

# Intrinsic Validation of Manifold Learning Algorithms

**Brittany Terese Fasy** ✉️ 🏠 

School of Computing & Department of Mathematical Sciences, Montana State University, USA.

**Benjamin Holmgren** ✉️


Department of Computer Science, Duke University, USA.

**Eli Quist** ✉️ 

Department of Mathematical Sciences, Montana State University, USA.

**Bastian Grossenbacher-Rieck**

AIDOS Lab, Helmholtz Munich and Technical University of Munich, Germany.

**Jordan Schupbach** ✉️ 

Department of Mathematical Sciences, Montana State University, USA.

## 1 Abstract

We present an intrinsic method to validate manifold learning techniques, adapting and extending Ripley's K-function. We categorize the extent to which an output of a manifold learning algorithm captures the structure of an ambient manifold, doing so in an unsupervised setting. As a consequence, we generalize the capabilities of the K-function to broad classes of Riemannian manifolds. In particular, we extend the K-function to general two-manifolds using the Gauss-Bonnet theorem, and demonstrate that the K-function for hypersurfaces is well approximated using the first Laplacian eigenvalue. Our approach has desirable convergence properties, is provably stable against noise, and shows experimental promise as a validation tool for manifold learning.

**2012 ACM Subject Classification** Mathematics of computing → Geometric topology

**Keywords and phrases** manifold learning, algorithm validation, Ripley's  $K$ -function, manifolds, hypersurfaces

**Funding** *Brittany Terese Fasy*: National Science Foundation grant numbers 1854336 and 2046730.

*Benjamin Holmgren*: U.S. Department of Energy grant number DE-SC0024386.

*Eli Quist*: National Science Foundation grant number 1854336, Montana State University Undergraduate Scholars Program.

## Acknowledgements

## 1 Introduction

Manifold learning is extremely well-studied in the machine learning, computational geometry, and computational topology literatures [17, 19, 39]. The uses of manifold learning, while generally categorized as a means for nonlinear dimensionality reduction [3, 4, 13, 30, 35], span application areas including shape recognition [20], image recognition [40, 5, Ch. 4], and motion planning [11]. Given the large number of techniques and their practical significance, natural questions concerning their *validation* are raised. That is, when has an algorithm effectively learned manifold structure within data? Moreover, how does one rigorously compare the performance of different manifold learning algorithms? When validating dimensionality reduction algorithms, it is especially pertinent to work in an unsupervised setting. In this paper, we thus assume little or no knowledge of the ground truth in data (such as knowledge of true geodesic distances). Validation in this unsupervised setting is called *intrinsic* validation, as observations are reliant only on core properties of the presented data.



© Brittany Terese Fasy and Benjamin Holmgren and Eli Quist and Jordan Schupbach; licensed under Creative Commons License CC-BY 4.0

42nd Conference on Very Important Topics (CVIT 2016).

Editors: John Q. Open and Joan R. Access; Article No. 23; pp. 23:1–23:25

Leibniz International Proceedings in Informatics



**LIPICs** Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Despite its prevalence and pivotal role in data analysis, validation techniques specific to manifold learning have been largely unexplored. One previous method, [24], is an analog to Precision-Recall, which is not intrinsic and requires knowledge of the underlying manifold. An intrinsic approach is given in [36] which uses persistent local homology to detect singularities in point cloud data. This paper takes a different perspective; giving a global assessment of the manifold properties of data with theoretical guarantees rooted in differential geometry.

## Contributions

Intuitively, discrete data exhibiting “manifold-like” properties could naturally be interpreted to mean that data locally resembles a uniform sample in  $\mathbb{R}^d$ . This manuscript makes such a notion rigorous. We examine local neighborhoods within a point cloud, and score how closely each neighborhood resembles a uniform sample in  $\mathbb{R}^d$  without any knowledge of the ground truth manifold. Our method is based on the natural idea that geodesic balls on a manifold  $\mathbb{X}$  should grow proportionately to balls in  $\mathbb{R}^n$ , up to the curvature of  $\mathbb{X}$ . To measure this behavior, we develop an extension of Ripley’s  $K$ -function [12], a popular method to categorize uniformity in a random sample. In this way, our work can be thought of as a manifold analog to the silhouette coefficient for clustering [29]. Specifically, we adapt Ripley’s  $K$ -function to broad classes of manifolds using principles from differential geometry, most notably the Gauss-Bonnet theorem and hypersurface inequalities relating scalar curvature to first Laplacian eigenvalue. In particular, this paper presents the following:

1. A stable, efficiently computable, intrinsic validation score for Riemannian manifolds (the manifold density function,  $K_{\mathbb{X}}$ ).
2. A stable, efficiently computable approximation for  $K_{\mathbb{X}}$  on two-manifolds, which adjusts for curvature using the Euler characteristic.
3. A stable, efficiently computable, approximation for  $K_{\mathbb{X}}$  on hypersurfaces using the first Laplacian eigenvalue.

The *impact* of our work is two-fold, providing both an intrinsic manifold validation method, and a way to assess the uniformity of data.

## 2 Background

We begin by providing the requisite definitions from geometry, topology, and statistics.

### 2.1 Fundamental Definitions From Topology and Geometry

We begin with the definitions from topology and geometry. For an additional discussion, we direct readers to Appendix B.2. Recall that a metric space  $(\mathbb{X}, d)$  is written as a set  $\mathbb{X}$  paired with a distance metric  $d : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_{\geq 0}$ . We write  $\mathbb{B}_d(x, r)$  to denote the open ball of radius  $r$  centered at  $x \in (\mathbb{X}, d)$ . For ease of notation, we often write  $\mathbb{X}$ , with the distance assumed. We say that  $\mathbb{X}$  is an  $n$ -dimensional manifold if, at every point in  $\mathbb{X}$ , there exists a neighborhood homeomorphic to the unit ball in  $\mathbb{R}^n$ .

In particular, we are interested in Riemannian manifolds in this work. Let  $\mathbb{X}$  be a smooth manifold, and let  $d : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_{\geq 0}$  be a distance. We say that  $(\mathbb{X}, d)$  is a *Riemannian manifold* if  $d$  is a Riemannian metric. In this case,  $\mathbb{X}$  has a well-defined Lebesgue measure  $\mu$  [1, §VII.1], making  $(\mathbb{X}, d, \mu)$  a metric measure space. For example, shortest-path *geodesic* distances on  $\mathbb{X}$ , which we denote by  $d_{\mathbb{X}}$ , is a Riemannian metric. Core to our methods are uniform samples of a manifold:

► **Definition 1** (Uniformly Sampled Manifold Representation). *Let  $\mathbb{X}$  be a manifold with geodesics defined. The geodesic distance function on  $\mathbb{X}$  is  $d_{\mathbb{X}} : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_{\geq 0}$  where  $d_{\mathbb{X}}(x, y)$  denotes the geodesic distance between  $x$  and  $y$ . For a uniform, finite sample  $S \subset \mathbb{X}$ , we can encode  $S$  by the pairwise distances  $d_{\mathbb{X}}(x, y) \forall x, y \in S$  within a distance matrix  $D \in M^{|S| \times |S|}(\mathbb{R}_{\geq 0})$ .*

Sometimes, instead of having  $D$  itself, we might have an approximation (based on neighborhood graphs, for instance) of it, denoted  $\tilde{D}$  throughout this manuscript. In this paper, properties of balls in  $\mathbb{R}^n$  provide a natural backdrop to study more interesting spaces. It is well known that in Euclidean space, an open ball of radius  $r$  under the standard Euclidean metric (denoted  $\mathbb{B}_2(x, r)$ ) has volume

$$\text{Vol}(\mathbb{B}_2(x, r)) := \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} r^n, \quad (1)$$

where  $\Gamma$  is the gamma function. Let  $\mathbb{X}$  be a Riemannian manifold. We write the Gaussian curvature at a point  $x \in \mathbb{X}$  as  $\mathcal{G}(x)$ , and the scalar curvature at  $x \in \mathbb{X}$  as  $\mathcal{S}(x)$ . Note that for two-manifolds,  $\mathcal{S}(x) = 2\mathcal{G}(x)$ . We often compare the volume of balls in  $\mathbb{R}^n$  with balls on curved manifolds, whose relationship is expressed by the following ratio:

► **Theorem 2** (Relation Between Volume and Curvature). *Assume  $(\mathbb{X}, d_{\mathbb{X}})$  is a Riemannian  $n$ -manifold. Let  $x \in \mathbb{X}$  a point with scalar curvature  $\mathcal{S}(x)$ . For sufficiently small  $r > 0$ , the ratio between the volume of the ball  $\mathbb{B}_{d_{\mathbb{X}}}(x, r) \subseteq \mathbb{X}$  and the volume of the Euclidean ball  $\mathbb{B}_2(0, r) \subset \mathbb{R}^n$  centered at zero is:*

$$\frac{\text{Vol}(\mathbb{B}_{d_{\mathbb{X}}}(x, r))}{\text{Vol}(\mathbb{B}_2(0, r))} = 1 - \frac{\mathcal{S}(x) \cdot r^2}{6(n+2)} + o(r^2).$$

Theorem 2 allows us to inspect the relationship between volume and curvature in Riemannian manifolds, and is proven in classical textbooks, including [15, p. 168] and [6, p. 317]. Likewise, the seminal Gauss-Bonnet theorem describes the relationship between the total curvature of  $\mathbb{X}$  and its topology.

► **Theorem 3** (Gauss-Bonnet[23]). *Let  $\mathbb{X}$  be a compact Riemannian two-manifold, and let  $x \in \mathbb{X}$ . Then, the total curvature on  $\mathbb{X}$  is  $\int_{\mathbb{X}} \mathcal{G} dA = 2\pi\chi(\mathbb{X})$ , where  $dA$  is the area element on  $\mathbb{X}$  and  $\chi(\mathbb{X})$  is the Euler characteristic.*

In higher dimensions, generalizations of the Gauss-Bonnet theorem become much more intricate and are only defined for even dimensions. Hence, in dimensions larger than two, we use properties of the total mean curvature of submanifolds of  $\mathbb{R}^n$ , given in [7]. In particular, we relate scalar curvature, mean curvature, and the second fundamental form using the Gauss-Codazzi equations for hypersurfaces; see Appendix B.4 for definitions.

► **Theorem 4** (Gauss-Codazzi Equations for Hypersurfaces [16]). *Let  $\mathbb{X}$  be a Riemannian  $n$ -dimensional manifold immersed in  $\mathbb{R}^m$ . Then, the scalar curvature  $\mathcal{S}$  of  $\mathbb{X}$  at a point  $p \in \mathbb{X}$  is expressed by  $\mathcal{S} = H^2 - \|h\|^2$ , where  $H$  is the mean curvature of  $\mathbb{X}$  at  $p$ , and  $h$  is the second fundamental form of  $\mathbb{X}$  at  $p$ ; see Definition 28.*

## 2.2 Fundamental Definitions from (Spatial) Statistics

Let  $(\mathbb{X}, d, \mu)$  be a metric measure space; that is,  $\mathbb{X}$  is a topological space,  $d: S \times S \rightarrow \mathbb{R}$  is a metric, and  $\mu$  is a measure over a suitable collection of subsets of  $\mathbb{X}$ . Then, let  $X = \{X_i\}_{i=1}^m$  be  $m$  samples drawn iid from the uniform distribution over  $S$ . Let  $A$  be a measurable

subset of  $S$ , and, for each  $i$ , let  $A_i$  be the random variable that is one if  $X_i \in A$  and zero otherwise. Since each  $X_i$  is drawn iid from the uniform distribution over  $\mathbb{S}$ , we know that in any realization of  $X$ , the expected number of points landing in  $A$  is:

$$\mathbb{E}[\#(X \cap A)] = \sum_{i=1}^m \mathbb{E}[A_i] = \sum_{i=1}^m \mathbb{P}[X_i \in A] = \frac{m\mu(A)}{\mu(S)}. \quad (2)$$

Similar to the Buffon needle problem (e.g., [31, pp. 71-2]), we can use this property to estimate geometric properties of  $S$  (and of  $A$ ) using the law of large numbers.

