi)}(c(\mathcal{Y})) \subset B_{\rho}(\mathcal{P})$ , then  $c(\mathcal{Y})$  induces a nontrivial cycle in  $H_k(C(\mathcal{P}, \rho))$ .*

The proof of Lemma 7.8 follows the same structure as a proof in [5] that shows  $\Theta$ -cycles induce nontrivial homology. For completeness we include the proof of Lemma 7.8 in Appendix B.

**Definition 7.9** ( $\Theta$ -like-cycle). A critical point  $c(\mathcal{Y})$  which satisfies the conditions of the above lemma is a  $\Theta$ -like-cycle at scale  $\psi(\mathcal{Y}, \varphi)$ .

Note that if  $\partial\Delta(\mathcal{Y})$  intersects  $A_0(c(\mathcal{Y})) \setminus A_0^{(\varphi)}(c(\mathcal{Y}))$  then for all  $\varepsilon > 0$  then  $\psi(\mathcal{Y}, \varphi) = 0$ . We do not want to count the critical points induced by such  $\mathcal{Y}$  since we cannot prove that they introduce nontrivial homological cycle. We sketch an example of a critical point inducing a  $\Theta$ -like-cycle in Figure 4, contrast this with an example of a critical point not inducing a  $\Theta$ -like-cycle.

### 7.3 | $\Theta$ -Like-cycle lower bound computation

Let us modify our indicator functions from the previous section to count  $\Theta$ -like-cycles with critical values in the range  $[r_1, r]$ . We count cycles whose critical point lies at a distance to the boundary in the range  $[\delta, 2\delta]$ , for suitably chosen  $\delta$ .

- $h_r^{\varepsilon, \delta}(\mathcal{Y}) = h_{r_1, r}(\mathcal{Y}) \mathbf{1}\{\psi(\mathcal{Y}, \varphi) \geq \varepsilon\} \mathbf{1}\{\delta \leq \text{dist}(c(\mathcal{Y}), \partial M) \leq 2\delta\}$ .
- $g_r^{\varepsilon, \delta}(\mathcal{Y}, \mathcal{P}) = h_r^{\varepsilon, \delta}(\mathcal{Y}) \mathbf{1}\{B_{r_2}(c(\mathcal{Y})) \cap (\mathcal{P} \setminus \mathcal{Y}) = \emptyset\} \mathbf{1}\{A_{\varepsilon}^{(\varphi)}(c(\mathcal{Y})) \subset B_{\rho(\mathcal{Y})}(\mathcal{P})\}$ .

Recall that the parameter  $r_2$  is used to ensure that the other points of the Poisson process are sufficiently distant that from the  $\Theta$ -like-cycles we count with critical value in the range  $[r_1, r]$  that they not annihilated by the introduction of another simplex with critical value in the range  $[r_1, r]$ . In particular  $r_2$  must satisfy the hypotheses of Lemma 7.2.

*Remark:* Before we embark on our computation of the above probabilities, we give a heuristic argument for our choices of  $\delta$  and  $\varphi$  and why we think that, using our methods, this is optimal.

We shall choose  $\delta$  to maximize our count of  $\Theta$ -like-cycles. The dominant term in the lower bound of the count of  $\Theta$ -like-cycles is the exponential term:  $\exp(-n\alpha \text{Vol}(B_{r_2}(c)))$  where  $\alpha = \frac{1}{2} + O(r + \frac{\delta}{r})$ . In order to match the dominant exponential term of the upper bound of critical points (Section 6) which dictates the leading term of the upper threshold, we require  $\exp(-n\alpha \text{Vol}(B_{r_2}(c))) \sim \exp(-\frac{1}{2}\Lambda)$ . It follows that the optimal choice for  $\delta$  to maximize the lower threshold is  $\delta \sim \frac{r}{\log n}$ .

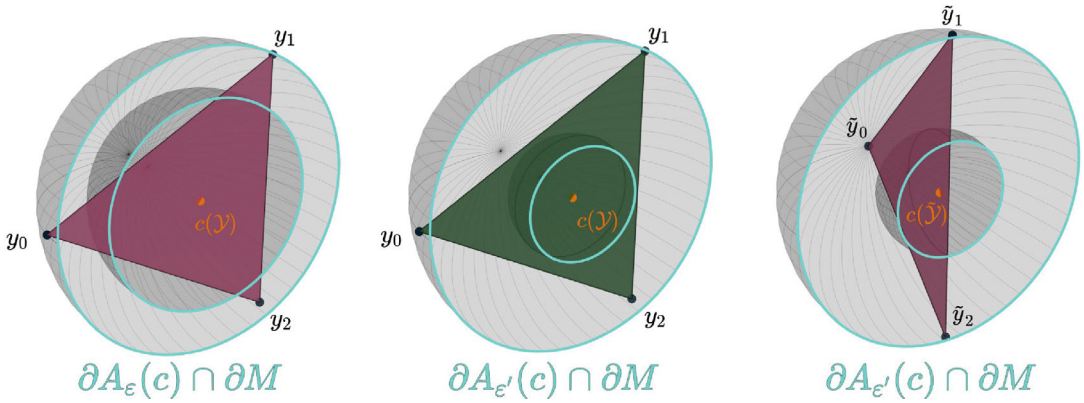


FIGURE 4 A sketch of an index 2 critical point near the boundary which induces a  $\Theta$ -like-cycle (left), and an index 2 critical point near the boundary which does not inducing a  $\Theta$ -like-cycle (right). Suppose  $\varphi > 0$  is such that  $A_{\varepsilon'}^{(\varphi)}(c) \cap \partial M = \emptyset$  for all  $\varepsilon > 0$ . In the leftmost sketch we depict the critical point for the set  $\mathcal{Y} = \{y_0, y_1, y_2\}$ . We highlight the intersection of the annulus  $A_\varepsilon(c)$  with the boundary of the manifold. The boundary of the simplex  $\mathcal{Y}$  is not contained in the annulus  $A_\varepsilon(c)$ . Since the points  $\mathcal{Y}$  lie in a hyperplane  $V$  tangential to the boundary the value  $\psi(\mathcal{Y}, \varphi) > 0$  is dependent on the distribution of the points  $\mathcal{Y}$  on the sphere  $\mathbb{S}_{\rho(\mathcal{Y})}(V)$ , but independent of how close  $c(\mathcal{Y})$  is to the boundary of the manifold. The annulus  $A_{\varepsilon'}^{(\varphi)}(c)$  witnesses that this critical point induces a  $\Theta$ -like-cycle. In contrast, in the rightmost sketch we depict the critical point for the set  $\tilde{\mathcal{Y}} = \{\tilde{y}_0, \tilde{y}_1, \tilde{y}_2\}$ . Since the points  $\tilde{\mathcal{Y}}$  lie in a hyperplane orthogonal to the boundary the boundary of the simplex  $\tilde{\mathcal{Y}}$  is not contained in  $A_{\varepsilon'}^{(\varphi)}(c(\tilde{\mathcal{Y}}))$  for all  $\varepsilon' > 0$  and so this critical point does not induce a  $\Theta$ -like-cycle

The parameter  $\varphi$  is chosen to be as large as possible while satisfying the condition that the partial annulus  $A_\varepsilon^{(\varphi)}(c)$  is covered w.h.p. for a critical point  $c$  with  $\text{dist}(c, \partial M) \in [\delta, 2\delta]$ .

Lemma 7.8 shows that the indicator function  $g_r^{\varepsilon, \delta}(\mathcal{Y}, \mathcal{P})$  is supported on subsets  $\mathcal{Y} \subset \mathcal{P}$  which introduce erroneous cycles. If  $\mathcal{P}_n$  is our Poisson process on  $M$  then we count the  $\Theta$ -like-cycles of index  $k$  with the following sum of indicator functions:

$$\Theta_k^{\varepsilon, \partial M}(r_1, r) = \sum_{|\mathcal{Y}|=k+1} g_r^{\varepsilon, \delta}(\mathcal{Y}, \mathcal{P}_n).$$

### 7.3.1 | Partial $\varepsilon$ -annulus coverage

Let us denote the probability of covering the partial  $\varepsilon$ -annulus of a critical point  $c(\mathbf{y})$  conditioned on the critical point having no points but  $\mathbf{y}$  in an  $r_2$  neighborhood as:

$$p_{\varepsilon, \varphi}(\mathbf{y}) = \mathbb{P}(A_\varepsilon^{(\varphi)}(\mathcal{Y}') \subset B_{\rho(\mathcal{Y}')}(\mathcal{P}_n) \mid \mathcal{Y}' = \mathbf{y}, \mathcal{P}_n \cap B_{r_2}(c(\mathcal{Y}')) = \mathcal{Y}').$$

As in previous sections we use Palm Theory (Theorem 2.4) applied to the Poisson process on  $M$  to attain an integral expression:

$$\begin{aligned} \mathbb{E} \left[ \left| \Theta_k^{\varepsilon, \partial M} \right| \right] &= \frac{n^{k+1}}{(k+1)!} \mathbb{E} [g_r^{\varepsilon, \delta}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)] \\ &= \frac{n^{k+1}}{(k+1)!} \int_{DM^{k+1}} h_r^{\varepsilon, \delta}(\mathbf{y}) p_{\varepsilon, \varphi}(\mathbf{y}) e^{-n \text{Vol}(B_{r_2}(\mathbf{y}) \cap M)} |\text{dvol}_g(\mathbf{y})|. \end{aligned}$$

The expectation depends on  $\varphi$  in the support of the functions  $\mathbf{1}\{\psi(\mathcal{Y}, \varphi) \geq \varepsilon\}$ ,  $\mathbf{1}\{A_\varepsilon^{(\varphi)} \subset B_{\rho(\mathcal{Y})}(\mathcal{P})\}$ . Increasing  $\varphi$ , augments the support of  $\mathbf{1}\{\psi(\mathcal{Y}, \varphi) \geq \varepsilon\}$  and diminishes the support of  $\mathbf{1}\{A_\varepsilon^{(\varphi)} \subset$

$B_{\rho(\mathcal{Y})}(\mathcal{P})\}$ . The term  $\mathbf{1}\{\psi(\mathcal{Y}, \varphi) \geq \varepsilon\}$  enforces that the points  $\mathcal{Y}$  must lie in a hyperplane approximately tangent to the boundary, and the influence of this term on the lower bound is computed in the following section. Meanwhile the term  $\mathbf{1}\{A_\varepsilon^{(\varphi)} \subset B_{\rho(\mathcal{Y})}(\mathcal{P})\}$  insists that the partial annulus is covered by the point process and is tracked by  $p_{\varepsilon, \varphi}(\mathbf{y})$ .

Recall our asymptotic notation  $f(n) \sim g(n)$  if  $f(n) = O(g(n))$  and  $f(n) = \Omega(g(n))$ . Our next lemma shows that if  $\varphi \sim (\log n)^{-1}$  and  $\delta(\mathbf{y}) \sim \frac{r}{\log n}$  then the term  $p_{\varepsilon, \varphi}(\mathbf{y}) \rightarrow 1$  uniformly as  $\Lambda \rightarrow \infty$  for fixed  $\varepsilon$ .

