SIAM J. MATH. DATA SCI.
Vol. 7, No. 4, pp. 1882–1903

HADES: Fast Singularity Detection with Local Measure Comparison*

Uzu Lim†, Harald Oberhauser‡, and Vidit Nanda‡

Abstract. We introduce HADES, an unsupervised algorithm to detect singularities in data. This algorithm employs a kernel goodness-of-fit test, and as a consequence it is much faster and far more scalable than the existing topology-based alternatives. Using tools from differential geometry and optimal transport theory, we prove that HADES correctly detects singularities with high probability when the data sample lives on a transverse intersection of equidimensional manifolds. In computational experiments, HADES recovers singularities in synthetically generated data, branching points in road network data, intersection rings in molecular conformation space, and anomalies in image data.

Key words. manifold learning, machine learning, topological data analysis

MSC codes. 55N31, 32S50

DOI. 10.1137/24M1665763

1. Introduction.

The manifold hypothesis asserts that high-dimensional datasets encountered in practice tend to concentrate near smooth manifolds of low intrinsic dimension. It is often used to justify the effectiveness of machine learning algorithms in high-dimensional settings, since the curse of dimensionality can be circumvented if the data concentrates on a low-dimensional manifold. It is, however, evident that several low-dimensional (and hence, visualizable) datasets do not satisfy the manifold hypothesis. Instead, such data can have *singularities*—points at which the local geometry does not resemble n-dimensional Euclidean space for any n. Prime examples of singular loci of datasets include branching points in neurons and cosmic filaments. Furthermore, standard image datasets (such as MNIST and CIFAR-10) are known to have nonconstant intrinsic dimension [16], whereas a connected manifold must possess the same intrinsic dimension throughout.

Whenever such nonmanifold behavior within datasets is of interest, it becomes natural to wonder whether it can be accurately and automatically identified. Particularly in large, high-dimensional datasets where visual inspection is impossible, we seek tools to identify and locate singularities within datasets. Our focus here is on unsupervised singularity detection, where one has neither recourse to a plethora of training data nor the opportunity to regenerate samples along an unknown probability measure.

This paper. Here we propose HADES, a *hypothesis-testing algorithm for the detection and exploration of singularities.* The basic philosophy is rooted in two elementary observations. First, by definition, an n-manifold locally resembles a standard Euclidean n-dimensional disk;

*Received by the editors June 4, 2024; accepted for publication (in revised form) July 23, 2025; published electronically November 19, 2025.
https://doi.org/10.1137/24M1665763

†Queen Mary University of London, Mile End Road, London E1 4NS, UK (sung.h.lim@qmul.ac.uk).
‡Mathematical Institute, University of Oxford, Radcliffe Observatory, Andrew Wiles Building, Woodstock Road, Oxford OX2 6GG, UK (oberhauser@maths.ox.ac.uk, nanda@maths.ox.ac.uk).

and second, this resemblance can be precisely quantified by measuring the distance between (the local restriction of) an empirical measure and the uniform measure on the n-disk. HADES employs a goodness-of-fit test to measure this distance; we are therefore able to obtain a p-value for rejecting the null hypothesis that a given data point lies in the nonsingular locus of the underlying space.

Before proceeding to the details, we highlight three important features of the proposed algorithm below.

(1) **Efficiency:** HADES uses an explicit formula for kernel MMD (maximum mean discrepancy) to perform its goodness-of-fit test; this has a linear time complexity in the dimension of data, which forms a substantial improvement on the exponential complexity of the existing topological methods.
(2) **Correctness:** We show in Theorem 7 that HADES correctly identifies the singular set arising from the union of two transversely intersecting equidimensional submanifolds of Euclidean space. The proof uses tools from differential geometry, optimal transport theory, and concentration inequalities.
(3) **Validation:** In section 5, we run HADES on several synthetic and real datasets. On synthetic data, we observe that singularities are correctly detected (Figures 2, 3). And in the real datasets where we have no access to ground truth, the singularities identified by HADES exhibit interesting and observable anomalous behavior when compared to their nonsingular counterparts.

HADES is publicly released on GitHub, accessible by the following link: https://github.com/uzulim/hades.

Related work. Identifying nonmanifold points and studying their structure often goes under the name of *stratified learning,* which attempts to model data using stratified spaces, instead of manifolds. An early example of studying nonmanifold behavior in data is seen in [24], where a Poisson mixture model was used to measure locally evaluated intrinsic dimension that may vary across data. Follow-up works considered data sampled from a union of multiple manifolds. In multimanifold clustering, one starts with a data sampled from a union of intersecting manifolds and clusters data by separating them into the individual manifolds [44, 51, 43, 4, 2, 3]. Evidence for real-world data containing multiple manifolds of mixed dimension has been recently studied [16, 17, 34]. We remark that unions of manifolds only constitute a small subset of all stratified spaces. While our algorithm doesn't recover the structural information of manifolds, it detects more diverse types of singularities not present in a union of manifolds.