► **Example 5 (Estimating Areas).** Consider  $A \subseteq S \subset \mathbb{R}^2$ , such that both  $S$  and  $A$  are Lebesgue-measurable. Then, for  $m$  sufficiently large, the number of points that land in  $A$  is approximately the ratio of the area of  $A$  to the area of  $S$ :

$$|A \cap X| \approx \frac{n \cdot \text{Area}(A)}{\text{Area}(\mathbb{X})}. \quad (3)$$

Multiplying both sides by  $\text{Vol}(\mathbb{X})$ , we see that  $\text{Area}(A) \approx \frac{n_A}{m} \cdot \text{Area}(\mathbb{X})$ , where  $n_A$  denotes the number of sample points that land in  $A$ . So, if we know  $\text{Area}(S)$ , we can use a realization of  $X$  to estimate  $\text{Area}(A)$ . This generalizes to Riemannian manifolds.

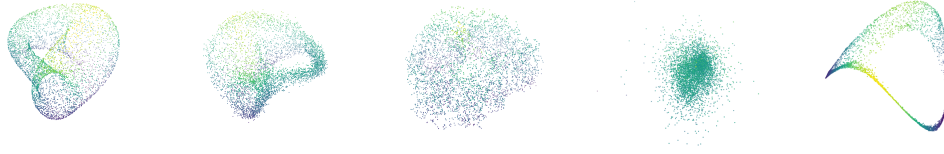
We note here that this only holds if the points are sampled iid from the uniform distribution over  $S$ . If the points are sampled from some other distribution, then this area estimation technique would not work for all measurable subspaces  $A$ .

## 2.3 Manifold Learning and Validation

Having given the definitions relevant to manifolds themselves, we lay out the paradigm on manifold learning and its validation. Broadly speaking, manifold learning operates on the assumption that data is sampled uniformly from a manifold  $\mathbb{X}$ , and attempts to learn  $\mathbb{X}$  by approximating geodesic distances [14, 26]. Because of the variety in approaches, there is no rigid definition of a manifold learning algorithm. We give a general definition of *manifold learning* that will be used throughout this work, which is informed by the core structures of manifolds themselves. That is, we consider arbitrary manifold learning methods as a map from set of data to a distance matrix:

► **Definition 6 (Manifold Learning).** A manifold learning *model*  $F: S \rightarrow M^{|S| \times |S|}(\mathbb{R}_{\geq 0})$  is a map from a set of data  $S$  (assumed to be sampled uniformly from an  $n$ -dimensional ambient manifold) to an  $|S| \times |S|$  real-valued distance matrix. We can think of  $F$  as a map approximating geodesic distances, so we typically write  $F(S) = \tilde{D}$ , and we write  $D$  as the true geodesic distance matrix for all points of  $S$ .

It should be noted that many manifold learning techniques compute an *embedding* rather than a distance matrix, wherein geodesics are implicit. In such cases, we can compute a geodesic distance matrix by constructing a neighbor graph in the space of the embedding and computing graph distances. We refer the readers to [19, 39, 17] for descriptions of many different manifold learning techniques and frameworks. In particular, we work closely with ISOMAP [35], leveraging its computation of  $\tilde{D}$  as an intermediary step. Generally, our paper considers manifold learning in an *unsupervised* setting, assuming no knowledge of the ground truth geodesic distances,  $D$ . In this setting, validation must be done using properties *intrinsic* to the “learned manifold” approximated by  $\tilde{D}$  alone. In general, intrinsic validation approaches have been quite successful in other areas of machine learning, for example silhouette coefficients for clustering [29], which rely only on properties intrinsic to clusters themselves.



**Figure 1** Constructing 3D embeddings of a Klein bottle lifted in ten dimensions, with varying success. From left to right, an output from PCA, ISOMAP, t-SNE, LLE, and spectral embedding. From visual inspection, PCA appears to have performed best in this example, followed by ISOMAP.

### 3 An Intrinsic Manifold Validation Method

Let  $(\mathbb{X}, d_{\mathbb{X}})$  be a Riemannian  $n$ -manifold. Let  $X \subset \mathbb{X}$  be a uniform sample on  $\mathbb{X}$ . By Theorem 2, for  $x \in \mathbb{X}$  and  $r > 0$  sufficiently small, we can write the volume of a ball  $\mathbb{B}_2(0, r)$  in  $\mathbb{R}^n$  as a function of the scalar curvature at  $x$  and the volume of the ball  $\mathbb{B}_{d_{\mathbb{X}}}(x, r)$  in  $\mathbb{X}$  for small  $r$ :

$$\text{Vol}(\mathbb{B}_2(0, r)) = \frac{\text{Vol}(\mathbb{B}_{d_{\mathbb{X}}}(x, r))}{\left(1 - \frac{\mathcal{S}(x) \cdot r^2}{6(n+2)} + o(r^2)\right)}. \quad (4)$$

Furthermore, if  $\text{Vol} \mathbb{X}$  is known, we can estimate By Equation (2) and the law of large numbers, we know that if  $|X|$  is sufficiently large, then the number of points landing in any measurable region  $A \subset \mathbb{X}$  is equal to  $\frac{|A|}{\text{Vol}(\mathbb{X})}$ . Thus, if we know  $\text{Vol}(\mathbb{X})$ , we can use the sample to estimate  $\text{Vol}(A)$ , and vice versa. Putting these two observations together, we define the manifold density function:

**Definition 7** (Manifold Density Function And Its Estimators). *Let  $(\mathbb{X}, d_{\mathbb{X}})$  be a Riemannian  $n$ -manifold, and let  $X \subset \mathbb{X}$  be a uniform sample on  $\mathbb{X}$ . Let  $p \in \mathbb{X}$ , and let  $R > 0$  be sufficiently small. Then, the manifold density function is the function  $K_{\mathbb{X}}: [0, R] \rightarrow \mathbb{R}$  defined by*

$$K_{\mathbb{X}}(r) := \frac{\text{Vol}(\mathbb{B}_2(0, r))}{\text{Vol}(\mathbb{X})} \quad (5)$$

with local estimator  $\hat{K}_p: [0, R] \rightarrow \mathbb{R}$  defined by

$$\hat{K}_p(r) := \left(1 - \frac{\mathcal{S}(p) \cdot r^2}{6(n+2)}\right)^{-1} \cdot \frac{1}{|X|} \sum_{x \in X} I(x \in B_r(p)) \quad (6)$$

and an aggregated estimator  $\hat{K}: [0, R] \rightarrow \mathbb{R}$  defined by

$$\hat{K}(r) := \frac{1}{|X|^2} \sum_{p \in X} \left( \left(1 - \frac{\mathcal{S}(p) \cdot r^2}{6(n+2)}\right)^{-1} \cdot \sum_{x \in X} I(x \in B_r(p)) \right) \quad (7)$$

If  $\mathbb{X}$  is flat, then  $\mathcal{S}(p) = 0$ , simplifying  $\hat{K}_p$  and  $\hat{K}$ . That is, we remove the first term of  $\hat{K}_p$  and  $\hat{K}$ , giving  $\hat{K}_p(r) := \frac{1}{|X|} \sum_{x \in X} I(x \in B_r(p))$  and  $\hat{K}(r) := \frac{1}{|X|^2} \sum_{p \in X} \sum_{x \in X} I(x \in B_r(p))$ .

The manifold density function is related to Ripley's  $K$ -function, which we define in Appendix A. Historically, the Ripley's  $K$ -function is used to study the uniformity of distributions in Euclidean space and in spheres.

### 3.1 The Manifold Score

We call the  $L_2$ -distance between a function and its estimate, for example,  $\|K_{\mathbb{X}} - \hat{K}\|_2$ , the error of the estimate. Since the estimators drop an  $o(r^2)$  term from Equation (4), we note that  $\hat{K}_p$  and  $\hat{K}$  are biased estimators of  $K$ ; that is, for all  $r \geq 0$ , we have  $\hat{K}_p(r) \geq K(r)$  and there exists cases where equality doesn't hold.

Finally, we define a score by taking a simple normalization of the error of an estimator of  $K_{\mathbb{X}}$ .

► **Definition 8 (Manifold Score).** Let  $\mathbb{X}$  be a compact Riemannian manifold, and  $X \subset \mathbb{X}$  a finite sample. The local manifold score for  $X$  as representing the manifold  $\mathbb{X}$  at  $p \in X$  is

$$M_p(X) = 1 - \frac{1}{|X|} \cdot \|K_{\mathbb{X}} - \hat{K}_p\|_2, \quad (8)$$

and the aggregated manifold score is

$$M(X) = 1 - \frac{1}{|X|} \cdot \|K_{\mathbb{X}} - \hat{K}\|_2. \quad (9)$$

This score indicates the extent to which a sample resemles a uniformly distributed set of points on a manifold.

Before proceeding further, we prove some basic but important properties of manifold density function, and the associated error from the theoretical manifold density function.

► **Lemma 9 (Range of Local Manifold Density Functions).** Let  $\mathbb{X}$  be a compact Riemannian manifold, let  $X \subset \mathbb{X}$  a finite uniform sample, and fix a point  $p \in X$ . For the local manifold density function, we have  $\hat{K}_p(r) \in [0, 1]$  for each radius  $r$ .

**Proof.** If  $r = 0$ , then  $\mathbb{B}_{\mathbb{X}}(p, 0) = \emptyset$ . Thus, by definition,  $\hat{K}_p(r) = 0$ . Additionally, by the compactness of  $\mathbb{X}$ , there exists  $r_{max} \geq 0$  such that  $\mathbb{B}_{\mathbb{X}}(p, r_{max})$  covers all of  $X$ . This implies  $\sum_{x \in X} I(x \in \mathbb{B}_{\mathbb{X}}(p, r_{max})) = |X|$ . Increasing  $r_{max}$  by  $\delta > 0$  continues to have  $\mathbb{B}_{\mathbb{X}}(p, r_{max} + \delta)$  covering all of  $X$ , so the maximum value attained by  $\hat{K}_p(r)$  is  $|X|/|X| = 1$ . ◀

The same result holds for aggregated manifold density function.

► **Lemma 10 (Range of Aggregated Manifold Density Function).** Let  $\mathbb{X}$  be a compact Riemannian manifold, let  $X \subset \mathbb{X}$  a finite uniform sample, and fix a point  $p \in X$ . For the aggregated manifold density function, we have  $\hat{K} \in [0, 1]$  for each radius  $r$ .

**Proof.** Let  $p, x \in X$ . Again, the minimum value attained by  $\hat{K}$  is vacuously 0, by setting  $r = 0$ . By Lemma 9, the range of  $\hat{K}_p(r)$  is  $[0, 1]$  for every  $p \in X$ . Denote  $r_{max}^x$  as the minimum radius needed such that  $\mathbb{B}_{\mathbb{X}}(x, r_{max}^x)$  covers all of  $X$ . Next, find the  $x \in X$  such that  $r_{max}^x \geq r_{max}^p$  for any other  $p \neq x \in X$ . Then  $\hat{K}_p(r_{max}^x) = 1$  for every  $p \in X$ , so we have  $\hat{K}_{D_X}(r_{max}^x) = \frac{|X| \cdot 1}{|X|} = 1$ . Hence, the range of  $\hat{K}$  is again  $[0, 1]$ . ◀

With the possible range of  $\hat{K}_p$  and  $\hat{K}$  established, we now bound the error of any  $\hat{K}_p$  and  $\hat{K}$  with respect to the theoretical manifold density function,  $K_{\mathbb{X}}$ .