**Lemma 7.10.** *Let  $M$  be a smooth compact manifold with boundary. There exists a constant  $C$  such that if  $\varphi = \frac{C}{\log n}$  and  $\varepsilon \in (0, 1)$  then for all  $c(\mathbf{y}) \in M$  with  $\text{dist}(c(\mathbf{y}), \partial M) \in [\frac{r}{\log n}, \frac{2r}{\log n}]$  the probability  $p_{\varepsilon, \varphi}(\mathbf{y}) \rightarrow 1$  uniformly as  $\Lambda \rightarrow \infty$ .*

*Proof.* First observe that for any angle  $\varphi$  and any center  $c$  we have that  $\text{Vol}(A_\varepsilon^{(\varphi)}(c)) \sim \rho^d(1 - \varepsilon^d)$  and so there is a constant  $C'$  dependent only on the metric  $g$  such that there is an  $\frac{\varepsilon \rho}{10}$ -net of  $A_\varepsilon^{(\varphi)}(c)$ ,<sup>2</sup>  $S$ , with  $|S| \leq C' \frac{1 - \varepsilon^d}{\varepsilon^d} = C'_\varepsilon$ . If for all  $s \in S$  we have  $\mathcal{P}_n \cap B_{\rho(1 - \varepsilon/10)}(s) \neq \emptyset$  then  $A_\varepsilon^{(\varphi)}(c) \subset B_\rho(\mathcal{P}_n)$ . Thus conditioning on the event that  $\{\mathcal{P}_n \cap B_{r_2}(c) = \emptyset\}$  we may bound  $p_{\varepsilon, \varphi}$  from below:

$$p_{\varepsilon, \varphi} \geq 1 - \sum_{s \in S} e^{-n \text{Vol}(B_{\rho(1 - \varepsilon/10)}(s) \setminus B_{r_2}(c))} \geq 1 - C'_\varepsilon \max_{s \in S} e^{-n \text{Vol}(B_{\rho(1 - \varepsilon/10)}(s) \setminus B_{r_2}(c))}.$$

In order to show that  $p_{\varepsilon, \varphi}(\mathbf{y}) \rightarrow 1$  we will show that there is a constant  $C$  independent of  $\mathbf{y}$  such that if  $\varphi = \frac{C}{\log n}$  then the volume  $\text{Vol}(B_{\rho(1 - \varepsilon/10)}(x) \setminus B_{r_2}(c(\mathbf{y})))$  is bounded below by  $\alpha_d \varepsilon^d r^d$  for all  $x \in A_\varepsilon^{(\varphi)}(c(\mathbf{y}))$ , where  $\alpha_d > 0$  is a constant depending on the dimension  $d$ .

For the next part of the proof we will show there is a constant  $C$  independent of  $\mathbf{y}$  such that for all  $c(\mathbf{y}) \in M$  with  $\text{dist}(c(\mathbf{y}), \partial M) \in [\frac{r}{\log n}, \frac{2r}{\log n}]$  and for all  $x \in A_\varepsilon^{(\frac{C}{\log n})}(c(\mathbf{y}))$  the geodesic emanating from  $c(\mathbf{y})$  passing through  $x$  of length  $2r$  is completely contained in  $M$  and does not intersect  $\partial M$ .

Fix some critical point  $c = c(\mathbf{y})$  with  $\text{dist}(c(\mathbf{y}), \partial M) \in [\frac{r}{\log n}, \frac{2r}{\log n}]$  and let  $p \in \partial M$  be the unique point on the boundary closest to  $c$ , and let  $\mathbf{n} \in T_c M$  be the normal vector at such that  $\exp_c(\mathbf{n}) = p$ . This induces the tangent hyperplane  $W$  at  $c$ ;  $W \subset T_c M$  such that  $\mathbf{n} \cdot W = 0$  (see Figure 5A).

Let us define  $U_c = \{\mathbf{u} \in T_c M : \exp_c(\mathbf{u}) \in \partial M \text{ and } \|\mathbf{u}\| = 2r\} = \exp_c^{-1}(S_{2r}(c) \cap \partial M)$  and define  $\varphi_{\mathbf{u}}$  to be the acute angle formed between  $\mathbf{u}$  and  $W$ , and let  $\varphi = \min_{\mathbf{u} \in U_c} \varphi_{\mathbf{u}}$ . Suppose the minimum is attained in direction  $\mathbf{u}$  so that  $\varphi = \varphi_{\mathbf{u}}$  and the point  $q = \exp_c(\mathbf{u})$  lies on the boundary and is distance  $2r$  from  $c(\mathbf{y})$ .

We form a right angled triangle in  $T_c M$  by adding a construction line perpendicular to  $\mathbf{n}$  passing through  $\mathbf{u}$  and intersecting the line  $\{t\mathbf{n} : t \in \mathbb{R}\}$  at the point  $t_0 \mathbf{n}$ . This triangle has vertices  $\mathbf{0}, \mathbf{u}, t_0 \mathbf{n}$  and the angle at vertex  $\mathbf{u}$  is  $\varphi$  (see Figure 5B). Thus  $\sin \varphi = \frac{\|t_0 \mathbf{n}\|}{\|\mathbf{u}\|} = \frac{\|t_0 \mathbf{n}\|}{2r}$ .

We can compute  $\|t_0 \mathbf{n}\|$  using the Euclidean metric on the tangent space  $T_c M$  as an approximation for  $g$  the metric of the manifold. In the local  $2r$ -neighborhood of  $c$  the Euclidean metric differs from  $g$  by a term of order  $O(r^2)$  (Equation (1)).

We compute that  $\|t_0 \mathbf{n}\| = \text{dist}(c(\mathbf{y}), \partial M) \pm d(q - p, T_p M) + O(r^2)$  (see Figure 5b). Using Theorem 2.8 we bound  $\text{dist}(q - p, T_p M) \leq \frac{\|q - p\|^2}{2\tau_M} \leq \frac{4r^2}{2\tau_M}$ . By our standing assumption that  $\Lambda \rightarrow \infty$ , we have  $\Lambda = \omega_d n r^d = O(\log n)$ , hence  $r = O((\frac{\log n}{n})^{1/d})$ , so in particular  $r^2 = o(\frac{r}{\log n})$  and hence  $\|t_0 \mathbf{n}\| = \text{dist}(c(\mathbf{y}), \partial M) + O(r^2) \sim \frac{r}{\log n}$ . It follows that  $\sin \varphi \sim \varphi \sim \frac{1}{\log n}$ . Hence there exists a constant  $C$  such

<sup>2</sup>Thus every point in  $A_\varepsilon^{(\varphi)}(c)$  is at most  $\frac{\varepsilon \rho}{10}$  far away from a point in the set  $S$ .

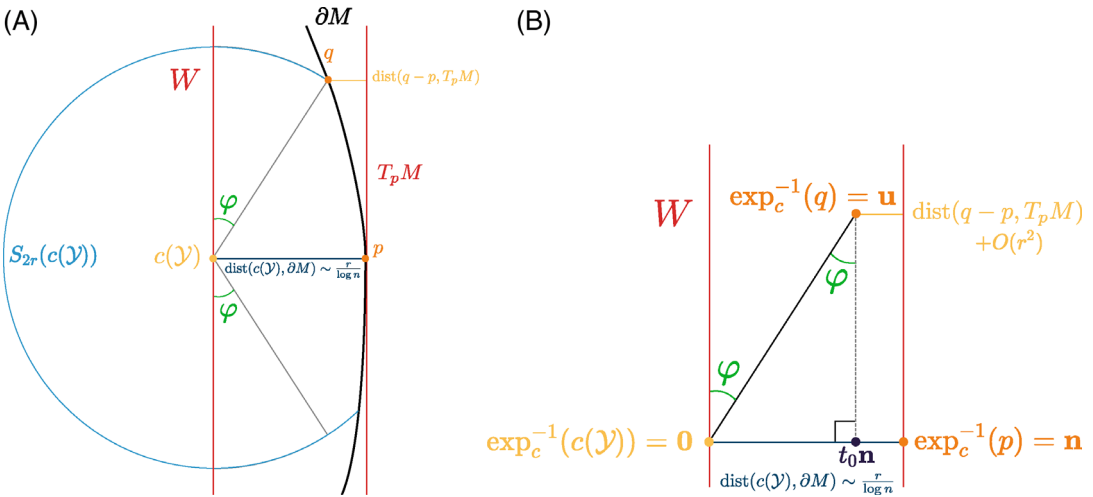


FIGURE 5 We sketch the geometric constructions present in the argument of Section 7.3.1. (A) We sketch  $p \in \partial M$  the closest point to  $c(\mathcal{Y})$  lying on the boundary and the point  $q \in \partial M \cap S_{2r}(c(\mathcal{Y}))$  which is such that the angle  $\varphi$  between  $\mathbf{u} = \exp_c^{-1}(q)$  and the tangent hyperplane  $W \subset T_p M$  is minimal over all points in the intersection  $\partial M \cap S_{2r}(c(\mathcal{Y}))$ . (B) We sketch the preimages of  $p$  and  $q$  in the tangent space  $T_p M$  and use the fact that the metric on the tangent space deviates from the Riemannian metric on  $M$  by a term of order  $O(r^2)$  in order to compute that  $\varphi \sim \sin \varphi \sim \frac{\|t_0 \mathbf{n}\|}{\|\mathbf{u}\|} = \frac{\text{dist}(c(\mathcal{Y}), \partial M) \pm d(q-p, T_p M) + O(r^2)}{2r} \sim \frac{1}{\log n}$

that for every center  $c(\mathbf{y})$  with  $\text{dist}(c(\mathbf{y}), \partial M) \in [\frac{r}{\log n}, \frac{2r}{\log n}]$ , for any  $x \in A_{\epsilon}^{(\frac{C}{\log n})}(c(\mathbf{y}))$ , the radial geodesic emanating from  $c(\mathbf{y})$  passing through  $x$  of length  $2r$  does not intersect  $\partial M$ .

Next let us show that for all  $x \in A_{\epsilon}^{(\frac{C}{\log n})}(c)$  the volume  $\text{Vol}(B_{r(1-\epsilon/10)}(x) \setminus B_{r_2}(c)) \geq \alpha_d \epsilon^d r^d$ , where  $\alpha_d > 0$  is a constant depending on the dimension  $d$  only. We have depicted the constructions in this argument in Figure 6 for clarity. Let  $\mathbf{v}$  be a unit vector making acute angle  $\frac{C}{\log n}$  with the tangent hyperplane  $W$  normal to  $\mathbf{n}$ .

In order to show that  $\text{Vol}(B_{r(1-\epsilon/10)}(x) \setminus B_{r_2}(c)) \geq \alpha_d \epsilon^d r^d$  for all  $x \in A_{\epsilon}^{(\frac{C}{\log n})}(c)$ , it suffices to show this property for  $x = \exp_c(\epsilon r \mathbf{v})$  since the volume  $\text{Vol}(B_{r(1-\epsilon/10)}(x) \setminus B_{r_2}(c))$  increases as  $x$  moves away from  $c$  in the radial direction.

Let  $t = r_2 + \frac{1}{2}((1 + \frac{9\epsilon}{10})r - r_2)$  and consider the volume of  $B_{(t-r_2)}(v)$  the radius  $t - r_2$  ball centered at  $v = \exp_c(t\mathbf{v})$ . At least half of this ball does not intersect the boundary and since  $r_2 \rightarrow r$  as  $n \rightarrow \infty$  we may assume  $t - r_2 \geq \frac{\epsilon r}{20}$  thus the volume of this ball is at least  $\alpha_d \epsilon^d r^d$ , for sufficiently small  $\alpha_d > 0$ . By construction  $B_{(t-r_2)}(v)$  is completely contained in  $B_{r(1-\epsilon/10)}(x) \setminus B_{r_2}(c)$  as  $\rho(c, v) \geq r_2 + (t - r_2)$ , and  $\rho(x, v) + (t - r_2) = 2t - \epsilon r - r_2 = r(1 - \frac{\epsilon}{10})$ . Thus  $\text{Vol}(B_{r(1-\epsilon/10)}(x) \setminus B_{r_2}(c)) > \alpha_d \epsilon^d r^d$ .