► **Lemma 11 (Bounding Error).** Let  $\mathbb{X}$  be a compact Riemannian manifold, and  $X \subset \mathbb{X}$  a finite uniform sample. The maximum error of  $\hat{K}$  is  $\|K_{\mathbb{X}} - 1\|_2$ .

**Proof.** From Lemma 10, the maximum value attained by  $\hat{K}$  is 1, which is possible to attain for every  $r \geq 0$  if all  $x \in X$  coincide. ◀

From Lemma 9, Lemma 10, and Lemma 11 we guarantee the well-definedness and normalization of  $M$ . Hence, Definition 8 quantifies how “manifold-like” a distance matrix  $D_X$  is on the interval  $[0, 1]$ . Intuitively, a manifold score of zero indicates that an uncovered sample from some manifold learning technique was completely nonuniform (with all points of  $D_X$  coinciding), and therefore the manifold learning model can be thought to have performed poorly. On the other hand, a manifold score of one says that an uncovered sample was perfectly uniform, that is, identical to the theoretical manifold density function. By assumption, since  $D_X$  was an approximation of a uniform sample  $X$  on an ambient manifold  $\mathbb{X}$ , we consider  $D_X$  to be a good approximation in this case.

### 3.2 Manifold Density Functions For Manifolds With Curvature

We now introduce a primary theoretical result of the paper, which is an adaptation of manifold density function to manifolds with curvature. We begin with a brief discussion of local manifold density functions on two-manifolds and their shortcomings, and then give stronger results for aggregated manifold density functions on two-manifolds using the Gauss-Bonnet theorem. We then provide a higher dimensional generalization for hypersurfaces, using inequalities related to the first Laplacian eigenvalue. Related extensions for Ripley’s  $K$ -function to more general settings have been attempted before, and include extensions for spheres [22, 38], and for fibers [32]. We present the first extension of its kind for general Riemannian two-manifolds and hypersurfaces.

At its core, our approach is designed with the specific intent to assess the uniformity of a sample lying on arbitrary manifolds rather than within Euclidean space alone. By its very formulation, for an arbitrary point  $p$  on a Riemannian manifold  $\mathbb{X}$ , the local manifold density function is defined in terms of the scalar curvature  $\mathcal{S}(p)$  at  $p$ . If we are equipped with knowledge of the scalar curvature  $\mathcal{S}(p)$  at  $p$ , then we can compute the local manifold density function directly. A straightforward, but desirable, consequence of this definition is the following theorem, which allows us to understand the local manifold density function in terms of Euclidean balls:

► **Theorem 12.** *Let  $\mathbb{X}$  be a compact Riemannian manifold, and let  $X \subset \mathbb{X}$  be a uniform sample on  $\mathbb{X}$ . Let  $p \in X$ . Then, for sufficiently small  $r > 0$  and large  $|X|$ ,  $\hat{K}_p(r)$  converges to the theoretical manifold density function  $K_{\mathbb{X}}(r)$ .*

**Proof.** Expanding definitions,

$$\hat{K}_p(r) = \left(1 - \frac{\mathcal{S}(p) \cdot r^2}{6(n+2)}\right)^{-1} \cdot \frac{1}{|X|} \sum_{x \in X} I(x \in B_r(p)) \quad (10)$$

$$= \frac{\text{Vol}(\mathbb{B}_2(r, 0))}{\text{Vol}(\mathbb{B}_{\mathbb{X}}(r, x))} \cdot \frac{1}{|X|} \sum_{x \in X} I(x \in B_r(p)) \text{ for sufficiently small } r \quad (11)$$

$$\rightarrow \frac{\text{Vol}(\mathbb{B}_2(r, 0))}{\text{Vol}(\mathbb{B}_{\mathbb{X}}(r, x))} \cdot \frac{\text{Vol}(\mathbb{B}_{\mathbb{X}}(r, x))}{\text{Vol}(\mathbb{X})} \text{ by the law of large numbers} \quad (12)$$

$$= \frac{\text{Vol}(\mathbb{B}_2(r, 0))}{\text{Vol}(\mathbb{X})} \quad (13)$$

$$= K_{\mathbb{X}}(r). \quad (14)$$

Where we use the standard area estimation procedure using sampling on Riemannian manifolds, see Example 5. It follows that  $\hat{K}_p(r)$  converges to  $K_{\mathbb{X}}(r)$  for large sample size  $|X|$ , and sufficiently small  $r$ . ◀



However, one should note that there are a few undesirable properties that go along with Definition 7 in the local setting. Most notably, our focus is on *intrinsic* validation, and knowing the scalar curvature at any point  $p \in X$  violates the true spirit of an intrinsic method. Additionally, Theorem 2 only technically holds for balls with sufficiently small radius  $r > 0$ . As such, the scaling in Definition 7 becomes less accurate in approximating  $K_{\mathbb{X}}$  as  $r$  increases. The latter problem can be avoided by considering the manifold score restricted to small enough radii, and the former is resolved by considering global properties of the aggregated manifold density function.

### 3.3 Aggregated Manifold Density Functions for Surfaces

Let  $\mathbb{X}$  a compact Riemannian two-manifold, and let  $X \subset \mathbb{X}$  be a uniform sample of  $X$ . In the aggregated setting, since  $\hat{K}$  is a global average of every manifold density function on  $X$ , we are able to avoid the shortcomings in the local setting by exploiting fundamental results in differential geometry. Namely, we first recall Theorem 3, the Gauss-Bonnet theorem, which categorizes the relationship between geometry and topology for two-manifolds, establishing that the total curvature of a manifold is a function of its Euler characteristic. In addition, we recall that the scalar curvature at a point  $p \in \mathbb{X}$  on a two-manifold is twice the Gaussian curvature:  $\mathcal{G}(p) = 2\mathcal{S}(p)$ . Considering the total curvature over all of  $\mathbb{X}$  and the ratio given in Theorem 2, we can relate the total area of balls of radius  $r > 0$  in  $\mathbb{X}$  to the total volume of radius  $r$  balls in  $\mathbb{R}^2$  for sufficiently small  $r$ .

**► Lemma 13 (Average Volume Distortion of Balls).** *Let  $\mathbb{X}$  be a compact Riemannian two-manifold. Then, on average, the ratio  $\frac{\text{Vol}(\mathbb{B}_{\mathbb{X}}(p, r))}{\text{Vol}(\mathbb{B}_2(0, r))}$  between the volume of Euclidean and geodesic balls for a point  $p \in \mathbb{X}$  is  $1 - \frac{\pi\chi(\mathbb{X})}{24 \cdot A} \cdot r^2 + o(r^2)$ , where  $A$  is the area form of  $\mathbb{X}$  and  $\chi(\mathbb{X})$  is the Euler characteristic.*

**Proof.** We expand definitions and integrate over the area form  $dA$  of  $\mathbb{X}$ :

$$\frac{1}{A} \int_{\mathbb{X}} \frac{\text{Vol}(\mathbb{B}_{d_{\mathbb{X}}}(x, r))}{\text{Vol}(\mathbb{B}_2(0, r))} dA = \frac{1}{A} \left[ A - \int_{\mathbb{X}} \frac{\mathcal{S}(x)}{6(n+2)} \cdot r^2 dA + Ao(r^2) \right] \text{ by Equation (4)} \quad (15)$$

$$= 1 - \frac{1}{A} \int_{\mathbb{X}} \frac{\mathcal{G}(x)}{48} \cdot r^2 dA + o(r^2) \quad (16)$$

$$= 1 - \frac{\pi\chi(\mathbb{X})}{24 \cdot A} \cdot r^2 + o(r^2), \quad (17)$$

giving a formula for the average ratio between the volume of Euclidean and geodesic balls in terms of the Euler characteristic, accompanied by an error term with value  $o(r^2)$ . ◀

If we take  $A \approx r^2$  and drop the error term,  $o(r^2) = 0$ , we obtain a simple approximation for the ratio in terms of the Euler characteristic, i.e.

$$\frac{1}{A} \int_{\mathbb{X}} \frac{\text{Vol}(\mathbb{B}_{d_{\mathbb{X}}}(x, r))}{\text{Vol}(\mathbb{B}_2(0, r))} dA \approx 1 - \frac{\pi\chi(\mathbb{X})}{24}.$$

We thus have a definition of total area distortion due to curvature on a two-manifold, which informs our scaling procedure of the aggregated manifold density function on a two-manifold. Namely, as a consequence of Lemma 13, the aggregated manifold density function on a general two-manifold with curvature is *invariant* of its Euler characteristic, and converges to the standard manifold density function as the sample size increases:

**► Theorem 14.** *Let  $\mathbb{X}$  be a compact Riemannian two-manifold, and let  $X \subset \mathbb{X}$  be a uniform sample on  $\mathbb{X}$ . For large  $|X|$  and sufficiently small  $r > 0$ ,  $\hat{K}(r)$  converges to  $K_{\mathbb{X}}(r)$ .*



289 **Proof.** We examine the result of Theorem 12 when considered globally on  $\mathbb{X}$ .

$$\begin{aligned}
 290 \quad \hat{K}(r) &= \frac{1}{|X|^2} \sum_{p \in X} \left( \left( 1 - \frac{\mathcal{S}(p) \cdot r^2}{6(n+2)} \right)^{-1} \cdot \sum_{x \in X} I(x \in B_r(p)) \right) \\
 291 \quad &\rightarrow \frac{1}{A^2} \int_{\mathbb{X}} \frac{\text{Vol}(\mathbb{B}_{d_{\mathbb{X}}}(p, r)) dA}{\left( 1 - \frac{\mathcal{S}(p)}{6(n+2)} * r^2 \right)} \text{ for large } |X| \\
 292 \quad &\rightarrow \frac{1}{A^2} \int_{\mathbb{X}} \text{Vol}(\mathbb{B}_2(0, r)) dA \text{ for small enough } r, \text{ by Equation (4)} \\
 293 \quad &= \frac{\text{Vol}(\mathbb{B}_2(0, r))}{A^2} \cdot A \\
 294 \quad &= \text{Vol}(\mathbb{B}_2(0, r)) / A = K_{\mathbb{X}}(r).
 \end{aligned}$$

295 Where we integrate over the area form  $A$  of  $\mathbb{X}$ . Moreover, the second to last equality  
 296 follows from the fact that the volume of Euclidean balls  $\text{Vol}(\mathbb{B}_2(0, r))$  is taken as a constant  
 297 over  $\mathbb{X}$  for a fixed radius  $r$ , and thus we integrate only over  $dA$ . Hence, for large  $|X|$ ,  
 298  $\hat{K}(r) \rightarrow K_{\mathbb{X}}(r)$ .  $\blacktriangleleft$

299 Examining the integral in the denominator of Line 2 in Theorem 14 gives a strong  
 300 approximation of the aggregated manifold density function  $\hat{K}$  on a surface, which depends  
 301 on the global topology of  $\mathbb{X}$  due to the Gauss-Bonnet theorem and making use of the fact  
 302 that  $\mathcal{G}(p) = 2\mathcal{S}(p)$  for surfaces:

303 **► Corollary 15.** *Let  $\mathbb{X}$  be a compact Riemannian manifold, and let  $X \subset \mathbb{X}$  be a uniform*  
 304 *sample on  $\mathbb{X}$ . We approximate:*

$$305 \quad \hat{K}(r) := \left( 1 - \frac{\pi \chi(\mathbb{X}) \cdot r^2}{24A} \right)^{-1} \cdot \frac{1}{|X|^2} \sum_{p \in X} \left( \sum_{x \in X} I(x \in B_r(p)) \right)$$

306 *Then as  $|X|$  increases,  $\hat{K}(r) \rightarrow K_{\mathbb{X}}(r)$ . Moreover, taking  $A = O(r^2)$ , we have the approxima-*  
 307 *tion dependent only on the Euler characteristic:*

$$308 \quad \hat{K}(r) := \left( 1 - \frac{\pi \chi(\mathbb{X})}{24} \right)^{-1} \cdot \frac{1}{|X|^2} \sum_{p \in X} \left( \sum_{x \in X} I(x \in B_r(p)) \right)$$

309 Consequently, in the aggregated setting we scale the average of local manifold density  
 310 functions by a function of the Euler characteristic and the area of  $\mathbb{X}$ . In fact, as we demonstrate  
 311 experimentally, the heuristic approximation of  $\hat{K}(r)$  given by taking  $A = O(r^2)$  is typically  
 312 sufficient, allowing us to scale the manifold density function using the Euler characteristic  
 313 alone. Assuming no knowledge of the Euler characteristic, we can simply estimate  $\chi(\mathbb{X})$  by  
 314 selecting an integer that minimizes  $M(\hat{K}(r))$ .

### 315 3.4 Aggregated Manifold Density Functions in High Dimensions

316 As is indicated by the ratio given in Theorem 2, the volume form of a Riemannian manifold  $\mathbb{X}$   
 317 depends on the scalar curvature, and in two dimensions we are able to use the Gauss-Bonnet  
 318 theorem alone due to the direct relation between  $\mathcal{S}(x)$  and  $\mathcal{G}(s)$  on surfaces. Unfortunately,  
 319 higher-dimensional generalizations of the Gauss-Bonnet theorem rely on the Pfaffian of  
 320  $\mathbb{X}$ , which is a function of Ricci, Riemannian, and scalar curvature. Instead, we scale the  
 321 aggregated manifold density function by a different global constant of manifolds, namely  $\lambda_1$ ,

the first eigenvalue of their Laplacian operator. We focus our attention where bounds to  $\lambda_1$  are known, thus dealing with orientable, compact hypersurfaces. In particular, two useful bounds on  $\lambda_1$  are given by the following theorems, which relate  $\lambda_1$  to the mean curvature  $H$ , and to the squared length of the second fundamental form  $h$ . See Appendix B.4 for more details.

► **Theorem 16** (Total Mean Curvature [7]). *Let  $\mathbb{X}$  be an  $n$ -dimensional, compact submanifold in  $\mathbb{R}^m$ , and  $H$  the mean curvature at any point  $p \in \mathbb{X}$ . Then*

$$\int_M |H|^k dV \leq \left( \frac{\lambda_q}{n} \right)^{\frac{k}{2}} \text{Vol}(\mathbb{X}),$$

where  $\lambda_q$  denotes the  $q$ th Eigenvalue of the Laplacian operator.

In addition, the length of  $h$  is related to  $\lambda_1$  in the following way:

► **Theorem 17** (Total Length of the Second Fundamental Form [7]). *Let  $\mathbb{X}$  be a compact orientable  $n$ -dimensional submanifold of  $\mathbb{R}^m$  with arbitrary codimension. Then,*

$$\int_{\mathbb{X}} \|h\|^2 dV \geq \lambda_1 \text{Vol}(\mathbb{X}).$$

Equipped with these results, we have a mechanism to scale manifold density functions in higher dimensions. Let  $\mathbb{X}$  be a compact,  $n$ -dimensional orientable manifold embedded in  $\mathbb{R}^{n+1}$ , i.e.,  $\mathbb{X}$  is a hypersurface in Euclidean space. Let  $r > 0$ , where  $r$  is sufficiently small to satisfy the ratio given in Theorem 2. As was done in Section 3.3, we compute the ratio of the total volume of balls of radius  $r$  on  $\mathbb{X}$  to the total volume of balls of radius  $r$  in  $\mathbb{R}^n$ . The following integral is taken over the volume form  $dV$  of  $\mathbb{X}$ , given formally in Appendix B.3:

$$\begin{aligned} \int_{\mathbb{X}} \frac{\text{Vol}(\mathbb{B}_{d_{\mathbb{X}}}(x, r))}{\text{Vol}(\mathbb{B}_2(0, r))} dV &= \text{Vol}(\mathbb{X}) - \int_{\mathbb{X}} \frac{\mathcal{S}(x)}{6(n+2)} \cdot r^2 dV + \text{Vol}(\mathbb{X})o(r^2) \\ &= \text{Vol}(\mathbb{X}) - \int_{\mathbb{X}} \frac{H^2 - \|h\|^2}{6(n+2)} \cdot r^2 dV + \text{Vol}(\mathbb{X})o(r^2) \quad (\text{by Theorem 4}) \\ &= \text{Vol}(\mathbb{X}) - \frac{r^2}{6(n+2)} \left( \int_{\mathbb{X}} H^2 dV - \int_{\mathbb{X}} \|h\|^2 dV \right) + \text{Vol}(\mathbb{X})o(r^2). \end{aligned}$$

Now, notice from Theorem 16 and Theorem 17:

$$\int_{\mathbb{X}} H^2 dV - \lambda_1 \text{Vol}(\mathbb{X}) \leq \int_{\mathbb{X}} H^2 dV - \int_{\mathbb{X}} \|h\|^2 dV \leq \frac{\lambda_1}{2n} \text{Vol}(\mathbb{X}) - \int_{\mathbb{X}} \|h\|^2 dV.$$

This leads to the following heuristic for the average ratio between geodesic and Euclidean balls:

$$\begin{aligned} &\frac{1}{\text{Vol}(\mathbb{X})} \int_{\mathbb{X}} \frac{\text{Vol}(\mathbb{B}_{d_{\mathbb{X}}}(x, r))}{\text{Vol}(\mathbb{B}_2(0, r))} dV \\ &\approx 1 - \frac{r^2}{6(n+2)\text{Vol}(\mathbb{X})} \left( \frac{\lambda_1}{2n} \text{Vol}(\mathbb{X}) - \lambda_1 \text{Vol}(\mathbb{X}) \right) + o(r^2) \\ &= 1 - \left( \frac{r^2 \cdot \lambda_1 (1-n)}{6(n)(n+2)} \right) + o(r^2) \end{aligned}$$

Consequently, we obtain a desired analog to the scaled manifold density function for two-manifolds by scaling the aggregated manifold density function of general hypersurfaces in accordance with their first Laplacian eigenvalue, and again dropping the error term.

► **Definition 18** (Hypersurface Aggregated Manifold Density Function). *Let  $\mathbb{X}$  a compact, orientable Riemannian  $n$ -manifold immersed in  $\mathbb{R}^{n+1}$ , and let  $X \subset \mathbb{X}$  be a uniform sample of  $\mathbb{X}$ . The aggregated manifold density function for hypersurfaces  $\hat{K}(r)$  is approximated by:*

$$\hat{K}(r) := \left(1 - \frac{r^2 \cdot \lambda_1(2n-1)}{12(n)(n+2)}\right)^{-1} \cdot \frac{1}{|X|^2} \sum_{p \in X} \left( \sum_{x \in X} I(x \in B_r(p)) \right)$$

Note that for many common hypersurfaces,  $\lambda_1$  is either known, or there exists a “reasonable” approximation; see Appendix B.5 for examples. Given any fundamental knowledge of the ambient hypersurface, (e.g., if  $\mathbb{X}$  can reasonably be assumed to lie on a hypersphere, or similar manifold), we choose  $\lambda_1$  explicitly from the known approximations. Otherwise, analogously to tuning the manifold density function by choosing the Euler characteristic  $\chi(\mathbb{X})$  minimizing  $M(\hat{K})$ , we choose the approximation for  $\lambda_1$  minimizing  $M(\hat{K})$ .

## 4 Desiderata

We remark on desirable properties when validating algorithms in machine learning, and assess our manifold validation technique in this context.

### 4.1 Robustness

It is pertinent when attempting the validation of random samples to achieve stability against noise. Intuitively, given a sample  $X \subset \mathbb{X}$  on a (flat or curved) Riemannian manifold  $\mathbb{X}$ , this implies that subjecting a point  $p \in X$  to small perturbation  $\delta > 0$  should change the validation score of  $M(p)$  very little. We thus examine the local and aggregated manifold scores  $M(\hat{K}_p)$  and  $M(K_{\mathbb{X}})$  with respect to their stability after subjecting points of  $X$  to noise. In the aggregated setting, it is not difficult to make robustness guarantees about the manifold density function, and hence the manifold score. Specifically, for a point  $p \in X$ , examine  $p \pm \delta$ . Indeed, for a fixed radius  $r > 0$ , in the worst case,  $\hat{K}_p(r)$  could change at most by  $|X|$ , and for every other  $p' \neq p \in \mathbb{X}$ ,  $\hat{K}_{p'}(r)$  could change by at most 1. This alters the aggregated manifold density function,  $K_{\mathbb{X}}$ , by at most  $\pm 2/|X|$ . However, a small change by  $\delta$  to  $r$  gives the identical manifold density function prior to the perturbation, i.e.,  $\hat{K}_p(r \pm \delta) = \hat{K}_p(r)$ . The worst-case bound remains the same if we allow perturbations of every  $p \in \mathbb{X}$  by  $0 \leq \delta' \leq \delta$  for both  $K_{\mathbb{X}}$  and  $\hat{K}_p(r)$  when considering a fixed radius  $r$ . Hence, we can consider the aggregated manifold density function to be resistant to noise for fixed radii, and both the local and aggregated manifold density functions can be considered resistant to noise when allowing a change in radius  $r \pm \delta$ . Consequently, our manifold score  $M(K_{\mathbb{X}})$  is robust to noise up to the equivalent thresholds.

### 4.2 Computation

Let  $X \subset \mathbb{X}$ , where  $\mathbb{X}$  is an  $n$ -dimensional Riemannian manifold with  $|X| = m$ . It is immediate from our construction that adjusting the computation of the aggregated manifold density function to arbitrary manifolds only adds a constant factor in runtime complexity, under reasonable assumptions about the complexity of a manifold in relation to  $m$ . For surfaces, assuming that  $\chi(\mathbb{X})$  is a fixed constant, we could have  $O(1)$  manifold density function estimations, each of which adds only an  $O(1)$  factor in finding the value for  $\chi(\mathbb{X})$  minimizing  $M(K_{\mathbb{X}})$ . For hypersurfaces, again we search over  $O(1)$  known formulas for  $\lambda_1$ , each of which scale computation of  $K_{\mathbb{X}}$  by an  $O(1)$  factor. Given an algorithm to compute Ripley’s

$K$ -function, we can compute manifold density function adding only an  $O(1)$  factor. Naively, computing the disaggregated manifold density function is done for each point  $p \in X$  with a scan of points within radius  $r$  of  $p$  in  $\Theta(n)$  time, taking  $\Theta(n^2)$  time overall to compute the aggregated manifold density function. More efficient implementations are likely possible by efficiently embedding the given distance matrix in  $\mathbb{R}^n$ , and then using an  $n$ -dimensional range tree to compute  $\hat{K}_p(r)$  for a given radius  $r$ . Doing so is possible in  $O(\log^n m + k)$  time, where  $k = \hat{K}_p(r)$ . Finding  $K_X$  overall is done by computing  $\hat{K}_p(r)$  for every  $p \in X$ , which takes  $O(m(\log^n m + k_{max}))$  time, where  $k_{max}$  is the largest number of points in any ball of radius  $r$  throughout computation of the manifold density function. If taken throughout the entire range of  $r$ , this is  $O(m^2 \log^n m)$  time; but in general we can restrict the manifold density function to much smaller radii to satisfy the scaling from Theorem 2, which could improve the bound depending on the distribution of  $X$ . In practice, optimized implementations for Ripley’s  $K$ -function exist for a variety of applications, such as [34, 33, 32, 18].