Hence given we have shown for any  $c(\mathbf{y})$  that  $\text{Vol}(B_{r(1-\epsilon/10)}(x) \setminus B_{r_2}(c)) \geq \alpha_d \epsilon^d r^d$  for all  $x \in A_{\epsilon}^{(\frac{C}{\log n})}(c)$  then it follows that:

$$p_{\epsilon, \varphi} \geq 1 - C'_\epsilon \max_{s \in S} e^{-n \text{Vol}(B_{r(1-\epsilon/10)}(s) \setminus B_{r_2}(c))} \geq 1 - C'_\epsilon e^{-n \alpha_d \epsilon^d r^d} = 1 - C'_\epsilon e^{-\alpha_d \epsilon^d \Lambda / \omega_d} \rightarrow 1$$

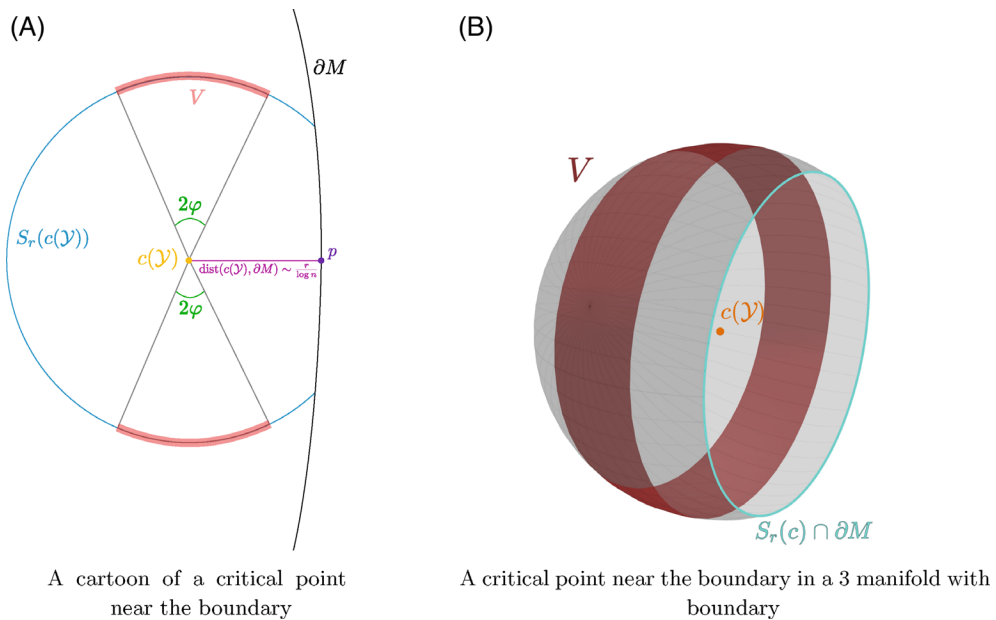
uniformly as  $\Lambda \rightarrow \infty$  for fixed  $\epsilon$ . ■

Since  $p_{\epsilon, \varphi}(\mathbf{y}) \rightarrow 1$  uniformly as  $\Lambda \rightarrow \infty$  for fixed  $\epsilon$ , we may remove the term  $p_{\epsilon, \varphi}$  from the integral and replace it by a constant.

$$\mathbb{E}[|\Theta_k^{\epsilon, \partial M}|] \geq C \frac{n^{k+1}}{(k+1)!} \int_{DM^{k+1}} h_r^{\epsilon, \delta}(\mathbf{y}) e^{-n \text{Vol}(B_{r_2}(c(\mathbf{y})) \cap M)} |\text{dvol}_g(\mathbf{y})|.$$







**FIGURE 7** (A) We sketch a critical point  $c(\mathcal{Y})$  close to the boundary to highlight the important quantities in computing a lower bound for the number of  $\Theta$ -like-cycles. We highlight a subset  $V \subset S_r(c(\mathcal{Y}))$  such that if every point of  $\mathcal{Y}$  lies in  $V$  the simplex introduced by the critical point  $c$  is a  $k$ -dimensional lid on the sliced-annulus centered around  $c$  and so induces a  $\Theta$ -like cycle. The corresponding subset  $\mathcal{V} \subset \text{Gr}(k, T_c M)$  consists of those  $k$ -planes whose intersection with the sphere lies in  $V \subset S_r(c(\mathcal{Y}))$ . (B) We give a concrete example for  $M$  a 3-manifold with boundary. We highlight where the sphere intersects the boundary. If the points  $\mathcal{Y}$  inducing an index 2 critical point at  $c$  lie in the highlighted subset they induce a  $\Theta$ -like-cycle as in Figure 4

Taking sufficiently small  $r$  we may assume the determinant term is arbitrarily close to 1. Moreover given  $\psi(\mathbf{y}, \varphi) \geq \varepsilon$  it is clear the volume  $Y(\mathbf{w})$  is bounded below by a term of order  $\text{dist}(A_\varepsilon^{(\varphi)}, c(\mathcal{Y}))^k = \Omega(\varepsilon^k)$ .

Let us calculate a lower bound for the volume of the subspace of the Grassmannian  $\text{Gr}(k, d)$  for which the points inducing our center must lie in order that  $c(\mathcal{Y})$  induces a  $\Theta$ -like cycle of order  $k$  (see Figure 7). Since we are using the invariant measure  $\mu_{k,d}$  on  $\text{Gr}(k, d)$  as in the Blaschke-Petkantschin formula we can appeal to Theorem 13.1.5 from *Stochastic and Integral Geometry* [26], and compute a bound for this volume using the action of  $SO(d)$  on  $\text{Gr}(k, d)$ .

For sufficiently small  $r$  there is a unique closest point to  $c$  on the boundary  $p \in \partial M$ . Let  $\mathbf{n} \in T_c M$  be such that  $\exp_c(\mathbf{n}) = p$  and let  $e_d = \mathbf{n}/\|\mathbf{n}\|$  be the unit vector *normal* to the boundary at  $c$ .

Define  $V$  to be a subset of the sphere in the tangent space at  $c$  as sketched in Figure 7:

$$V := \{\mathbf{v} \in T_c M : \|\mathbf{v}\| = 1, |\langle \mathbf{v}, \mathbf{n} \rangle| \leq \sin(\varphi)\} \subset T_{c(\mathcal{Y})} M.$$

This induces a subset of the Grassmannian:

$$\mathcal{V} := \{W \in \text{Gr}(k, d) : W = \langle v_1, \dots, v_k \rangle, v_i \in V\} \subset \text{Gr}(k, d)$$

For our suitably chosen  $\varphi$ , a critical point induced by points lying in  $V$  will introduce a simplex which lies approximately tangential to the boundary and thus will induce a  $\Theta$ -like-cycle. In order to calculate the volume of the associated subset of the Grassmannian  $\mathcal{V}$  we shall use Theorem 7.11.

**Theorem 7.11** (Theorem 13.1.5 [26]). *Let  $G$  be a compact group operating continuously and transitively on a Hausdorff space  $E$ , and suppose  $G$  and  $E$  have countable bases. Let  $\nu$  be a Haar measure on  $G$  with  $\nu(G) = 1$ . Then there exists a unique  $G$ -invariant Borel measure  $\rho$  on  $E$  with  $\rho(E) = 1$  defined by:*

$$\rho(B) = \nu(\{g \in G : g \cdot x_0 \in B\}), \quad B \in \mathcal{B}(E)$$

for arbitrary  $x_0 \in E$ .

We observe that the compact group  $SO(d)$  acts continuously and transitively on  $\text{Gr}(k, d)$ . We can parametrize  $SO(d)$  with  $\binom{d}{2}$  angles  $\{\phi_{i,j}\}_{1 \leq i < j \leq d-1}$  [8,14].

Let us briefly recall this parametrization. Denote the augmentation of vectors and matrices via a superscript  $a$ :

$$R^a = \begin{bmatrix} R & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (\nu_1, \dots, \nu_l)^a = (\nu_1, \dots, \nu_l, 0).$$

Recall the hyperspherical coordinate system of the unit  $d$ -sphere  $S^d$  given by  $d$  angles  $\theta_1, \dots, \theta_d, \Sigma_d : \mathbb{R}^d \rightarrow S^d$ . Let  $\pi_i$  denote the projection onto the  $i$ th coordinate.

$$\pi_i(\Sigma_d(\theta_1, \dots, \theta_d)) = \begin{cases} \sin \theta_1 \dots \sin \theta_d & \text{if } i = 1 \\ \cos \theta_{i-1} \sin \theta_i \dots \sin \theta_d & \text{if } 2 \leq i \leq d+1. \end{cases}$$

We have an orthonormal basis  $\{\mathbf{u}_{d,1}, \dots, \mathbf{u}_{d,d}, \Sigma_d(\theta_1, \dots, \theta_d)\}$ , where  $\mathbf{u}_{d,i}$  is the unit vector in the direction  $\frac{\partial \Sigma_d(\theta_1, \dots, \theta_d)}{\partial \theta_i}$ . Let the rotation  $R_{d+1}(\theta_1, \dots, \theta_d) \in SO(d+1)$  be the matrix:

$$R_{d+1}(\theta_1, \dots, \theta_d) = \begin{bmatrix} \mathbf{u}_{d,1} & \dots & \mathbf{u}_{d,d} & \Sigma_d(\theta_1, \dots, \theta_d) \end{bmatrix}.$$

Observe that  $R_{d+1}$  maps the standard basis vector  $\mathbf{e}_{d+1}$  to the point  $\Sigma_d(\theta_1, \dots, \theta_d) \in S^d$ .

We inductively construct  $\Phi_d \in SO(d)$  from  $\binom{d}{2}$  angles  $\{\phi_{i,j}\}_{1 \leq i < j \leq d-1}$  as follows:

$$\Phi_2(\phi_{1,1}) = \begin{bmatrix} \cos \phi_{1,1} & \sin \phi_{1,1} \\ -\sin \phi_{1,1} & \cos \phi_{1,1} \end{bmatrix}.$$

$$\Phi_d(\{\phi_{i,j}\}_{1 \leq i < j \leq d-1}) = R_d(\phi_{1,d-1}, \dots, \phi_{d-1,d-1})(\Phi_{d-1}(\{\phi_{i,j}\}_{1 \leq i < j \leq d-2}))^a.$$

$$\phi_{1,j} \in [0, 2\pi], \quad \phi_{i,j} \in [0, \pi] \quad \text{for } 2 \leq i \leq d-1, 1 \leq j \leq d-1.$$

The Haar measure on  $SO(d)$  is then given up to normalizing constant by:

$$d\nu = \left( \prod_{1 \leq i < j \leq d-1} \sin^{i-1} \phi_{i,j} \right) d\phi_{1,1} \dots d\phi_{d-1,d-1}.$$

Let  $P_k$  denote the projection onto the  $k$ -plane  $x_0$ . Observe that if  $\|P_k \Phi_d e_d\| \leq \sin \varphi$  then  $\Phi_d x_0 \in \mathcal{V}$ . Recall  $x_0 = \langle e_1, \dots, e_k \rangle$ , where  $e_i$  are standard basis vectors. Using the inductive construction we

observe that  $\|P_k \Phi_d e_d\| = \|P_k \Sigma_{d-1}(\phi_{1,d-1}, \dots, \phi_{d-1,d-1})\| \leq \sin \phi_{k,d-1}$  since the first  $k$  coordinates of  $\Sigma_{d-1}(\phi_{1,d-1}, \dots, \phi_{d-1,d-1})$  contain a factor of  $\sin \phi_{k,d-1}$ .

Thus we attain the following estimate for  $\mu_{k,d}(\mathcal{V})$ :

$$\begin{aligned} \mu_{k,d}(\mathcal{V}) &= \int_{\Phi \in SO(d)} \mathbf{1}\{\Phi \cdot x_0 \in \mathcal{V}\} d\nu \sim \int_{\phi_{i,j}} \mathbf{1}\{\Phi \cdot x_0 \in \mathcal{V}\} \left( \prod_{1 \leq i \leq j \leq d-1} \sin^{i-1} \phi_{i,j} \right) d\phi_{1,1} \dots d\phi_{d-1,d-1} \\ &\geq C \int_0^{\varphi/2} \sin^{k-1} \phi_{k,d-1} d\phi_{k,d-1} = \Omega(\varphi^k). \end{aligned}$$

Thus we have a lower bound estimate for the Grassmannian volume in terms of the angle  $\varphi$ .

Given a critical point  $c$  induced by  $\mathcal{Y} \subset A_\varepsilon^{(\varphi)}(c)$  and associated critical simplex  $\Delta$ , for sufficiently small  $\varepsilon$  we have  $\partial\Delta \subset A_\varepsilon^{(\varphi)}(c)$  and thus  $c$  induces a  $\Theta$ -like cycle.

We can compute a lower bound for the support of the term  $\mathbf{1}\{\psi(\mathbf{y}, \varphi) \geq \varepsilon\}$  by observing that:

$$\mathbf{1}\{\psi(\mathbf{y}, \varphi) \geq \varepsilon\} \geq \mathbf{1}\{\phi(\mathbf{y}) \geq \varepsilon\} \mathbf{1}\{\mathbf{v} \in V \subset T_{c(\mathbf{y})}M\}.$$

The support of the term  $\mathbf{1}\{\mathbf{v} \in V \subset T_{c(\mathbf{y})}M\}$  may be bounded below using the Grassmannian volume calculation above and as observed in [4] the support of the term  $\mathbf{1}\{\phi(\mathbf{y}) \geq \varepsilon\}$  is bounded below by a constant for fixed  $\varepsilon$ . Hence we conclude that for sufficiently small  $\varepsilon$  we have  $D_{k,\varphi}^\varepsilon = \Omega((\log n)^{-k})$ .