## 5 Experimental Results

We present our experimental findings, comparing the manifold score of samples on popular manifolds. For a complete elaboration on our experiments, we refer to Appendix C.

### 5.1 Results on Flat Manifolds

As an initial verification, we test the framework on an embedding of the flat 2-torus, where we expect the manifold score to be nearly perfect (close to one) for a uniform sample. We compare this against a stratification on the flat torus, sampled uniformly in the form of a “cross,”  $\{(x, y) : x = \frac{1}{2} \text{ or } y = \frac{1}{2}\} \subset [0, 1) \times [0, 1)$ . We also consider a sample of the “cross”-stratification with noise, where ten percent of the sample’s points are uniformly sampled from the domain graph of the flat torus. Table 1 gives the average aggregated manifold score across ten trials on the flat 2-torus for these three samples. As anticipated, the manifold score is nearly perfect for the uniform sample, worse for the stratification with noise, and even worse for the pure stratification.

**Table 1** Manifold Score on the Flat Torus

|      | Sample Size | Uniform Sample      | Strat. with Noise   | Stratification      |
|------|-------------|---------------------|---------------------|---------------------|
| 100  |             | $0.9840 \pm 0.0056$ | $0.8807 \pm 0.0158$ | $0.6768 \pm 0.0490$ |
| 200  |             | $0.9921 \pm 0.0023$ | $0.8769 \pm 0.0178$ | $0.6410 \pm 0.0339$ |
| 500  |             | $0.9962 \pm 0.0010$ | $0.8763 \pm 0.0071$ | $0.6537 \pm 0.0210$ |
| 1000 |             | $0.9980 \pm 0.0009$ | $0.8728 \pm 0.0074$ | $0.6545 \pm 0.0241$ |

### 5.2 Surfaces with Curvature

We compare the manifold score of stratifications vs. uniform samples on both the sphere and the Klein bottle. We generate a distance matrix using ISOMAP with six neighbors, conducting ten trials for each experiment. For the sphere, we scale according to the approximation following Corollary 15 setting  $A \approx r^2$ . A uniform sampling and “cross”-stratification were considered, and the results are presented in Table 2. We note the impact of proper scaling



**Figure 2** From left to right: (1) uniform sample vs. (2) minor noise vs (3) stratification on the Klein bottle with  $|X| = 2000$ . The right plot compares the normalized manifold density functions of each, with (1) in blue, (2) in orange, (3) red, and the theoretical manifold density function in green.

on the manifold score. For the Klein bottle, we use no scaling since the Euler characteristic is zero. We again consider a uniform sample and “cross”-stratification, as well as a noisy sample obtained by sampling with probabilities defined by a normalized sine wave across the surface. As demonstrated by Table 3, the manifold score is effective in evaluating the sample’s representation of the Klein bottle.

For both surfaces, the manifold score converges to one (a perfect score) as the uniform sample size increases, which is explained by Theorem 12. However, this convergence is slower than for the flat torus, which we attribute to the introduction of a neighbors graph rather than using true geodesics, as well as the heuristic nature of the definition of the aggregated manifold density function for general two-manifolds.

**Table 2** Results on the Sphere

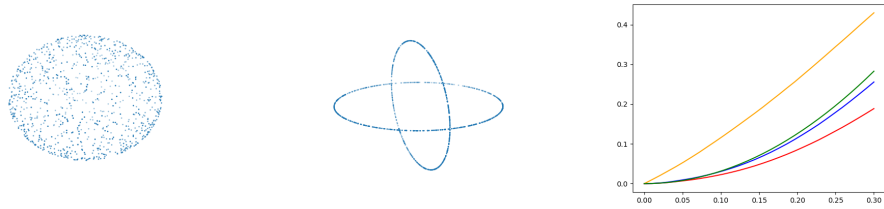
| Sample Size | Scaled Unif.        | Unscaled Unif.      | Stratification      |
|-------------|---------------------|---------------------|---------------------|
| 100         | $0.9415 \pm 0.0169$ | $0.7770 \pm 0.0147$ | $0.6116 \pm 0.1801$ |
| 500         | $0.9443 \pm 0.0083$ | $0.7834 \pm 0.0256$ | $0.6101 \pm 0.1842$ |
| 1000        | $0.9583 \pm 0.0151$ | $0.7840 \pm 0.0123$ | $0.6597 \pm 0.2110$ |
| 2000        | $0.9730 \pm 0.0096$ | $0.7963 \pm 0.0110$ | $0.6310 \pm 0.2382$ |

**Table 3** Results on the Klein Bottle

| Sample Size | Uniform             | Minor Noise         | Stratification      |
|-------------|---------------------|---------------------|---------------------|
| 200         | $0.8834 \pm 0.0571$ | $0.8114 \pm 0.0760$ | $0.3578 \pm 0.0926$ |
| 500         | $0.9109 \pm 0.0334$ | $0.8942 \pm 0.0386$ | $0.4048 \pm 0.1521$ |
| 1000        | $0.9430 \pm 0.0105$ | $0.8417 \pm 0.1286$ | $0.5054 \pm 0.0558$ |
| 2000        | $0.9529 \pm 0.0081$ | $0.8735 \pm 0.0520$ | $0.3075 \pm 0.0630$ |

### 5.3 Hypersurfaces

In higher dimensions, we use the results of Section 3.4 to examine uniform samples on hyperspheres in six, eight, ten, and fifteen dimensions. We demonstrate the manifold score of a uniform sample pre and post scaling, and the analogous stratification to the one in Figure 3



**Figure 3** From left to right: (1) uniform sample vs. (2) a stratification on the sphere, with  $|X| = 1000$ . The right plot compares the manifold density functions of each, with (1) scaled in blue, (1) unscaled in red, (2) in orange, and the theoretical manifold density function in green.

of two  $d/2$ -dimensional hyperspheres (scaled equivalently). For each experiment, we work with samples of size 1000 over ten trials. In higher dimensions the volume of Euclidean balls decreases, making the manifold score increasingly sensitive; yet we are still able to delineate between uniform samples and stratifications with the needed specificity. We provide the complete experiment and analysis for hypersurfaces in Appendix C.

**Table 4** Results on Hyperspheres of Differing Dimension

| Dimension | Scaled Unif.        | Unscaled Unif.      | Stratification      |
|-----------|---------------------|---------------------|---------------------|
| 6         | $0.9747 \pm 0.0022$ | $0.9143 \pm 0.0056$ | $0.8033 \pm 0.0023$ |
| 8         | $0.9803 \pm 0.0010$ | $0.9256 \pm 0.0034$ | $0.8782 \pm 0.0043$ |
| 10        | $0.9895 \pm 0.0004$ | $0.9458 \pm 0.0019$ | $0.9239 \pm 0.0029$ |
| 15        | $0.9968 \pm 0.0003$ | $0.9692 \pm 0.0033$ | $0.9697 \pm 0.0014$ |

## 6 Discussion

In this paper, we adapt and extend Ripley’s  $K$ -function to build a stable, efficiently computable, intrinsic framework for manifold validation. We then demonstrate how the manifold score can be used to evaluate and compare manifold learning algorithms. We make strong theoretical guarantees in the case for surfaces, and present a well-motivated heuristic for hypersurfaces. The effectiveness of the manifold density function for a variety of settings is verified experimentally. Further extensions to this work include consideration of broader classes of manifolds and surfaces, such as manifolds with boundary. Additionally, to further understand the use of this tool in applied settings, we wish to test the results of this framework in a wide array of datasets popular in applied manifold learning, especially biochemical data.

## References

- Herbert Amann, Joachim Escher, and Gary Brookfield. *Analysis III*. Springer, 2009.
- Adrian J Baddeley, Jesper Møller, and Rasmus Waagepetersen. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350, 2000.

- 496   **3**   Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction  
497       and data representation. *Neural Computation*, 15(6):1373–1396, 2003. doi:10.1162/  
498       089976603321780317.
- 499   **4**   Christopher M. Bishop, Markus Svensén, and Christopher K. I. Williams. Gtm: The gen-  
500       erative topographic mapping. *Neural Computation*, 10(1):215–234, 1998. doi:10.1162/  
501       089976698300017953.
- 502   **5**   Gunnar Carlsson, Tigran Ishkhanov, Vin De Silva, and Afra Zomorodian. On the local  
503       behavior of spaces of natural images. *International Journal of Computer Vision*, 76(1):1–12,  
504       2008.
- 505   **6**   Isaac Chavel. *Eigenvalues in Riemannian Geometry*. Academic Press, 1984.
- 506   **7**   Bang-Yen Chen. *Total Mean Curvature and Submanifolds of Finite Type*. 05 1984.
- 507   **8**   Sung Nok Chiu, Dietrich Stoyan, Wilfrid S Kendall, and Joseph Mecke. *Stochastic geometry*  
508       *and its applications*. John Wiley & Sons, 2013.
- 509   **9**   Noel Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.
- 510   **10**   Daryl J Daley, David Vere-Jones, et al. *An introduction to the theory of point processes:*  
511       *volume I: elementary theory and methods*. Springer, 2003.
- 512   **11**   Alireza Dirafzoon, Alper Bozkurt, and Edgar Lobaton. Geometric learning and topological  
513       inference with biobotic networks. *IEEE Transactions on Signal and Information Processing*  
514       *over Networks*, 3(1):200–215, 2016.
- 515   **12**   Philip M. Dixon. *Ripley’s K Function*. John Wiley & Sons, Ltd, 2014. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat07751>,  
516       arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118445112.stat07751>,  
517       doi:<https://doi.org/10.1002/9781118445112.stat07751>.
- 518       stat07751, doi:<https://doi.org/10.1002/9781118445112.stat07751>.
- 519   **13**   David Donoho and Carrie Grimes. Hessian eigenmaps: Locally linear embedding techniques for  
520       high-dimensional data. *proc. national academy of science (pnas)*, 100, 5591–5596. *Proceedings*  
521       *of the National Academy of Sciences of the United States of America*, 100:5591–6, 06 2003.  
522       doi:10.1073/pnas.1031596100.
- 523   **14**   Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis.  
524       *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.
- 525   **15**   S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Universitext. Springer Berlin  
526       Heidelberg, 2004. URL: [https://books.google.com/books?id=6F4Umpws\\_gUC](https://books.google.com/books?id=6F4Umpws_gUC).
- 527   **16**   Li Haizhong. Hypersurfaces with constant scalar curvature in space forms. *Mathematische*  
528       *Annalen*, 305(1):665–672, 1996.
- 529   **17**   Felix Hensel, Michael Moor, and Bastian Rieck. A survey of topological machine learning  
530       methods. *Frontiers in Artificial Intelligence*, 4, 2021. URL: <https://www.frontiersin.org/articles/10.3389/frai.2021.681108>, doi:10.3389/frai.2021.681108.  
531       org/articles/10.3389/frai.2021.681108, doi:10.3389/frai.2021.681108.
- 532   **18**   Alexander Hohl and Peilin Chen. Spatiotemporal simulation: local ripley’s k function para-  
533       meterizes adaptive kernel density estimation. In *Proceedings of the 2nd ACM SIGSPATIAL*  
534       *International Workshop on GeoSpatial Simulation*, pages 16–23, 2019.
- 535   **19**   Xiaoming Huo and Andrew Smith. A survey of manifold-based learning methods. *Recent*  
536       *Advances in Data Mining of Enterprise Data*, 01 2008. doi:10.1142/9789812779861\_  
537       0015.
- 538   **20**   Miao Jin, Wei Zeng, Ning Ding, and Xianfeng Gu. Computing Fenchel-Nielsen coordinates  
539       in Teichmüller shape space. In *2009 IEEE International Conference on Shape Modeling and*  
540       *Applications*, pages 193–200. IEEE, 2009.
- 541   **21**   Alan Karr. *Point processes and their statistical inference*. Routledge, 2017.



- 542 **22** Thomas Joseph Lawrence. Point pattern analysis on a sphere. Master's thesis, Master's thesis,  
543 The University of Western Australia, 2018.
- 544 **23** J.M. Lee. *Introduction to Riemannian Manifolds*. Graduate Texts in Mathematics.  
545 Springer International Publishing, 2019. URL: [https://books.google.com/books?id=](https://books.google.com/books?id=UIPltQEACAAJ)  
546 [UIPltQEACAAJ](https://books.google.com/books?id=UIPltQEACAAJ).
- 547 **24** Meghana Madhyastha, Gongkai Li, Veronika Strnadová-Neeley, James Browne, Joshua T.  
548 Vogelstein, Randal Burns, and Carey E. Priebe. Geodesic forests. In *Proceedings of the*  
549 *26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*,  
550 KDD '20, page 513–523, New York, NY, USA, 2020. Association for Computing Machinery.  
551 doi:10.1145/3394486.3403094.
- 552 **25** Jesper Møller and Ege Rubak. Functional summary statistics for point processes on the sphere  
553 with an application to determinantal point processes. *Spatial Statistics*, 18:4–23, 2016.
- 554 **26** Hariharan Narayanan and Sanjoy Mitter. Sample complexity of testing the manifold hy-  
555 pothesis. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors,  
556 *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc.,  
557 2010. URL: [https://proceedings.neurips.cc/paper\\_files/paper/2010/file/](https://proceedings.neurips.cc/paper_files/paper/2010/file/8a1e808b55fde9455cb3d8857ed88389-Paper.pdf)  
558 [8a1e808b55fde9455cb3d8857ed88389-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2010/file/8a1e808b55fde9455cb3d8857ed88389-Paper.pdf).
- 559 **27** Brian D. Ripley. Locally finite random sets: Foundations for point process theory. *The Annals*  
560 *of Probability*, pages 983–994, 1976.
- 561 **28** Brian D Ripley. *Statistical inference for spatial processes*. Cambridge university press, 1988.
- 562 **29** Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster  
563 analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987. URL: [https://](https://www.sciencedirect.com/science/article/pii/0377042787901257)  
564 [www.sciencedirect.com/science/article/pii/0377042787901257](https://www.sciencedirect.com/science/article/pii/0377042787901257), doi:[https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- 566 **30** Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear  
567 embedding. *Science*, 290(5500):2323–2326, 2000. URL: [https://www.science.org/doi/](https://www.science.org/doi/abs/10.1126/science.290.5500.2323)  
568 [abs/10.1126/science.290.5500.2323](https://www.science.org/doi/abs/10.1126/science.290.5500.2323), arXiv:[https://www.science.org/doi/](https://www.science.org/doi/pdf/10.1126/science.290.5500.2323)  
569 [pdf/10.1126/science.290.5500.2323](https://www.science.org/doi/pdf/10.1126/science.290.5500.2323), doi:10.1126/science.290.5500.2323.
- 570 **31** Luis A. Santaló. *Integral Geometry and Geometric Probability Addison Wesley*. Cambridge  
571 University Press, 2nd edition, 2004.
- 572 **32** Jon Sporring, Rasmus Waagepetersen, and Stefan Sommer. Generalizations of ripley's k-  
573 function with application to space curves. In *Information Processing in Medical Imaging: 26th*  
574 *International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26*,  
575 pages 731–742. Springer, 2019.
- 576 **33** Kevin Streib and James W Davis. Using ripley's k-function to improve graph-based clustering  
577 techniques. In *CVPR 2011*, pages 2305–2312. IEEE, 2011.
- 578 **34** Wenwu Tang, Wenpeng Feng, and Meijuan Jia. Massively parallel spatial point pattern  
579 analysis: Ripley's k function accelerated using graphics processing units. *International Journal*  
580 *of Geographical Information Science*, 29(3):412–439, 2015.
- 581 **35** Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric  
582 framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.  
583 URL: <https://www.science.org/doi/abs/10.1126/science.290.5500.2319>,  
584 arXiv:<https://www.science.org/doi/pdf/10.1126/science.290.5500.2319>,  
585 doi:10.1126/science.290.5500.2319.
- 586 **36** Julius von Rohrscheidt and Bastian Rieck. Topological singularity detection at multiple scales,  
587 2023. arXiv:2210.00069.
- 588 **37** S Ward, HS Battey, and EAK Cohen. Nonparametric estimation of the intensity function of a  
589 spatial point process on a riemannian manifold. *Biometrika*, page asad012, 2023.

- 590 **38** Scott Ward. *Analysis of Spatial Point Patterns on Surfaces*. PhD thesis, Imperial College  
591 London, 2021.
- 592 **39** Nanning Zheng and Jianru Xue. *Manifold Learning*, pages 87–119. Springer London, London,  
593 2009. doi:10.1007/978-1-84882-312-9\_4.
- 594 **40** Nanning Zheng and Jianru Xue. *Statistical Learning and Pattern Analysis for Image and*  
595 *Video Processing*. Springer Science & Business Media, 2009.

## 596 **A** Ripley's $K$ -Function

597 Here, we provide a few concepts from spatial statistics, but defer the reader to [9], [8], [10] and [21]  
 598 for a more comprehensive overview of the relevant statistical definitions. Given  $n \in \mathbb{N}$  and a Borel  
 599 space  $S$  with measure  $\mu$ , a (binomial) point process  $\Phi$  of  $n$  points is a random variable valued in  
 600 functions  $f: S \rightarrow \mathbb{N}$ , where  $\|f\|_1 := \sum_{s \in S} f(s) = n$ . In other words, one way to think of  $\Phi$  is as a  
 601 random variable valued in finite multisets (i.e., points) of size  $n$  in  $S$ ; here, usually,  $S$  is a compact  
 602 subspace of  $\mathbb{R}^d$ , but, more generally, we allow  $S$  to be a Borel-measurable metric space [27]. In this  
 603 light, we write  $\Phi$  as the sum of  $n$  Dirac delta functions:  $\Phi = \sum \delta_{X_i}$ , where  $X_i$  is a random variable  
 604 valued in  $S$ .

605 One common tool to study spatial point patterns, such as a set of points sampled from a  
 606 manifold, is Ripley's  $K$ -function [12]. In particular, the  $K$ -function is a tool to assess complete  
 607 spatial randomness (homogeneity) of a point process. As we saw in Section 2.2, if the sample is  
 608 truly uniformly distributed, then the proportion of the sample points that land within a specific  
 609 region is proportional to the volume of that region.

612 ► **Definition 19 (Ripley's  $K$ -Functions).** Let  $n \in \mathbb{N}$ ,  $R \geq 0$ , and  $\Phi = \sum_{i=1}^n \delta_{X_i}$  be a binomial point  
 613 process with  $n$  points over a Lebesgue-measurable, compact set  $S \subset \mathbb{R}^d$ . Define the set  $S_R := \{x \in$   
 614  $S \mid d_2(x, \partial D) \leq R\}$ , where  $\partial D$  is the boundary of  $D$ . For each  $p \in S_R$ , Ripley's local  $K$ -function  
 615 for  $\Phi$  at  $p$  is the function  $\mathcal{K}_p: [0, R] \rightarrow \mathbb{R}$  defined by<sup>1</sup>

$$616 \quad \mathcal{K}_p(r) = \frac{\text{Vol}(S)}{n} \mathbb{E} \left[ |\mathbb{B}_d(p, r) \cap \{X_i\}_{i=1}^n| \right], \quad (18)$$

617 where  $\mathbb{E}[\cdot]$  denotes expectation.

618 The aggregated  $K$ -function is the function  $\mathcal{K}: [0, R] \rightarrow \mathbb{R}$  defined by

$$619 \quad \mathcal{K}(r) := \int_{p \in S} \mathcal{K}_p(r) d\mu(p) \quad (19)$$

620 In short, Ripley's  $K$ -function for a random sample is proportional to the expected number of  
 621 points that land in  $\mathbb{B}_d(p, r)$ . If the points are taken iid from the uniform distribution over  $D$ , then,  
 622 by Equation (2), we have  $\mathbb{E}[\mathcal{K}_p(r)] = \frac{\text{Vol}(S)}{n} \cdot \frac{n \text{Vol}(\mathbb{B}_d(p, r))}{\text{Vol}(S)} = \pi r^2$ .

623 Let  $X$  be a realization of  $\Phi$ . To estimate  $\mathcal{K}_p$  in practice, we use the empirical  $K$ -function, which  
 624 is the function  $\hat{\mathcal{K}}_p: [0, R] \rightarrow \mathbb{R}$  defined by:

$$625 \quad \hat{\mathcal{K}}_p(r) = \hat{\lambda}^{-1} \left[ \sum_{x \in X} I(x \in \mathbb{B}_2(p, r)) \right], \quad (20)$$

626 where  $I$  is the indicator function and  $\hat{\lambda}$  is an estimator of density per unit area ( $n/\text{Vol}(S)$ ). Often,  
 627 it is more practical to divide this quantity by  $\text{Vol}(S)$ , resulting in  $\hat{\mathcal{K}}_p(r) = \frac{1}{\text{Vol}(S)} \cdot \hat{\mathcal{K}}_p(r)$ , representing  
 628 the proportion of sample points that land in  $\mathbb{B}_2(p, r)$ . The empirical local  $K$ -function is the collection  
 629 of functions  $\{\hat{\mathcal{K}}_x\}_{x \in X}$ . Analogously, the aggregated empirical  $K$ -function  $\hat{\mathcal{K}}: [0, R] \rightarrow \mathbb{R}$  is defined  
 630 by:

$$631 \quad \hat{\mathcal{K}}(r) = \frac{1}{n} \sum_{x \in X} \hat{\mathcal{K}}_x(r).$$

632 Note that these definitions assume homogeneity of the point process, and hence  $\lambda$  is a constant.  
 633 If, however, this assumption is not met, we can adjust our estimate in Equation 20 by weighting  
 634 points in the point process by the estimated intensity measure at that point (see e.g. [2]):

$$635 \quad \hat{\mathcal{K}}_p(r) = \left[ \sum_{x \in X} \hat{\lambda}(x_i)^{-1} I(x \in \mathbb{B}_2(p, r)) \right], \quad (21)$$

---

610 <sup>1</sup> More generally, the  $K$ -function can be defined for all points in  $D$ ; however, care must be taken when  
 611 considering points near the boundary of  $D$ .

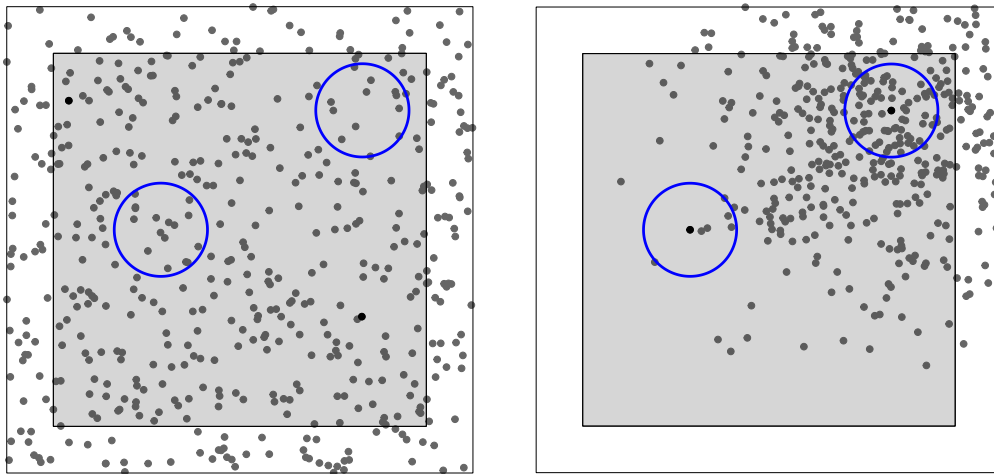


Figure 4 Two realizations of a Binomial point processes. Left shows a homogeneous point process and right shows an inhomogeneous point process.