We can calculate a lower bound using this estimate and with  $r_1 = r(1 - \frac{\varepsilon^2}{2c_g^2})$ ,  $r_2 = r(1 + \xi)$ :

$$\begin{aligned} \mathbb{E} \left[ \left| \Theta_k^{\varepsilon, \partial M} \right| \right] &\geq C(1 - c_R r^2)(1 + s_{\min} \Lambda_{r_2} r^2) D_{k,\varphi}^\varepsilon n^{k+1} e^{-\alpha \Lambda_{r_2}} \int_{\partial M_{[\delta, 2\delta]}} |\text{dvol}_g(c)| \int_{r_1}^r du u^{d-1} \\ &\geq C D_{k,\varphi}^\varepsilon n^{k+1} e^{-\alpha \Lambda_{r_2}} \delta r^{dk} \int_{r_1/r}^1 s^{dk} ds \\ &\geq C D_{k,\varphi}^\varepsilon n \Lambda^k e^{-\alpha \Lambda_{r_2}} \delta \xi^2. \end{aligned}$$

For  $\xi = \Lambda^{-1}$  we observe that asymptotically  $e^{-\alpha \Lambda_{r_2}} = \Omega(e^{-\alpha \Lambda})$  and so:

$$\begin{aligned} \mathbb{E} \left[ \left| \Theta_k^{\varepsilon, \partial M} \right| \right] &\geq C D_{k,\varphi}^\varepsilon n \Lambda^{k-2} e^{-\alpha \Lambda_{r_2}} \delta \\ &\geq C n \Lambda^{k-2} e^{-\alpha \Lambda} r (\log n)^{-(k+1)}. \end{aligned}$$

Thus we see that  $\mathbb{E}[\left| \Theta_k^{\varepsilon, \partial M} \right|] = \Omega(n \Lambda^{k-2} e^{-\alpha \Lambda} r (\log n)^{-(k+1)})$

In order to complete the argument for the lower threshold we need to verify that the second moment arguments in [4] carry over to our modified notion of a  $\Theta$ -like-cycle, and so w.h.p. the number of  $\Theta$ -like-cycles is bounded below by the same regime as the expected number of  $\Theta$ -like-cycles. We compute the second moments in Appendix C and verify that indeed w.h.p. the number of  $\Theta$ -like-cycles is bounded below by the same regime as the expected number of  $\Theta$ -like-cycles.

## 8 | MAIN RESULT

Collecting the results contained in the previous sections we attain thresholds in terms of  $\Lambda$  for the  $k$ th homological connectivity of a compact manifold with boundary  $M$ . For  $\Lambda$  greater than the upper

threshold we recover the  $k$ th homology of  $M$  with high probability, and for  $\Lambda$  less than the lower threshold we do not recover the  $k$ th homology of  $M$  with high probability.

A recent coverage result from [6] yields a sharp coverage threshold for Riemannian manifolds with boundary. For  $\Lambda = (2 - \frac{2}{d}) \log n + 2(d - 2) \log \log n + w(n)$  the manifold is covered with high probability. Our result shows that the homological connectivity threshold for lower homology groups occurs before coverage and that the first-order term for homological connectivity is  $(2 - \frac{2}{d}) \log n$ .

**Theorem 8.1** (Homological connectivity of a Riemannian manifold with boundary). *Let  $M$  be a unit volume compact Riemannian manifold with smooth nonempty boundary. Let  $d \geq 2$  be the dimension of  $M$ ,  $\Lambda = \omega_d n r^d$  and  $\mathcal{P}_n$  a Poisson process of intensity  $n$  on  $M$ . Suppose  $w(n)$  is any function with  $w(n) \rightarrow \infty$  as  $n \rightarrow \infty$ . Then for  $1 \leq k \leq d - 1$*

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

*Proof.* (Upper threshold). Let  $r_0 = r(\frac{\omega_d}{\kappa}(1 + |\log r|))^{1/d}$  so that the conditions of Lemma 6.2 are met. Asymptotically  $\Lambda_{r_0} \geq 2 \log n$ . Using our asymptotic coverage result Theorem 4.2 we observe that  $M \subset B_{r_0}(\mathcal{P}_n)$  w.h.p. and so for sufficiently small  $r_0$ , by the Nerve Lemma ( $H_k(C(n, r_0)) \cong H_k(M)$ ) w.h.p.

Moreover for  $\Lambda$  as described in the upper threshold we observe that  $n\Lambda^k e^{-\Lambda}, n^{1-\frac{1}{d}} \Lambda^k e^{-\frac{1}{2}\Lambda} \rightarrow 0$  so by Lemma 6.2 the expected number of  $k$ -critical and  $(k + 1)$ -critical points with critical value in the range  $(r, r_0]$  goes to zero:

$$\Lambda = \left(2 - \frac{2}{d}\right) \log n + 2k \log \log n + w(n), \Rightarrow n\Lambda^k e^{-\Lambda} = O\left(\frac{n(\log n)^k}{n^{(2-2/d)(\log n)^{2k} e^{w(n)}}}\right) \rightarrow 0.$$

$$\Lambda = \left(2 - \frac{2}{d}\right) \log n + 2k \log \log n + w(n), \Rightarrow n^{1-\frac{1}{d}} \Lambda^k e^{-\frac{1}{2}\Lambda} = O\left(\frac{n^{1-\frac{1}{d}} (\log n)^k}{n^{1-1/d} (\log n)^{2k} e^{w(n)}}\right) \rightarrow 0.$$

Using Morse theory for manifolds with boundary we yield that  $H_k(C(n, r)) \cong H_k(C(n, r_0))$  w.h.p. since the probability of there being a critical point in the range  $(r, r_0]$  tends to zero, and thus  $H_k(C(n, r)) \cong H_k(M)$  w.h.p. for  $\Lambda = (2 - \frac{2}{d}) \log n + 2k \log \log n + w(n)$ . ■

*Proof.* (Lower threshold). Recall that by Lemma 7.6 we have  $\mathbb{E}[\Theta_k^{\epsilon, \partial M}] = \Omega(e^{-\alpha\Lambda} n \Lambda^{k-2} r(\log n)^{-(k+1)})$  for  $\alpha = \frac{1}{2} + O((\log n)^{-1})$ . Second moment calculations in Appendix C verify that w.h.p.  $\Theta_k^{\epsilon, \partial M} > \frac{1}{2} \mathbb{E}[\Theta_k^{\epsilon, \partial M}]$  and thus w.h.p.  $|\beta_k(r)| = \Omega(e^{-\alpha\Lambda} n \Lambda^{k-2} r(\log n)^{-(k+1)})$ . Recall that by our standing assumptions  $\Lambda \rightarrow \infty$ , and we must have  $w(n) = O(\log n)$ . If  $w(n) = o(\log n)$ , then with  $\Lambda = (2 - \frac{2}{d}) \log n + 2(k - 2 - (k + 1 - \frac{1}{d})) \log \log n - w(n)$ :

$$e^{-\alpha\Lambda} n \Lambda^{k-2} r(\log n)^{-(k+1)} \sim \frac{n^{1-\frac{1}{d}} (\log n)^{k-2+\frac{1}{d}-(k+1)}}{e^{\frac{\Lambda}{2}}} \sim \frac{n^{1-\frac{1}{d}} (\log n)^{k-2+\frac{1}{d}-(k+1)}}{n^{1-\frac{1}{d}} (\log n)^{k-2+\frac{1}{d}-(k+1)} e^{-\frac{w(n)}{2}}} \rightarrow \infty.$$

Else, we have  $w(n) = \Omega(\log n)$ . Note also that we must have in any case  $\Lambda(\log n)^{-1} = O(1)$ . We then have

$$\begin{aligned} e^{-\alpha\Lambda} n \Lambda^{k-2} r(\log n)^{-(k+1)} &= \omega_d^{-1/d} n^{1-1/d} \Lambda^{k-2+1/d} (\log n)^{-(k+1)} e^{-\frac{1}{2}\Lambda+O(1)} \\ &= \omega_d^{-1/d} (\Lambda)^{k-2+1/d} (\log n)^{-(k+1)} (\log n)^{3-1/d} e^{\frac{1}{2}w(n)+O(1)} \\ &= \omega_d^{-1/d} \left(\frac{\Lambda}{\log n}\right)^{k-2+1/d} e^{\frac{1}{2}w(n)+O(1)} \\ &\geq \omega_d^{-1/d} \exp(-(k-2+1/d) \log \log n + \frac{1}{2}w(n) + O(1)) \rightarrow \infty, \end{aligned}$$

where we use, in the last line, that  $\Lambda \rightarrow \infty \geq 1$  and that  $w(n) = \Omega(\log n) = \omega(\log \log n)$ .

Thus for  $\Lambda$  in this regime the Betti numbers tend to infinity w.h.p. and so we do not recover the homology of  $M$ . ■

## 9 | CONCLUSION

Having adapted the techniques from [4] in order to count critical points near to the boundary we have attained thresholds similar to those of the homological connectivity theorem from [4], applicable to compact, closed Riemannian manifolds.

**Theorem 9.1** (Homological connectivity thresholds for compact manifolds without boundary [4]). *Let  $M$  be a unit volume compact, Riemannian manifold without boundary. Suppose that as  $n \rightarrow \infty$ ,  $w(n) \rightarrow \infty$ . Then for  $1 \leq k \leq d-1$*

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

The first point of similarity to note between Theorems 8.1 and 9.1 is that while neither theorem identifies a sharp threshold, both identify the first-order term for the transition to homological connectivity:  $(2 - \frac{2}{d}) \log n$  and  $\log n$  respectively. It is worth noting how the geometric differences between building a Čech complex on a manifold with boundary, rather than a closed manifold, inform the differences in the coefficients of the terms in the homological connectivity thresholds.

Our analysis shows that the distance function of a Poisson point process yields a large number of critical points near to the boundary, and this results in the disparity between the constant factor of the first order terms. For a collection of points  $\mathcal{Y}$  of the Poisson process to induce a critical point at the center of these points  $c(\mathcal{Y})$ , one requires that no other point of the Poisson point process lies in the ball of radius  $\rho(\mathcal{Y})$  centered at  $c(\mathcal{Y})$ . The existence of critical points near to the boundary is made more likely by the fact that the ball of radius  $\rho(\mathcal{Y})$  centered at  $c(\mathcal{Y})$  is cut by the boundary, and so it is more likely that no other point of the Poisson point process lies in this cut ball. In the most extreme case the ball's volume is cut in half by the boundary and this introduces the factor of 2 in the leading term of the threshold. Since this phenomenon occurs only for critical points near to the boundary we have to scale our count by the volume of an  $r$ -collar neighborhood of the boundary. This volume behaves like  $\text{Vol}(\partial M_r) \sim r \sim (\frac{\log n}{n})^{\frac{1}{d}}$ , and in particular the factor  $n^{-\frac{1}{d}}$  introduces the term  $-\frac{1}{d}$  to the leading coefficient in Theorem 8.1.

The second-order disparity is identified to lie in the range  $[(k-2) \log \log n, k \log \log n]$  by Theorem 9.1 and in the range  $[2(\frac{1}{d}-3) \log \log n, 2k \log \log n]$  by Theorem 8.1. The factor of 2 in the second-order terms is again induced by the effect of the boundary cutting volumes in half. The second

order term of the lower threshold is affected by counting special erroneous cycles which occur near the boundary which we call  $\Theta$ -like-cycles. Since we only count points near the boundary, again a  $\frac{1}{d}$  term is introduced to the coefficient of the second order term. The coefficient of the second order term is further impacted by a lower bound for the volume of a subset of a Grassmannian, and may be able to be improved if this bound is sharpened.

*Relative homology.* Here we studied the homological connectivity of a compact Riemannian manifold with boundary in terms of the absolute homology. Equally we could have chosen to study the homology of the manifold relative to its boundary. We could perform the same analysis to attain connectivity thresholds for the relative homology by counting critical points of a Morse function induced by the Point process. Recall that the Morse complex of the Morse function recovers absolute homology if our Morse function attains a maximum on the boundary, and relative homology if our Morse function attains a minimum on the boundary (Theorem 3.4). Thus the negative of the distance function from the point process would induce a Morse complex which calculates the relative homology.

Taking the negative of the distance function converts index  $k$  critical points into index  $d - k$  critical points. With this setup we lose the geometric interpretation that the Morse complex at scale  $r$  computes the homology of the union of radius  $r$  balls built around the point process. This correspondence is a result analogous to Lefschetz duality combined with the Universal Coefficient Theorem for Cohomology for calculating the relative homology, which makes it clear that given  $H_k(M)$  and  $H_{k-1}(M)$  we can compute  $H_{d-k}(M, \partial M)$ .

There is a geometric interpretation for the dual to our  $\Theta$ -like-cycles when we consider the Morse complex of the negative of the distance function. For a  $\Theta$ -like-cycle of index  $k$  we require a  $k$ -simplex to be introduced approximately tangential to the boundary. The corresponding orthogonal dual  $(d - k)$ -simplex crosses the part of the partial annulus cut by the boundary and introduces a new homological cycle when we take homology relative to the boundary. See Figure 4, for an illustration of such a dual simplex which introduces a relative cycle.

## ACKNOWLEDGMENTS

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

## LIST OF SYMBOLS

$M$	compact Riemannian manifold with boundary
$g$	smooth Riemannian metric
$\rho(\cdot, \cdot)$	distance induced by Riemannian metric
$\tau_M$	the reach of the manifold $M$
$\partial M_r$	an $r$ neighborhood of the boundary
$P$	a finite sample of points of a manifold
$\mathcal{P}, \mathcal{Q}$	a Poisson point process
$B_r(P)$	the union of radius $r$ balls centered at each $p \in P$
$C(n, r)$	Čech complex at radius $r$ on a point process of intensity $n$

$c(\mathcal{Y}), c(\mathbf{y})$	the center of a finite collection of points
$\rho(\mathcal{Y}), \rho(\mathbf{y})$	the critical value of a finite collection of points
$B(\mathcal{Y}), B(\mathbf{y})$	the ball centered at $c(\mathcal{Y})$ with radius $\rho(\mathcal{Y})$
$\beta_k(r)$	the $k$ th Betti number associated to a Čech complex at radius $r$
$\omega_d$	volume of a unit radius $d$ -dimensional ball
$\Lambda$	expected number of points lying in an $r$ -ball $\Lambda = \omega_d n r^d$
$C$	constant factor, a product of constant terms used to simplify expressions in inequalities
$C_k^{\rho_M}(r, r_0)$	set of index $k$ critical points with critical values in the range $[r, r_0)$
$\Theta_k^\varepsilon(r_1, r)$	set of index $k$ critical points with critical values in the range $[r_1, r)$ inducing $\Theta$ -cycles
$\Theta_k^{\varepsilon, \partial M}(r_1, r)$	set of index $k$ critical points with critical values in the range $[r_1, r)$ inducing $\Theta$ -like-cycles
$A_\varepsilon(c)$	$\varepsilon$ -annulus about center $c(\mathbf{y})$ of radius $\rho(\mathbf{y})$
$A_\varepsilon^{(\varphi)}(c)$	partial $\varepsilon$ -annulus about center $c(\mathbf{y})$ of radius $\rho(\mathbf{y})$
$\mathbb{S}_u(V)$	sphere of radius $u$ in the plane $V$

## ORCID

Henry-Louis de Kergorlay  <https://orcid.org/0000-0003-0520-3803>

Ulrike Tillmann  <https://orcid.org/0000-0002-8076-7660>

Oliver Vipond  <https://orcid.org/0000-0003-2355-5570>

## REFERENCES

1. A. Baddeley, "Spatial point processes and their applications," in W. Weil, Ed., *Stochastic geometry*, Lecture Notes in Mathematics, Vol **1892**, Springer, Berlin, Germany, 2007, pp. 1–75.
2. O. Bobrowski and M. Kahle, *Topology of random geometric complexes: A survey*, J. Appl. Comput. Topol. **1** (2018), no. 3, 331–364.
3. O. Bobrowski, M. Kahle, and P. Skraba, *Maximally persistent cycles in random geometric complexes*, Ann. Appl. Probab. **27** (2017), no. 4, 2032–2060.
4. O. Bobrowski and G. Oliveira, *Random Čech complexes on Riemannian manifolds*, Random Struct. Algorithms **54** (2019), no. 3, 373–412.
5. O. Bobrowski and S. Weinberger, *On the vanishing of homology in random Čech complexes*, Random Struct. Algorithms **51** (2017), no. 1, 14–51.
6. W. Chai, *Random topological structures*, PhD thesis, University of Chicago, 2018.
7. F. Chazal, D. Cohen-Steiner, and A. Lieutier, *A sampling theory for compact sets in Euclidean space*, Discrete Comput. Geom. **41** (2009), no. 3, 461–479.
8. P. Diaconis and P. J. Forrester, *Hurwitz and the origins of random matrix theory in mathematics*, Random Matrices Theory Appl. **6** (2017), no. 1, 1730001, 26.
9. H. Edelsbrunner and J. L. Harer, *Computational topology - An introduction*, American Mathematical Society, Providence, RI, 2010.
10. H. Federer, *Curvature measures*, Trans. Am. Math. Soc. **93** (1959), 418–491.
11. L. Flatto and D. J. Newman, *Random coverings*, Acta Math. **138** (1977), no. 3–4, 241–264.
12. V. Gershkovich and H. Rubinstein, *Morse theory for min-type functions*, Asian J. Math. **1** (1997), no. 4, 696–715.
13. A. Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, UK, 2002.
14. A. Hurwitz, *Über die Erzeugung der Invarianten durch Integration*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **71–2** (1897), 1897.
15. M. Kahle, *Sharp vanishing thresholds for cohomology of random flag complexes*, Ann. Math. **179** (2014), no. 3, 1085–1107.
16. J. M. Lee, *Introduction to smooth manifolds*, Graduate Texts in Mathematics, 2nd ed., Vol **218**, Springer, New York, NY, 2013.
17. N. Linial and R. Meshulam, *Homological connectivity of random 2-complexes*, Combinatorica **26** (2006), no. 4, 475–487.
18. R. E. Miles, *Isotropic random simplices*, Adv. Appl. Probab. **3** (1971), 353–382.
19. J. Milnor, *Lectures on the h-cobordism theorem*, Princeton University Press, Princeton, NJ, 1965.



20. J. Nash, *The imbedding problem for Riemannian manifolds*, Ann. Math. **63** (1956), 20–63.
21. P. Niyogi, S. Smale, and S. Weinberger, *Finding the homology of submanifolds with high confidence from random samples*, Discrete Comput. Geom. **39** (2008), no. 1-3, 419–441.
22. N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, *A roadmap for the computation of persistent homology*, EPJ Data Sci. **6** (2017), no. 1, 17.
23. R. Paul and S. K. Chalup, *A study on validating non-linear dimensionality reduction using persistent homology*, Pattern Recognit. Lett. **100** (2017), 160–166.
24. M. Penrose, *Random geometric graphs*, Oxford Studies in Probability, Oxford University Press, Oxford, UK, 2003.
25. V. Robins, *Betti number signatures of homogeneous Poisson point processes*, Phys. Rev. E **74** (2006), 061107.
26. R. Schneider and W. Weil, *Stochastic and integral geometry*, Probability and Its Applications, Springer-Verlag, Berlin/Heidelberg, Germany, 2008.
27. M. F. Smiley and J. Milnor, *Morse theory*, Am Math Monthly **71** (1964), no. 8, 936.
28. J. B. Tenenbaum, V. De Silva, and J. C. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science **290** (2000), no. 5500, 2319–2323.
29. Y. Wang and B. Wang, *Topological inference of manifolds with boundary*, Comput. Geom. **88** (2020), 101606.
30. A. Zomorodian and G. Carlsson, *Computing persistent homology*, Proc. 20th Annu. Sympos. Comput. Geom. SCG '04, ACM, New York, NY, 2004. pp. 347–356.

**How to cite this article:** H. -L. Kergorlay, U. Tillmann, and O. Vipond, *Random Čech complexes on manifolds with boundary*, Random Struct. Algorithms. **61** (2022), 309–352.  
<https://doi.org/10.1002/rsa.21062>

## APPENDIX A: UPPER THRESHOLD PROOFS

*Proof of Proposition 6.1.* Let us define  $\hat{\beta}_k(r) = \beta_k(r) - \beta_k(M)$  then use Lemma 6.2 to bound  $\mathbb{E}[\hat{\beta}_k(r)]$ . Let us condition on the event  $E = \{M \subset B_{r_0}(\mathcal{P}_n)\}$  that we cover our manifold with the radius  $r_0$  neighborhood of our Poisson Process. Conditioning on  $E$  we get:

$$\mathbb{E}[\hat{\beta}_k(r)] = \mathbb{E}[\hat{\beta}_k(r)|E]\mathbb{P}(E) + \mathbb{E}[\hat{\beta}_k(r)|E^c]\mathbb{P}(E^c).$$

We can bound the term  $\mathbb{E}[\hat{\beta}_k(r)|E]\mathbb{P}(E)$  by observing that any nontrivial  $k$ -cycle in  $C_r$  that is trivial in  $M$  must be annihilated by a  $(k+1)$ -critical point in  $C_{k+1}^{\rho_M}(r, r_0)$ . Thus assuming  $\Lambda \rightarrow \infty$ ,  $r = o(r_0)$  and  $\Lambda_{r_0} r_0^2 \rightarrow 0$ , using Lemma 6.2 we may bound this term by  $\mathbb{E}[|C_{k+1}^{\rho_M}(r, r_0)|] = O(n^{1-\frac{1}{d}} \Lambda^k e^{-\frac{1}{2}\Lambda} + n \Lambda^k e^{-\Lambda})$

The second term,  $\mathbb{E}[\hat{\beta}_k(r)|E^c]\mathbb{P}(E^c)$  can be bounded using an  $\frac{r_0}{2}$ -net to bound the non-coverage probability. Let us begin by bounding  $\mathbb{E}[\beta_k(r)|E^c]\mathbb{P}(E^c)$ . A coarse upper bound for  $\beta_k(r)$  is the number of  $k$ -dimensional faces of  $C_r$ :

$$\begin{aligned} \mathbb{E}[\beta_k(r)|E^c]\mathbb{P}(E^c) &\leq \mathbb{E}\left[\binom{|\mathcal{P}_n|}{k+1} | E^c\right] \mathbb{P}(E^c) \\ &= \sum_{k+1}^{\infty} \binom{m}{k+1} \mathbb{P}(|\mathcal{P}_n| = m | E^c) \mathbb{P}(E^c) \\ &= \sum_{k+1}^{\infty} \binom{m}{k+1} \mathbb{P}(E^c | |\mathcal{P}_n| = m) \mathbb{P}(|\mathcal{P}_n| = m). \end{aligned}$$

Since  $\mathcal{P}_n$  is a Poisson process of intensity  $n$  then  $\mathbb{P}(|\mathcal{P}_n| = m) = \frac{e^{-n} n^m}{m!}$  and conditioned on  $\{|\mathcal{P}_n| = m\}$  we may write  $\mathcal{P}_n$  as a set of  $m$  independent uniformly distributed random variables  $\chi_m = \{X_1, \dots, X_m\}$

$$\mathbb{P}(E^c \mid |\mathcal{P}_n| = m) = \mathbb{P}(B_{r_0}(\chi_m) \neq M).$$