Also, note that the definition we present for the  $K$ -function is biased-corrected using the boundary method (see [28] Ch. 3) by only considering the points  $S_R$  that is the set  $S$  eroded by distance  $R$ . See Figure 4 for a visualization of this for both a homogeneous and an inhomogeneous point process. Other bias correction methods exist to account for this bias. For example, one method weights points near the boundary by the proportion of  $\mathbb{B}_d(p, r)$  that lies within the manifold.

Further adjustments to the  $K$ -function are also needed when considering manifolds that are not  $\mathbb{R}^d$ . For example, Möller et. al develop the  $K$ -function over  $\mathbb{S}^d$  for the homogeneous and inhomogeneous cases in [25]. More recently, Ward et al. develop nonparametric estimation of intensity functions point processes over Riemannian manifolds [37]. However, functional summaries (e.g.  $K$ -function) were not considered in this work.

## B Definitions from Differential Geometry

### B.1 A Discussion of Intrinsic and Extrinsic Validation

In general, *model validation* takes a model and quantifies the correctness of its output. There are two main classes of validation algorithms: *extrinsic* and *intrinsic*.

In *extrinsic* validation, the labels in data are known. In this case, the validation process compares a model's output to the ground truth in order to quantify the model's correctness. Extrinsic validation is generally the easiest and strongest method for validation, one particularly ubiquitous variant is precision and recall which has been studied in the context for manifolds in [24].

In *intrinsic* validation, the ground truth correct output of a model is not known. Thus, the validation process must use properties of only the model's output in order to evaluate its correctness. A particularly notable example of intrinsic validation includes silhouette coefficients for clustering [29]. In practice, extrinsic validation typically accompanies supervised learning and intrinsic validation is often coupled with unsupervised learning. As the title alludes, this paper focuses on intrinsic validation.

### B.2 A Review of Basic Definitions in Differential Geometry

We present fundamental definitions from differential geometry in the following section. For a survey, we suggest [7] and [23].

We begin with a formal definition of Riemannian metrics, which are crucial in the definition of a Riemannian manifold

► **Definition 20 (Riemannian metric).** Let  $(\mathbb{X}, d)$  be a smooth manifold. We call  $d$  a Riemannian metric if  $d$  is a smooth covariant two-tensor field whose value  $d_x$  at each  $x \in \mathbb{X}$  is an inner product on the tangent space  $T_x \mathbb{X}$ .

On a manifold  $\mathbb{X}$ , we often study paths. A path between  $x, y \in \mathbb{X}$  is a continuous map  $\gamma : [0, 1] \rightarrow \mathbb{X}$ , where  $\gamma(0) = x$  and  $\gamma(1) = y$ . A path  $\gamma$  has length, which is defined as follows:

► **Definition 21 (Length).** Let  $(\mathbb{X}, d)$  be a topological space and let  $\gamma$  be a path in  $(\mathbb{X}, d)$ . Let  $\mathcal{P}$  be the set of all finite subsets  $P = \{t_i\}$  of  $[0, 1]$  such that  $0 = t_0 < t_1 < \dots < t_n = 1$ . The length  $L_d(\gamma)$  of  $\gamma$  is:

$$L_d(\gamma) := \sup_{P \in \mathcal{P}} \sum_{i=1}^n d(\gamma(t_i), \gamma(t_{i-1})).$$

The shortest paths between elements in  $\mathbb{X}$  are usually of particular interest, which are called geodesics.

► **Definition 22 (Geodesic).** Let  $(\mathbb{X}, d)$  be a topological space, let  $x, y \in \mathbb{X}$ , and let  $\Gamma$  be the set of all paths in  $(\mathbb{X}, d)$  from  $x$  to  $y$ . A geodesic  $\gamma^*$  from  $x$  to  $y$  is a path with shortest length:

$$\gamma^* = \inf_{\gamma \in \Gamma} L_d(\gamma)$$

Let  $U \subseteq \mathbb{X}$  be an open subset of  $\mathbb{X}$ , and let  $\phi : U \rightarrow \mathbb{R}^n$  be a coordinate chart of  $\mathbb{X}$  (that is,  $\phi$  is a homeomorphism). We say that  $\phi$  is smooth if for every  $u \in U$ , every component function of  $\phi$  has continuous partial derivatives of every order. Let  $U, V \subseteq \mathbb{X}$  be open subsets of  $\mathbb{X}$  and examine two coordinate charts  $\phi : U \rightarrow \mathbb{R}^n$  and  $\psi : V \rightarrow \mathbb{R}^n$ . We call  $\phi, \psi$  smoothly compatible if both  $\phi \circ \psi^{-1}$  and  $\psi \circ \phi^{-1}$  are smooth when defined on  $\phi(U \cap V)$  and  $\psi(U \cap V)$  respectively. An atlas of  $\mathbb{X}$  is a set of coordinate charts whose domains cover  $\mathbb{X}$ , and an atlas is a smooth atlas if any two charts in the atlas are smoothly compatible. A smooth structure is a smooth atlas not properly contained in another smooth atlas. Finally, a smooth manifold is a manifold  $\mathbb{X}$  equipped with a smooth structure.

Let  $x \in \mathbb{X}$  and let  $U$  be an open neighborhood of  $x$ . Let  $C^\infty(V, \mathbb{R})$  be the set of all smooth mappings  $f : U \rightarrow \mathbb{R}$ . Let  $f, g \in C^\infty(V, \mathbb{R})$ . A tangent vector  $T_x$  on  $\mathbb{X}$  at  $x$  is a linear map  $T_x : C^\infty(V, \mathbb{R}) \rightarrow \mathbb{R}$  satisfying the Leibniz law of products:

$$T_x(f \cdot g) = f(x)T_x(g) + g(x)T_x(f).$$

The tangent space on  $T_x \mathbb{X}$  on  $\mathbb{X}$  at  $x$  is the set of all tangent vectors on  $\mathbb{X}$  at  $x$ .

### B.3 A Review of Tensor Products and the Volume Form

We begin with a condensed review of tensors, and describe one of their most important constructions in differential geometry: the volume form. For a more detailed description, we refer readers to [23]. We begin with alternating tensors and wedge products, and then proceed to the volume form itself.

► **Definition 23 (Alternating Tensor).** Let  $V$  be a finite-dimensional vector space, and let  $F$  be a covariant  $k$ -tensor on  $V$ . We call  $F$  an alternating tensor if the interchanging of two arguments causes  $F$  to change sign:

$$F(v_1, \dots, v_i, \dots, v_j, \dots, v_k) = -F(v_1, \dots, v_j, \dots, v_i, \dots, v_k)$$

If  $F$  is an alternating  $k$ -tensor, then it follows that for an arbitrary permutation of arguments  $\sigma$ ,

$$F(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = \text{sgn } \sigma F(v_1, \dots, v_k).$$

Where  $\text{sgn } \sigma$  is the sign of  $\sigma$ , giving  $+1$  if  $\sigma$  is an even permutation and  $-1$  if  $\sigma$  is odd. Similarly, the alternation of  $F$  is given as follows:

706 ► **Definition 24** (Alternation). *Let  $F$  be a covariant  $k$ -tensor. Let  $S_k$  be the set of permutations of*  
 707  *$F$ . The alternation  $\text{Alt } F$  of  $F$  is given by the expression:*

$$708 \quad \text{Alt } F(v_1, \dots, v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} (\text{sgn } \sigma) F(v_{\sigma(1)}, \dots, v_{\sigma(k)})$$

709 It is easy to check that  $\text{Alt } F = F$  iff  $F$  is alternating. Using the Alt operation, we define the  
 710 wedge product of tensors:

711 ► **Definition 25** (Wedge Product). *Let  $V$  be a finite-dimensional vector space. Let  $F$  be an alternating*  
 712 *covariant  $k$ -tensor on  $V$ , and let  $G$  be an alternating covariant  $l$ -tensor on  $V$ . Then, their wedge*  
 713 *product is defined by:*

$$714 \quad F \wedge G = \frac{(k+l)!}{k!l!} \text{Alt}(F \otimes G)$$

715 Where  $F \otimes G$  is the usual tensor product.

716 Now, in order to understand the volume form rigorously, we present its definition below. Many  
 717 definitions give the same property as the one below, along with two equivalent expressions. For  
 718 simplicity we provide only the first property given in most definitions.

719 ► **Definition 26** (Volume Form). *Let  $(\mathbb{X}, g)$  be an oriented Riemannian manifold (with or without*  
 720 *boundary). Let  $(\varepsilon_1, \dots, \varepsilon_n)$  be a local oriented orthonormal coframe for  $T^*\mathbb{X}$ , the cotangent bundle of  $\mathbb{X}$ .*  
 721 *The volume form  $dV_g$ , or  $dV$  as a shorthand, is the unique  $n$ -form on  $\mathbb{X}$  such that  $dV_g = \varepsilon_1 \wedge \dots \wedge \varepsilon_n$ .*

722 By design, if we have a continuous function  $f : \mathbb{X} \rightarrow \mathbb{R}$  on a compact orientable Riemannian  
 723 manifold (with or without boundary), then  $f \cdot dV_g$  is a compactly supported  $n$ -form, and taking an  
 724 integral  $\int_{\mathbb{X}} f \cdot dV_g$  is well defined. Moreover, if  $\mathbb{X}$  is compact, the *volume* of  $\mathbb{X}$  is defined canonically:

$$725 \quad \text{Vol}(\mathbb{X}) = \int_{\mathbb{X}} dV_g.$$

## 726 B.4 The Gauss-Codazzi Equations for Hypersurfaces

727 As with basic tensor operations, this article assumes a some working knowledge of concepts of  
 728 curvature in differential geometry. For a detailed review of Riemannian, Ricci, Scalar, and Mean  
 729 curvatures, as well as the second fundamental form, we again refer readers to Lee [23].

730 In the results for hypersurfaces, we use the *second fundamental form* and the *length* of the second  
 731 fundamental form:

732 ► **Definition 27** (Second Fundamental Form). *Let  $M$  be an  $n$ -dimensional hypersurface in of a*  
 733 *Riemannian manifold  $M'$ , and let  $D$  and  $D'$  be their covariant derivatives. Then, for  $m \in M$  and*  
 734  *$u, v \in T_m M$  the second fundamental form  $h$  of  $M$  is given by:*

$$735 \quad h(u, v) := (D'_U V - D_u V)'_m$$

736 where  $U, V$  are vector fields tangent to  $M$  at any point of  $M$ , defined in a neighborhood of  $m'$  in  
 737  $M'$ , and with respective values  $u$  and  $v$  at  $m$ .

738 ► **Definition 28** (Length of Second Fundamental Form). *Let  $h$  be the second fundamental form at a*  
 739 *point  $p \in \mathbb{X}$ , a Riemannian manifold. The squared length  $\|h\|^2$  of  $h$  is given by:*

$$740 \quad \|h\|^2 = \sum_{i,j=1}^n h(E_i, E_j)^2$$

741 Where  $E_1, \dots, E_n$  is a local orthonormal frame of tangent vector fields over  $\mathbb{X}$ .

## 742 B.5 Hypersurfaces and their First Laplacian Eigenvalue

743 Naturally, after showing that the aggregated manifold density function can be scaled as a function  
744 of  $\lambda_1$  in high dimensions, one wonders how to actually compute  $\lambda_1$ . The following lays out the bulk  
745 of known results for  $\lambda_1$  with respect to a number of different manifolds.