Let  $\mathcal{N}$  be a  $\frac{r_0}{2}$ -net for  $M$  chosen such that  $|\mathcal{N}| \leq c_d r_0^{-d}$  for some constant  $c_d$  dependent only on the dimension of our manifold and the metric, and such that at least  $\frac{1}{2}$  of the volume of radius  $\frac{r_0}{2}$  ball lies in  $M$ . We then use this net to bound the non-coverage probability:

$$\begin{aligned} \mathbb{P}(B_{r_0}(\chi_m) \neq M) &\leq \sum_{x \in \mathcal{N}} \mathbb{P}(\rho_{\chi_m}(x) > \frac{r_0}{2}) \\ &\leq c_d r_0^{-d} \max_{x \in \mathcal{N}} \left\{ \left( 1 - \text{Vol} \left( B_{\frac{r_0}{2}}(x) \right) \right)^m \right\} \\ &\leq c_d r_0^{-d} (1 - \omega_d 2^{-(d+1)} (1 - s_{\max} r_0^2) r_0^d)^m \\ &\leq c_d r_0^{-d} (1 - \kappa r_0^d)^m, \end{aligned}$$

where  $s_{\max}$  is a constant intrinsic to the manifold as defined in Lemma 2.5. Thus we yield:

$$\begin{aligned} \mathbb{E}[\beta_k(r)|E^c] \mathbb{P}(E^c) &\leq \sum_{m=k+1}^{\infty} \binom{m}{k+1} \mathbb{P}(E^c \mid |\mathcal{P}_n| = m) \mathbb{P}(|\mathcal{P}_n| = m) \\ &\leq \sum_{m=k+1}^{\infty} \binom{m}{k+1} c_d r_0^{-d} (1 - \kappa r_0^d)^m \frac{e^{-n} n^m}{m!} \\ &\leq r_0^{-d} n^{k+1} (1 - \kappa r_0^d)^{k+1} e^{-\kappa n r_0^d} \sum_{j=0}^{\infty} \frac{c_d}{(k+1)!} \frac{e^{-n(1-\kappa r_0^d)} n^j (1 - \kappa r_0^d)^j}{j!} \\ &\leq C r_0^{-d} n^{k+1} (1 - \kappa r_0^d)^{k+1} e^{-\kappa n r_0^d} \leq C r_0^{-d} n^{k+1} e^{-\kappa n r_0^d}. \end{aligned}$$

Calculating the asymptotic behavior of  $r_0^{-d} n^{k+1} e^{-\kappa n r_0^d} / n \Lambda^k e^{-\Lambda}$  given the condition that  $r_0 \geq r(\frac{\omega_d}{\kappa}(1 + |\log r|))^{1/d}$  we see:

$$\begin{aligned} \frac{r_0^{-d} n^{k+1} e^{-\kappa n r_0^d}}{n \Lambda^k e^{-\Lambda}} &= \frac{n^k e^{\Lambda - \kappa n r_0^d}}{r_0^d \Lambda^k} \leq \frac{n^k e^{-|\log r| \Lambda}}{r^d \Lambda^k (\frac{\omega_d}{\kappa}(1 + |\log r|))} \\ &= \frac{n^k r^{\Lambda-d}}{\Lambda^k (\frac{\omega_d}{\kappa}(1 + |\log r|))} = \frac{n^k r^{\Lambda-d}}{n^k r^{dk} \omega_d^k (\frac{\omega_d}{\kappa}(1 + |\log r|))} = \frac{r^{\Lambda-d-dk}}{\omega_d^k (\frac{\omega_d}{\kappa}(1 + |\log r|))} \rightarrow 0. \end{aligned}$$

Hence  $\mathbb{E}[\beta_k(r)|E^c] \mathbb{P}(E^c) = o(n \Lambda^k e^{-\Lambda})$ . Then we note that  $\mathbb{P}(E^c) = o(n \Lambda^k e^{-\Lambda})$  by an almost identical argument and so  $\mathbb{E}[\hat{\beta}_k(r)|E^c] \mathbb{P}(E^c) = \mathbb{E}[\beta_k(r)|E^c] \mathbb{P}(E^c) - \beta_k(M) \mathbb{P}(E^c) = o(n \Lambda^k e^{-\Lambda})$ . ■

## APPENDIX B: LOWER THRESHOLD PROOFS

*Proof of Lemma 7.8.* We may assume  $\mathcal{P}$  to be generic in the sense that the critical values of each simplex are distinct, so there is some  $\rho_- < \rho$  such that  $C(\mathcal{P}, \rho) = C(\mathcal{P}, \rho_-) \cup \Delta$  where  $\Delta$  is the  $k$ -simplex  $\mathcal{V}$ . Moreover, the boundary  $\partial \Delta \in C(\mathcal{P}, \rho_-)$ , and  $\Delta$  is not the face of any higher simplex in  $C(\mathcal{P}, \rho)$  by the construction of the Čech complex.

Suppose that  $\partial\Delta$  is a boundary in  $C(\mathcal{P}, \rho_-)$  so that  $\partial\Delta = \partial\gamma$  for some  $k$  chain  $\gamma \in C(\mathcal{P}, \rho_-)$ . Then clearly  $\Delta - \gamma$  is a  $k$ -cycle in  $C(\mathcal{P}, \rho)$ . However  $\Delta - \gamma$  is not a boundary or homologous to another cycle since  $\Delta$  is not the face of any higher simplex in  $C(\mathcal{P}, \rho)$ , and so we introduce a new nontrivial  $k$ -cycle.

Thus it suffices to show that  $\partial\Delta$  is indeed a boundary. Consider the natural map from the simplicial chains to the singular chains  $\iota_k : C_k^\Delta(C(\mathcal{P}, \rho_-)) \rightarrow C_k^{\text{sing}}(B_{\rho_-}(\mathcal{P}))$ . For sufficiently small  $\rho_-$ , by the Nerve Lemma the induced map on homology is an isomorphism  $h_k : H_k(C(\mathcal{P}, \rho_-)) \cong H_k(B_{\rho_-}(\mathcal{P}))$ .

If  $\rho_-$  is sufficiently close to  $\rho$  then given that  $A_{2\psi}^{(\varphi)}(\mathcal{Y}) \subset B_\rho(\mathcal{P})$  we also have  $A_{2\psi}^{(\varphi)}(\mathcal{Y}) \subset B_{\rho_-}(\mathcal{P})$ . Our  $\psi$  is constructed in order that  $\iota_{k-1}(\partial\Delta) \in C_{k-1}^{\text{sing}}(A_{2\psi}^{(\varphi)}) \subset C_{k-1}^{\text{sing}}(B_{\rho_-}(\mathcal{P}))$ .  $A_{2\psi}^{(\varphi)}$  is homotopic to a  $d$ -dimensional annulus sliced by a  $d - 1$ -hyperplane and so  $H_{k-1}(A_{2\psi}^{(\varphi)}) = 0$ , and thus  $h_{k-1}(\partial\Delta) = 0$ . ■

## APPENDIX C: SECOND MOMENT CALCULATIONS

We have attained a lower bound for the expected number of  $\Theta$ -like-cycles near the boundary of the manifold. Let us now seek to show that w.h.p. the number of  $\Theta$ -like-cycles is bounded below by this regime. Some of the notation used in this section is defined in Section 7. Let us make the abbreviation  $\Theta_k^{\varepsilon, \partial M} = \Theta_k^{\varepsilon, \partial M}(r_1, r)$ . Using Chebyshev's inequality we observe that:

$$\mathbb{P}(\Theta_k^{\varepsilon, \partial M} \leq \gamma \mathbb{E}[\Theta_k^{\varepsilon, \partial M}]) \leq \frac{\text{Var}(\Theta_k^{\varepsilon, \partial M})}{(1 - \gamma)^2 \mathbb{E}[\Theta_k^{\varepsilon, \partial M}]^2}.$$

It suffices to show that the right-hand side goes to zero. In bounding the variance we will be interested in the following term:

$$(\Theta_k^{\varepsilon, \partial M}(r_1, r))^2 = \sum_{\substack{\mathcal{Y}_1, \mathcal{Y}_2 \subset \mathcal{P}_n \\ |\mathcal{Y}_i| = k+1}} g_r^{\varepsilon, \delta}(\mathcal{Y}_1, \mathcal{P}_n) g_r^{\varepsilon, \delta}(\mathcal{Y}_2, \mathcal{P}_n).$$

Let us separate the distinct situations in which the centers induced by  $\mathcal{Y}_i$  are either close together  $T_2$ , or far apart  $T_1$ . Let us define:

$$\begin{aligned} \Phi_r(\mathcal{Y}_1, \mathcal{Y}_2) &= \mathbf{1}\{B_r(c(\mathcal{Y}_1)) \cap B_r(c(\mathcal{Y}_2)) = \emptyset\}. \\ T_1 &= \Theta_k^{\varepsilon, \partial M}(r_1, r)^2 \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2) = \sum_{\substack{\mathcal{Y}_1, \mathcal{Y}_2 \subset \mathcal{P}_n \\ |\mathcal{Y}_i| = k+1}} g_r^{\varepsilon, \delta}(\mathcal{Y}_1, \mathcal{P}_n) g_r^{\varepsilon, \delta}(\mathcal{Y}_2, \mathcal{P}_n) \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2). \\ T_2 &= \Theta_k^{\varepsilon, \partial M}(r_1, r)^2 (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) = \sum_{\substack{\mathcal{Y}_1, \mathcal{Y}_2 \subset \mathcal{P}_n \\ |\mathcal{Y}_i| = k+1}} g_r^{\varepsilon, \delta}(\mathcal{Y}_1, \mathcal{P}_n) g_r^{\varepsilon, \delta}(\mathcal{Y}_2, \mathcal{P}_n) (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)). \end{aligned}$$

Thus we may express the variance as:

$$\text{Var}(\Theta_k^{\varepsilon, \partial M}) = \mathbb{E}[(\Theta_k^{\varepsilon, \partial M})^2] - \mathbb{E}[\Theta_k^{\varepsilon, \partial M}]^2 = (\mathbb{E}[T_1] - \mathbb{E}[\Theta_k^{\varepsilon, \partial M}]^2) + \mathbb{E}[T_2].$$

Using Palm Theory 2.4 to simplify the sums taken over  $(k + 1)$ -sized subsets we obtain:

$$\begin{aligned} \mathbb{E}[\Theta_k^{\varepsilon, \partial M}]^2 &= \frac{n^{2k+2}}{((k+1)!)^2} \mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}'_1 \cup \mathcal{P}_n) g_r^{\varepsilon, \delta}(\mathcal{Y}'_2, \mathcal{Y}'_2 \cup \mathcal{P}'_n)], \\ \mathbb{E}[T_1] &= \frac{n^{2k+2}}{((k+1)!)^2} \mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}' \cup \mathcal{P}_n) g_r^{\varepsilon, \delta}(\mathcal{Y}'_2, \mathcal{Y}' \cup \mathcal{P}_n)], \end{aligned}$$

where  $\mathcal{P}_n, \mathcal{P}'_n$  are i.i.d. Poisson processes,  $\mathcal{Y}'_1, \mathcal{Y}'_2 \subset M$  are i.i.d., and  $\mathcal{Y}' = \mathcal{Y}'_1 \cup \mathcal{Y}'_2$ . Notice that when  $\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2) \neq 0$  then  $g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}' \cup \mathcal{P}_n) = g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}'_i \cup \mathcal{P}_n)$  since  $g_r^{\varepsilon, \delta}$  is only dependent on the points in the second argument which are sufficiently close to the points in the first argument. Thus we can make the following simple bound by omitting the negative contribution of the third expectation term in the first expression:

$$\begin{aligned} \mathbb{E}[T_1] - \mathbb{E}[\Theta_k^{\varepsilon, \delta M}]^2 &= \frac{n^{2k+2}}{((k+1)!)^2} (\mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}' \cup \mathcal{P}_n)g_r^{\varepsilon, \delta}(\mathcal{Y}'_2, \mathcal{Y}' \cup \mathcal{P}_n)\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)] \\ &\quad - \mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}'_1 \cup \mathcal{P}_n)g_r^{\varepsilon, \delta}(\mathcal{Y}'_2, \mathcal{Y}'_2 \cup \mathcal{P}_n)\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)] \\ &\quad - \mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}'_1 \cup \mathcal{P}_n)g_r^{\varepsilon, \delta}(\mathcal{Y}'_2, \mathcal{Y}'_2 \cup \mathcal{P}'_n)(1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2))]) \\ &\leq \frac{n^{2k+2}}{((k+1)!)^2} (\mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}'_1 \cup \mathcal{P}_n)g_r^{\varepsilon, \delta}(\mathcal{Y}'_2, \mathcal{Y}'_2 \cup \mathcal{P}_n)\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)] \\ &\quad - \mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}'_1 \cup \mathcal{P}_n)g_r^{\varepsilon, \delta}(\mathcal{Y}'_2, \mathcal{Y}'_2 \cup \mathcal{P}'_n)\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)]) \\ &=: \frac{n^{2k+2}}{((k+1)!)^2} \mathbb{E}[\Delta g_r^{\varepsilon, \delta}]. \end{aligned}$$

Let us condition on the subsets  $\mathcal{Y}'_1, \mathcal{Y}'_2$  and note that whenever  $\Delta g_r^{\varepsilon, \delta} \neq 0$  these two subsets are sufficiently separated so as not to interact. Thus by the independence property of Poisson processes:

$$\begin{aligned} &\mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}'_1 \cup \mathcal{P}_n)g_r^{\varepsilon, \delta}(\mathcal{Y}'_2, \mathcal{Y}'_2 \cup \mathcal{P}_n)|\mathcal{Y}'_1, \mathcal{Y}'_2] \\ &= \mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}'_1 \cup \mathcal{P}_n)|\mathcal{Y}'_1, \mathcal{Y}'_2]\mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_2, \mathcal{Y}'_2 \cup \mathcal{P}_n)|\mathcal{Y}'_1, \mathcal{Y}'_2] \\ &= \mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_1, \mathcal{Y}'_1 \cup \mathcal{P}_n)|\mathcal{Y}'_1, \mathcal{Y}'_2]\mathbb{E}[g_r^{\varepsilon, \delta}(\mathcal{Y}'_2, \mathcal{Y}'_2 \cup \mathcal{P}'_n)|\mathcal{Y}'_1, \mathcal{Y}'_2]. \end{aligned}$$

Thus  $\mathbb{E}[\Delta g_r^{\varepsilon, \delta}] = \mathbb{E}[\mathbb{E}[\Delta g_r^{\varepsilon, \delta}|\mathcal{Y}'_1, \mathcal{Y}'_2]] = 0$ . It remains to bound the second term of the variance  $\mathbb{E}[T_2]$ . Let us split the term  $T_2$  into the separate cases in which  $\mathcal{Y}_1, \mathcal{Y}_2$  share  $j$  points:

$$T_2 = \sum_{j=0}^{k+1} \sum_{|\mathcal{Y}_1 \cap \mathcal{Y}_2|=j} g_r^{\varepsilon, \delta}(\mathcal{Y}_1, \mathcal{P}_n)g_r^{\varepsilon, \delta}(\mathcal{Y}_2, \mathcal{P}_n)(1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) = \sum_{j=0}^{k+1} I_j.$$

We next require a corollary of Theorem 2.4.

**Corollary C.1.** *Let  $(X, \rho)$  be a metric space,  $f : X \rightarrow \mathbb{R}$  a probability density and  $\mathcal{P}_n$  a Poisson process on  $X$  with intensity  $\lambda_n = nf$ . If  $h(\mathcal{Y}, \mathcal{X})$  is a measurable function for all finite subsets  $\mathcal{Y}, \mathcal{X} \subset X$ , with  $\mathcal{Y} \subset \mathcal{X}$  and  $|\mathcal{Y}| = k+1$  then:*

$$\mathbb{E} \left[ \sum_{\substack{|\mathcal{Y}_1|=|\mathcal{Y}_2|=k+1, \\ |\mathcal{Y}_1 \cap \mathcal{Y}_2|=j}} h(\mathcal{Y}_1, \mathcal{P}_n)h(\mathcal{Y}_2, \mathcal{P}_n) \right] = \frac{n^{2k-j}}{j!(k-j)!} \mathbb{E}[h(\mathcal{Y}'_1, \mathcal{Y}' \cup \mathcal{P}_n)h(\mathcal{Y}'_2, \mathcal{Y}' \cup \mathcal{P}_n)],$$

where  $\mathcal{Y}' = \mathcal{Y}'_1 \cup \mathcal{Y}'_2$  is a set of  $2k-j$  i.i.d points in  $X$  with density  $f$ , with  $|\mathcal{Y}'_1 \cap \mathcal{Y}'_2| = j$  and  $\mathcal{Y}'$  independent of  $\mathcal{P}_n$ .

Using the result given in Corollary C.1 yields:

$$\begin{aligned}\mathbb{E}[I_j] &= \sum_{|\mathcal{Y}_1 \cap \mathcal{Y}_2|=j} g_r^{\varepsilon, \delta}(\mathcal{Y}_1, \mathcal{P}_n) g_r^{\varepsilon, \delta}(\mathcal{Y}_2, \mathcal{P}_n) (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) \\ &\leq \frac{n^{2k+2-j}}{j!((k+1-j)!)^2} \int_{DM^{2k+2-j}} h_r^{\varepsilon, \delta}(\mathbf{y}_1) h_r^{\varepsilon, \delta}(\mathbf{y}_2) e^{-n \text{Vol}(\mathbf{y}_1, \mathbf{y}_2)} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) |\text{dvol}_g(\mathbf{y})|.\end{aligned}$$

We bound the volume term using the results of Section 2.4:

$$\begin{aligned}\text{Vol}(\mathbf{y}_1, \mathbf{y}_2) &= \text{Vol}(B_r(c(\mathbf{y}_1)) \cup B_r(c(\mathbf{y}_2))) \\ &\geq \left(\frac{1}{2} + \alpha\right) (1 - (d\sqrt{r} + \sqrt{r}^2)) \left(1 + \frac{\omega_{d-1} d(c_1, c_2)}{\omega_d r} + O\left(\frac{d(c_1, c_2)^2}{r}\right)\right) \omega_d r^d,\end{aligned}$$

where  $\alpha = O(\frac{\delta}{r}) + O(r)$ .

Next we require a multivariable Blaschke-Petkantschin change of variable formula in order to bound the expectations  $\mathbb{E}[I_j]$ . We show how to bound a change of variable formula when integrating over two variables in  $M^{k+1}$ , where  $M$  has nonempty boundary. This formula is already used (without proof) in [4] in the case where  $M$  is closed.

**Lemma C.2.** *Let  $M$  be a compact Riemannian manifold with nonempty boundary and let  $\mathbf{y}_1, \mathbf{y}_2 \in M^{k+1}$ . Denote the respective centers by  $c_1, c_2$ . Define*

$$\Omega := \left\{ (\mathbf{y}_1, \mathbf{y}_2) \in (M^{k+1})^2 \mid a \leq \rho_M(c_1, c_2) \leq b \right\}.$$

*Then there is some constant  $C_M$  dependent solely on  $M$  for which the following bound holds:*

$$\begin{aligned}& \int_{\Omega} f_1(\mathbf{y}_1) f_2(\mathbf{y}_2) \mathbf{1}_r(\mathbf{y}_1, \mathbf{y}_2) |\text{dvol}_g(\mathbf{y}_1, \mathbf{y}_2)| \\ & \leq C_M \int_M |\text{dvol}_g(c_1)| \int_a^b ds \int_{\mathbb{S}_1(T_{c_1}M)} s^{d-1} |\text{dvol}_{\mathbb{S}_1(T_{c_1}M)}(w)| \\ & \quad \times \prod_{i=1}^2 \int_0^r du_i u_i^{kd-1} \int_{\text{Gr}(k,d)} d\mu_{k,d}(V) \int_{\mathbb{S}_1(V)^{k+1}} |\text{dvol}_{\mathbb{S}_1(V)^{k+1}}(\mathbf{w}_i)| f_i(\exp_{c_i}(u_i \mathbf{w}_i)).\end{aligned}$$

*Proof.* We use the Blaschke-Petkantschin formula for integrals over one variable in  $M^{k+1}$  to attain the result. Given  $\mathbf{y}_1 \in M^{k+1}$  and  $c_1$  the induced center, let

$$\Omega(\mathbf{y}_1) := \left\{ \mathbf{y}_2 \in M^{k+1} \mid a \leq \rho_M(c_1, c_2) \leq b \right\}.$$

Denote the above integral on the LHS by  $I$ , we have

$$I = \int_{M^{k+1}} f_1(\mathbf{y}_1) \mathbf{1}_r(\mathbf{y}_1) \left( \int_{\Omega(\mathbf{y}_1)} f_2(\mathbf{y}_2) \mathbf{1}_r(\mathbf{y}_2) |\text{dvol}_g(\mathbf{y}_2)| \right) |\text{dvol}_g(\mathbf{y}_1)|.$$

We first compute the inner integral, for fixed  $c_1 \in M$ .

Note that

$$\int_{\Omega(\mathbf{y}_1)} f_2(\mathbf{y}_2) \mathbf{1}_r(\mathbf{y}_2) |\mathrm{dvol}_g(\mathbf{y}_2)| = \int_M f_2(\mathbf{y}_2) \mathbf{1}_r(\mathbf{y}_2) \mathbf{1}\{c_2 \in A_a^b(c_1)\} |\mathrm{dvol}_g(\mathbf{y}_2)|,$$

where  $A_a^b(c_1) := B_b(c_1) \setminus B_a(c_1)^o$ .

By the Blaschke-Petkantschin formula for manifolds with nonempty boundary, we then find

$$\begin{aligned} & \int_{\Omega(\mathbf{y}_1)} f_2(\mathbf{y}_2) \mathbf{1}_r(\mathbf{y}_2) |\mathrm{dvol}_g(\mathbf{y}_2)| \\ &= \int_{A_a^b(c_1)} |\mathrm{dvol}_g(c_2)| \int_0^r du_2 \int_{\mathrm{Gr}(k, T_{c_2} DM)} d\mu_{k,d}(V) \\ & \quad \times \left( \int_{(\mathbb{S}_u(V) \cap \exp_{c_2}^{-1}(M))^{k+1}} |\mathrm{dvol}_{\mathbb{S}_u(V)^{k+1}}(\mathbf{v}_2)| f_2(\exp_{c_2}(\mathbf{v}_2)) \Upsilon_u^{d-k}(\mathbf{v}_2) \prod_{j=1}^{k+1} \sqrt{\det(g_{\exp_{c_2}(v_j)})} \right) \\ & \leq C \int_{A_a^b(c_1)} |\mathrm{dvol}_g(c_2)| \int_0^r du_2 u_2^{d_{k-1}} \int_{\mathrm{Gr}(k,d)} d\mu_{k,d}(V) \\ & \quad \times \left( \int_{\mathbb{S}_1(V)^{k+1}} |\mathrm{dvol}_{\mathbb{S}_1(V)^{k+1}}(\mathbf{w}_2)| f_2(\exp_{c_2}(u_2 \mathbf{w}_2)) \Upsilon_1^{d-k}(\mathbf{w}_2) \prod_{j=1}^{k+1} \sqrt{\det(g_{\exp_{c_2}(u_2 w_j)})} \right). \end{aligned}$$

Using the compactness of  $M$ , the above is

$$\leq C \int_{A_a^b(c_1)} |\mathrm{dvol}_g(c_2)| \int_0^r du_2 u_2^{d_{k-1}} \int_{\mathrm{Gr}(k,d)} d\mu_{k,d}(V) \int_{\mathbb{S}_1(V)^{k+1}} |\mathrm{dvol}_{\mathbb{S}_1(V)^{k+1}}(\mathbf{w}_2)| f_2(\exp_{c_2}(u_2 \mathbf{w}_2)).$$

Furthermore, using the Riemannian approximation results and polar decomposition, we have

$$\begin{aligned} \int_{A_a^b(c_1)} |\mathrm{dvol}_g(c_2)| & \leq C \int_a^b ds \int_{\mathbb{S}_s(T_{c_1} DM)} |\mathrm{dvol}_{\mathbb{S}_s(T_{c_1} DM)}(w)| \\ & = \int_a^b ds \int_{\mathbb{S}_1(T_{c_1} DM)} s^{d-1} |\mathrm{dvol}_{\mathbb{S}_1(T_{c_1} DM)}(w)|. \end{aligned}$$

The outer integral in the expression of  $I$  is estimated again with the Blaschke-Petkantschin formula for manifolds with nonempty boundary. Doing so and combining the result with the above expression for the inner integral of  $I$  yields the claimed formula.  $\blacksquare$

We now make two separate change of variables separating the cases where  $j = 0$  and  $j \neq 0$ . We take the center  $c(\mathbf{y}_2)$  to be in polar coordinates around  $c(\mathbf{y}_1)$ . For details see Lemma C.2 and Section 5 on Blaschke-Petkantschin formulae. Let us further partition our integral into two regions  $\Omega_a, \Omega_b$ :

$$\Omega_a = \left\{ (\mathbf{y}_1, \mathbf{y}_2) : 0 \leq \frac{\rho(c(\mathbf{y}_1), c(\mathbf{y}_2))}{r} \leq \epsilon \right\}.$$

$$\Omega_b = \left\{ (\mathbf{y}_1, \mathbf{y}_2) : \epsilon \leq \frac{\rho(c(\mathbf{y}_1), c(\mathbf{y}_2))}{r} \leq 4 \right\}.$$