746 ► **Theorem 29** (Conformal Clifford Torus). *If  $M$  is a conformal Clifford torus, Then*

$$747 \quad \lambda_1 * \text{vol}(M) \leq 4 * \pi^2$$

748 ► **Theorem 30** (Veronese Surface). *If  $M$  is a conformal Veronese surface, then we have*

$$749 \quad \lambda_1 * \text{vol}(M) \leq 12 * \pi,$$

750 *with equality holding if and only if  $M$  admits an order 1 isometric embedding.*

751 ► **Theorem 31** (Submanifolds of Hyperspheres). *Let  $M$  be an  $n$ -dimensional compact submanifold of  
752 a hypersphere  $S^m(r)$  of radius  $r$  in  $\mathbb{R}^{m+1}$ . Then  $\lambda_1 \leq n$ , with equality holding iff  $M$  is of order 1.*

753 ► **Theorem 32** (Projective Plane). *Let  $M$  be a compact,  $n$ -dimensional, minimal submanifold of  
754  $\mathbb{R}P^m$ , where  $\mathbb{R}P^m$  is of constant sectional curvature 1. Then it follows that  $\lambda_1 \leq 2(n+1)$ , with  
755 equality holding iff  $M$  is a totally geodesic  $\mathbb{R}P^n$  in  $\mathbb{R}P^m$ .*

756 ► **Theorem 33** (Complex Projective Plane). *Let  $M$  be an  $n$ -dimensional ( $n \geq 2$ ), compact, minimal  
757 submanifold of  $\mathbb{C}P^m$ , where  $\mathbb{C}P^m$  is of constant holomorphic sectional curvature 4. Then we have  
758  $\lambda_1 \leq 2(n+2)$   
759 *with equality holding iff  $n$  is even,  $M$  is a  $\mathbb{C}P^{\frac{n}{2}}$ , and  $M$  is a complex totally geodesic submanifold  
760 of  $\mathbb{C}P^m$ .**

761 ► **Theorem 34** (Quaternion Projective Plane). *Let  $M$  be a compact,  $n$ -dimensional ( $n \geq 4$ ), minimal  
762 submanifold of  $\mathbb{Q}P^m$ , where  $\mathbb{Q}P^m$  is of constant quaternion sectional curvature 4. Then we have  
763  $\lambda_1 \leq 2(n+4)$ , with equality holding iff  $n$  is a multiple of 4,  $M$  is  $\mathbb{Q}P^{\frac{n}{4}}$ , and  $M$  is embedded in  $\mathbb{Q}P^m$   
764 as a totally geodesic quaternionic submanifold.*

765 ► **Theorem 35** (Cayley Plane). *Let  $M$  be a compact,  $n$ -dimensional, minimal submanifold of the  
766 Cayley Plane  $OP^2$ , where  $OP^2$  is of maximal sectional curvature 4. Then we have  $\lambda_1 \leq 4n$ .*

767 ► **Theorem 36** (CR Submanifolds of  $\mathbb{C}P^n$ ). *Let  $M$  be a compact,  $n$ -dimensional, minimal, CR-  
768 submanifold of  $P^m$ . Then we have  $\lambda_1 \leq 2(n^2 + n + 2a)/n$ , where  $a$  is the complex dimension of the  
769 holomorphic distribution.*

770 ► **Theorem 37** (CR Submanifolds of  $\mathbb{Q}P^m$ ). *Let  $M$  be a compact,  $n$ -dimensional, minimal CR-  
771 submanifold of  $\mathbb{Q}P^m$ . Then we have  $\lambda_1 \leq 2(n^2 + n + 12a)/n$  where  $a$  is the quaternionic dimension  
772 of the quaternion distribution.*

## 773 C Experiments in Detail

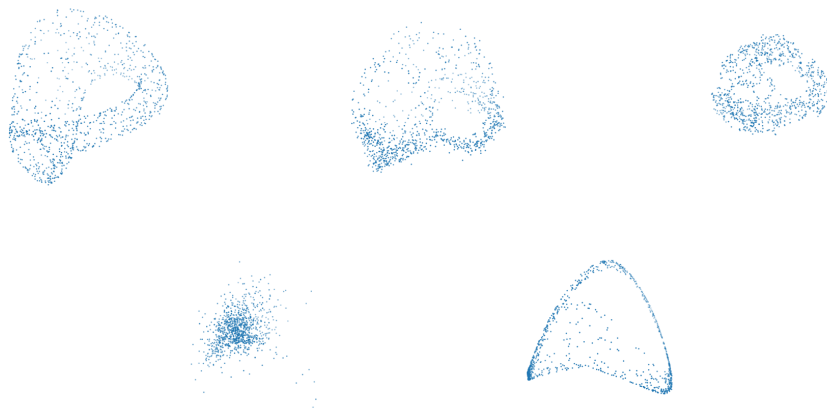
774 In the following section we outline additional experimental details left out for brevity in the main  
775 body of the paper. Our anonymized code is accessible at the following url for reproducibility:  
776 <https://anonymous.4open.science/r/intrinsic-manifold-validation/>

### 777 C.1 Additional Experimental Details for Surfaces

778 In the main paper, we presented results for surfaces delineating between uniform samples and  
779 stratifications. Here, we present more explicitly the details regarding how these samples were  
780 generated, alongside additional anecdotal experimental results not included in the main body.

781 We begin with a discussion of the manifold score on the Klein bottle, and implement the  
782 comparison of manifold learning algorithms shown in Figure ?? in Section 2. This anecdote provides  
783 a sense for both the intended positive behavior and the shortcomings of our use of the manifold density



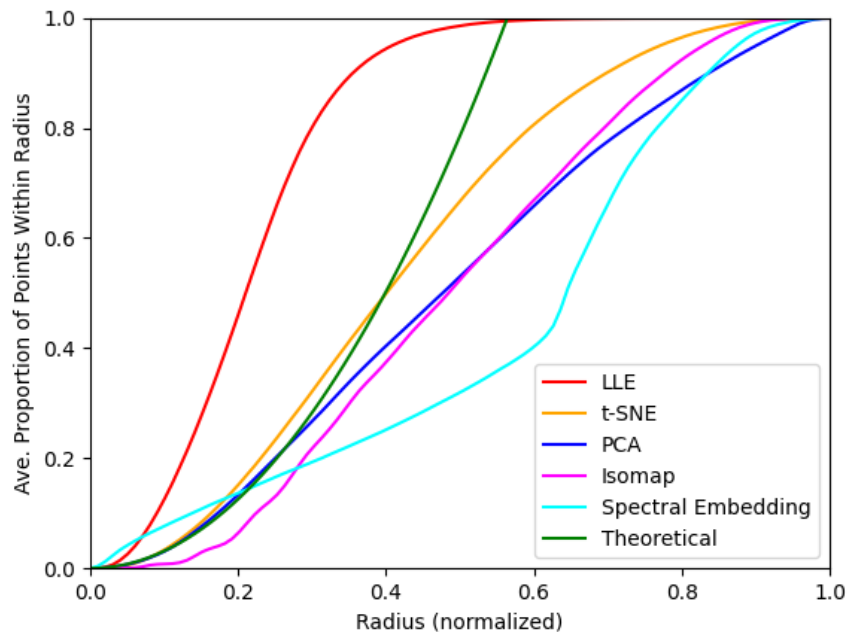


**Figure 5** Embeddings of the Klein bottle in  $\mathbb{R}^3$  after being projected from  $\mathbb{R}^{10}$  from left to right: PCA, ISOMAP, and t-SNE (top) locally linear embedding and spectral embedding (bottom).

function, which appear inherent to intrinsic validation in general. Namely, we lift a parameterization of the Klein bottle in  $\mathbb{R}^3$  to ten dimensions by assigning values from the interval  $[0, 2\pi]$  uniformly at random to each of the other seven coordinates. In total, this point cloud representation lying on an ambient Klein bottle in 3D has 1000 points. We then attempt to re-learn the 3D parameterization of the Klein bottle; computing embeddings in  $\mathbb{R}^3$  of the point cloud in  $\mathbb{R}^{10}$  using PCA, ISOMAP, t-SNE, LLE, and spectral embedding, each with  $n_{components} = 3$  and  $n_{neighbors} = 6$ . To the naked eye, PCA appears to perform the best, simply by dropping the other seven dimensions, and ISOMAP also appears to perform well, albeit a bit worse. In actuality, ISOMAP assigns much greater density to the handle region of the Klein bottle in this example, and achieves a somewhat unfavorable score. This provides an adversarial example where t-SNE outperforms ISOMAP, despite the visual preferability of the latter. The results when evaluating the manifold score on each algorithm are summarized in the following table, where we report the manifold score on the manifold density function restricted to radius  $r = 0.3$ :

**Table 5** Differing Manifold Scores from the Klein Bottle

| PCA    | ISOMAP | t-SNE  | LLE    | Spectral Embedding |
|--------|--------|--------|--------|--------------------|
| 0.9557 | 0.7179 | 0.9154 | 0.3988 | 0.6881             |



803 Upon visual inspection, the apparent best performing algorithm in this example (PCA) is indeed  
 804 detected as performing well, while the worse performing algorithms are generally detected as such.  
 805 However, the manifold score is not able to distinguish between ISOMAP and t-SNE in the example,  
 806 despite visually ISOMAP appearing to better uncover the manifold. This illuminates a shortcoming  
 807 of the manifold score; which is a consequence of working in the unsupervised setting. That is,  
 808 without knowledge of the ground truth manifold which we are attempting to uncover, there is no  
 809 way of distinguishing between a uniform sample of something more distantly resembling a desired  
 810 manifold, and a nonuniform sample closely resembling a desired manifold.

## 811 C.2 Experimental Details for Hypersurfaces

812 In what follows, we outline the resulting manifold density functions for hyperspheres, tabulated in  
 813 their corresponding dimension. We provide a more complete experiment here, including edge cases in  
 814 lower and higher dimensions than the cases included in experiments in the main body. Interestingly,  
 815 only upon reaching five dimensions does scaling the manifold density function for hypersurfaces  
 816 become beneficial, which we attribute to the fact that our scaling is indeed approximate; if possible  
 817 to work with two-manifolds the exact scaling afforded by the Gauss-Bonnet theorem is preferable.  
 818 In very high dimensions, as expected, it becomes increasingly subtle (but still possible outside of the  
 819 range of error) to distinguish between uniform samples nonuniform samples such as the stratification.  
 820 Once again, we employ a stratification of two  $d/2$  hyperspheres sampled on the surface of the  
 821  $d$ -dimensional hypersphere.

■ **Table 6** Results on Hyperspheres of Differing Dimension

| Dimension | Post Scaling           | Pre Scaling          | Stratification         |
|-----------|------------------------|----------------------|------------------------|
| 3         | $0.6801 \pm 0.0074$    | $0.7791 \pm 0.0093$  | $0.6251 \pm 0.0016$    |
| 4         | $0.9143 \pm 0.0043$    | $0.9785 \pm 0.0046$  | $0.8291 \pm 0.0899$    |
| 5         | $0.9823 \pm 0.0006$    | $0.9207 \pm 0.0049$  | $0.7936 \pm 0.0805$    |
| 20        | $0.9983 \pm 0.00002$   | $0.9732 \pm 0.0003$  | $0.9842 \pm 0.0015$    |
| 50        | $0.9997 \pm 0.000001$  | $0.9736 \pm 0.00012$ | $0.9982 \pm 0.000039$  |
| 100       | $0.9999 \pm 0.0000003$ | $0.9735 \pm 0.00013$ | $0.9995 \pm 0.0000085$ |