Under the assumption that  $\Lambda r \rightarrow 0$  and  $r_1 = (1 - \frac{1}{2c_g^2\Lambda^2})r$  we attain:

$$\begin{aligned}\mathbb{E}[I_0^{(a)}] &\leq \frac{n^{2k+2}}{((k+1)!)^2} \int_{\Omega_a} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-n\text{Vol}(\mathbf{y}_1, \mathbf{y}_2)} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) |\text{dvol}_g(\mathbf{y})| \\ &\leq Cn^{2k+2} \int_{\Omega_a} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-\frac{\Lambda}{2}(1-(d\eta' r + \eta r^2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) |\text{dvol}_g(\mathbf{y})| \\ &\leq Cn^{2k+2} e^{-\frac{\Lambda}{2}} \int_{\partial M_{[\delta, 2\delta]}} |\text{dvol}_g(c_1)| \int_0^{\varepsilon r} ds \int_{\mathbb{S}_1(T_{c_1}M)} s^{d-1} \text{dvol}_{\mathbb{S}_1(T_{c_1}M)}(w) \\ &\quad \times \prod_{i=1}^2 \int_{r_1}^r du_i \int_{\text{Gr}(k,d)} u_i^{k(d-k)} d\mu_{k,d}(V) \int_{(\mathbb{S}_1(V))^{k+1}} u_i^{(k-1)(k+1)} |\text{dvol}_{(\mathbb{S}_1(V))^{k+1}}(\mathbf{w}_i)| h_r^{\varepsilon,\delta}(\exp_{c_i}(u_1 \mathbf{w}_i)) \\ &\leq Cn^{2k+2} e^{-\frac{\Lambda}{2}} \delta \varepsilon^d r^{d(2k+1)} \Lambda^{-4} \leq Cne^{-\frac{\Lambda}{2}} \Lambda^{2k-3} \varepsilon^d r,\end{aligned}$$

$$\begin{aligned}\mathbb{E}[I_0^{(b)}] &\leq \frac{n^{2k+2}}{((k+1)!)^2} \int_{\Omega_b} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-n\text{Vol}(\mathbf{y}_1, \mathbf{y}_2)} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) |\text{dvol}_g(\mathbf{y})| \\ &\leq Cn^{2k+2} \int_{\Omega_b} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-\frac{\Lambda}{2}(1+(\frac{\varepsilon\omega_d}{\omega_{d-1}})(1-(d\eta' r + \eta r^2)))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) |\text{dvol}_g(\mathbf{y})| \\ &\leq Cn^{2k+2} e^{-\frac{\Lambda}{2}(1+(\frac{\varepsilon\omega_d}{\omega_{d-1}}))} \int_{\partial M_{[\delta, 2\delta]}} |\text{dvol}_g(c_1)| \int_{\varepsilon r}^{4r} ds \int_{\mathbb{S}_1(T_{c_1}M)} s^{d-1} \text{dvol}_{\mathbb{S}_1(T_{c_1}M)}(w) \\ &\quad \times \prod_{i=1}^2 \int_{r_1}^r du_i \int_{\text{Gr}(k,d)} u_i^{k(d-k)} d\mu_{k,d}(V) \int_{(\mathbb{S}_1(V))^{k+1}} u_i^{(k-1)(k+1)} |\text{dvol}_{(\mathbb{S}_1(V))^{k+1}}(\mathbf{w}_i)| h_r^{\varepsilon,\delta}(\exp_{c_i}(u_1 \mathbf{w}_i)) \\ &\leq Cn^{2k+2} e^{-\frac{\Lambda}{2}(1+(\frac{\varepsilon\omega_d}{\omega_{d-1}}))} \delta r^{d(2k+1)} \Lambda^{-4} \leq Cne^{-\frac{\Lambda}{2}} e^{-\frac{\Lambda\varepsilon\omega_d}{2\omega_{d-1}}} \Lambda^{2k-3} r,\end{aligned}$$

where  $c_2 = \exp_{c_1}(sw)$ ,  $\mathbf{y}_i = \exp_{c_i}(\mathbf{w}_i)$  and  $\mathbb{S}_1(V)$  is the unit sphere.

Similarly for  $j \neq 0$  we attain bounds:

$$\begin{aligned}\mathbb{E}[I_j^{(a)}] &\leq \frac{n^{2k+2-j}}{j!((k+1-j)!)^2} \int_{\Omega_a} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-n\text{Vol}(\mathbf{y}_1, \mathbf{y}_2)} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) |\text{dvol}_g(\mathbf{y})| \\ &\leq n^{2k+2-j} \int_{\Omega_a} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-\frac{\Lambda}{2}(1-(d\eta' r + \eta r^2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) |\text{dvol}_g(\mathbf{y})| \\ &\leq Cn^{2k+2-j} e^{-\frac{\Lambda}{2}} \int_{\partial M_{[\delta, 2\delta]}} |\text{dvol}_g(c_1)| \\ &\quad \times \int_{r_1}^r du_1 \int_{\text{Gr}(k,d)} u_1^{k(d-k)} d\mu_{k,d}(V_1) \int_{(\mathbb{S}_1(V_1))^{k+1}} u_i^{(k-1)(k+1)} \text{dvol}_{(\mathbb{S}_1(V_1))^{k+1}} \\ &\quad \times \int_0^{\varepsilon r} ds \int_{\mathbb{S}_1(T_{c_1}E)} s^{d-j} \text{dvol}_{\mathbb{S}_1(T_{c_1}E)}(w) \int_{r_1}^r du_2 \int_{\text{Gr}(k-j,d)} u_i^{(k-j)(d-(k-j))} d\mu_{k-j,d}(W) \\ &\quad \times \int_{(\mathbb{S}_1(V_2))^{k+1-j}} u_2^{(k-1)(k+1-j)} \text{dvol}_{(\mathbb{S}_1(V_2))^{k+1-j}} h_r^{\varepsilon,\delta}(\exp_{c_1}(u_1 \mathbf{w}_1)) h_r^{\varepsilon,\delta}(\exp_{c_2}(u_2 \mathbf{w}_2)) \\ &\leq Cn^{2k+2-j} e^{-\frac{\Lambda}{2}} \delta \varepsilon^{d-j+1} r^{d(2k+1-j)+j(k-j)+1} \Lambda^{-4} \leq Cne^{-\frac{\Lambda}{2}} \Lambda^{2k-3} \varepsilon^{d-j+1} r^{j(k-j)+1}.\end{aligned}$$

$$\mathbb{E}[I_j^{(b)}] \leq \frac{n^{2k+2-j}}{j!((k+1-j)!)^2} \int_{\Omega_b} h_r^{\varepsilon,\delta}(\mathbf{y}_1) h_r^{\varepsilon,\delta}(\mathbf{y}_2) e^{-n\text{Vol}(\mathbf{y}_1, \mathbf{y}_2)} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) |\text{dvol}_g(\mathbf{y})|$$



$$\begin{aligned}
&\leq n^{2k+2-j} \int_{\Omega_b} h_r^{\varepsilon, \delta}(\mathbf{y}_1) h_r^{\varepsilon, \delta}(\mathbf{y}_2) e^{-\frac{\Lambda}{2}(1+\frac{\varepsilon\omega_d}{\omega_{d-1}})(1-(d\eta' r + \eta r^2))} (1 - \Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) |\mathrm{dvol}_g(\mathbf{y})| \\
&\leq C n^{2k+2-j} e^{-\frac{\Lambda}{2}(1+\frac{\varepsilon\omega_d}{\omega_{d-1}})} \int_{\partial M_{[\delta, 2\delta]}} |\mathrm{dvol}_g(c_1)| \\
&\quad \times \int_{r_1}^r du_1 \int_{\mathrm{Gr}(k, d)} u_1^{k(d-k)} d\mu_{k, d}(V_1) \int_{(\mathbb{S}_1(V_1))^{k+1}} u_i^{(k-1)(k+1)} \mathrm{dvol}_{(\mathbb{S}_1(V_1))^{k+1}} \\
&\quad \times \int_{\varepsilon r}^{4r} ds \int_{\mathbb{S}_1(T_{c_1} E)} s^{d-j} \mathrm{dvol}_{\mathbb{S}^{d-1}}(w) \int_{r_1}^r du_2 \int_{\mathrm{Gr}(k-j, d)} u_i^{(k-j)(d-(k-j))} d\mu_{k-j, d}(W) \\
&\quad \times \int_{(\mathbb{S}_1(V_2))^{k+1-j}} u_2^{(k-1)(k+1-j)} \mathrm{dvol}_{(\mathbb{S}_1(V_2))^{k+1-j}} h_r^{\varepsilon, \delta}(\exp_{c_1}(u_1 \mathbf{w}_1)) h_r^{\varepsilon, \delta}(\exp_{c_2}(u_2 \mathbf{w}_2)) \\
&\leq C n^{2k+2-j} e^{-\frac{\Lambda}{2}(1+\frac{\varepsilon\omega_d}{\omega_{d-1}})} \delta r^{d(2k+1-j)+j(k-j)+1} \Lambda^{-4} \leq C n e^{-\frac{\Lambda}{2}} e^{-\frac{\Lambda\varepsilon\omega_d}{2\omega_{d-1}}} \Lambda^{2k-3} r^{j(k-j)+1}.
\end{aligned}$$

We note that our largest bounding function is for the expectation  $\mathbb{E}[I_0]$  and so we have the following bound for all  $j$  and for arbitrary  $\varepsilon \in (0, 1)$ :

$$\mathbb{E}[I_j] \leq C n e^{-\frac{\Lambda}{2}} \Lambda^{2k-3} r \left( e^{-\frac{\Lambda\varepsilon\omega_d}{2\omega_{d-1}}} + \varepsilon^d \right).$$

We may therefore use this same regime to bound  $\mathbb{E}[T_2]$ . Let us choose  $\varepsilon = \frac{2(2k-1)\omega_{d-1}}{\omega_d} \frac{\log \log n}{\log n}$ . Then for  $\Lambda = (2 - \frac{2}{d}) \log n + 2(k-2 - (k+1 - \frac{1}{d})) \log \log n - \omega(n)$ , (recalling that  $\mathbb{E}[\Theta_k^{\varepsilon, \partial M}] = \Omega(e^{-\alpha\Lambda} n \Lambda^{k-2} r (\log n)^{-(k+1)})$  and that  $\alpha = \frac{1}{2} + O((\log n)^{-1})$ ), we calculate:

$$\begin{aligned}
\frac{\mathrm{Var}(\Theta_k^{\varepsilon, \partial M})}{\mathbb{E}[\Theta_k^{\varepsilon, \partial M}]^2} &= \frac{\mathbb{E}[T_2]}{\mathbb{E}[\Theta_k^{\varepsilon, \partial M}]^2} \leq C \frac{n e^{-\frac{\Lambda}{2}} \Lambda^{2k-3} r (e^{-\frac{\Lambda\varepsilon\omega_d}{2\omega_{d-1}}} + \varepsilon^d)}{n^2 \Lambda^{2k-4} e^{-2\alpha\Lambda} r^2 (\log n)^{-2(k+1)}} \sim \frac{\Lambda e^{\frac{\Lambda}{2}} (\log n)^{2(k+1)} (e^{-\frac{\Lambda\varepsilon\omega_d}{2\omega_{d-1}}} + \varepsilon^d)}{nr} \\
&\sim e^{-\frac{\omega(n)}{2}} (\log n)^{2k} (e^{-\frac{\Lambda\varepsilon\omega_d}{2\omega_{d-1}}} + \varepsilon^d) \sim e^{-\frac{\omega(n)}{2}} \left( \frac{1}{\log n} + \frac{(\log \log n)^d}{(\log n)^d} \right) \rightarrow 0.
\end{aligned}$$

Thus we conclude that w.h.p.  $\Theta_k^{\varepsilon, \partial M} \geq \frac{1}{2} \mathbb{E}[\Theta_k^{\varepsilon, \partial M}]$ .