Stratification learning has received considerable attention from the topological data analysis community. The flagship tool here is *persistent homology,* which extracts topological information at multiple scales of data. In [8, 9, 10], persistent intersection homology was used to discover stratified structure of data. In [33, 13, 14], algorithms for recovering low-dimensional stratification structure and homotopy type of a stratified space have been studied. Discovering a stratification structure of a given simplicial complex [37] and a complex projective variety [25] has also been studied. In [45, 50], persistent homology was used to detect singularities in data, and their algorithms have the same objective as our algorithm. Compared to their algorithms, our algorithm has a significantly improved time complexity and theoretical foundation.

Dimension estimation and reduction are key steps in our algorithm, for which we simply apply principal component analysis (PCA) locally. Nevertheless there are many more advanced dimension estimation methods available, such as [29, 20, 19, 23, 27, 11, 50]. Dimension reduction methods in the literature include [7, 53, 46, 35, 47]. For a survey of dimension estimation and dimension reduction algorithms, see [18, 48].

2. Algorithm. The main idea of HADES is to perform the uniformity test at the neighborhood of each data point. The uniformity test measures resemblance of each neighborhood to a flat disk, and thus determines whether a data point is smooth or singular. Hyperparameters are required for the uniformity test, which can be chosen either manually or automatically. The output is then filtered and evaluated, and the best set of hyperparameters is chosen based on the output evaluation.

2.1. Uniformity test. The uniformity test works in two steps.

Step 1*: Dimensionality reduction.* We use PCA projection to reduce dimension of each *local* neighborhood, and this requires a threshold hyperparameter η. Suppose the points $\mathbf{z} \subset \mathbb{R}^D$ are plugged into the uniformity test. First its estimated dimension $\hat{d}$ is the number of principal components required to explain η of the total variance. Then the local neighborhood of data is projected to the $\hat{d}$ principal components, producing $\tilde{\mathbf{z}}$.

Step 2*: Goodness-of-fit test.* We use a kernel method to perform a goodness-of-fit test against the uniform distribution over a disk. We first compute the MMD and then compute the p-value associated to a null hypothesis. Given $\tilde{\mathbf{z}}$ obtained from projection, define the empirical measure $\hat{\mu}_z = n_z^{-1} \sum_{x \in \tilde{\mathbf{z}}} \delta_x$, where δ_x is the Dirac delta measure centered at x. Let $\mathfrak{u}_{\hat{d}}$ be the uniform measure over the unit $\hat{d}$-dimensional disk. Let Δ be the kernel MMD associated to a kernel κ. Let $S = \Delta(\mathbf{X}_n, \mathfrak{u}_d)$ be the null statistic, where $\mathbf{X}_n$ is an independent and identically distributed (i.i.d.) sample of size n drawn from $\mathfrak{u}_d$. We define the *singularity score* $\sigma(z)$ and the *singularity p-value* $p_\sigma(z)$ as follows:

$$\sigma(z) = \Delta(\hat{\mu}_z, \mathfrak{u}_{\hat{d}}),$$
$$p_\sigma(z) = \mathbb{P}\left[S \geq \sigma(z)\right].$$

The output of the uniformity test is the tuple $(\sigma(z), p_\sigma(z))$, and one of the main jobs done by HADES is to compute these for $z \in \{x_1, \ldots x_n\}$.

Intuitive explanation. Suppose a data distribution μ is a uniform distribution on $M \subseteq \mathbb{R}^D$. Given x, consider $\mu_{x,r}$ = the restriction of μ to the ball of radius r centered at x. There are two possible outcomes of the uniformity test:

Case 1*: M is smooth at x.* The support of $\mu_{x,r}$ is a slightly curved disk when r is sufficiently small. The dimensionality reduction flattens the slightly curved disk into $\mu_{x,r}^{\perp}$, which is supported on a flat disk and has a small nonuniformity. Then $\mu_{x,r}^{\perp} \approx \mathfrak{u}_{\hat{d}}$ and thus the goodness-of-fit test fails to reject the null hypothesis. The point x is declared as highly *unlikely* to be singular.

Case 2*: M is singular at x.* The support of $\mu_{x,r}$ does not resemble a flat disk no matter how small r is. The dimensionality reduction identifies a low-dimensional subspace containing the $\tilde{\mu}_{x,r}$, but $\tilde{\mu}_{x,r}$ only takes up a small portion of the Euclidean ball it spans. As such, the

goodness-of-fit test will reject the null hypothesis. The point x is declared as highly *likely* to be singular.

Therefore, a large singularity score σ (resp., small singularity p-value p_σ) corresponds to a high likelihood of the data point lying near the singular locus of the underlying stratifying space, and vice versa for a small singularity score (resp., large p_σ) and nonsingular (smooth) locus.

Technical details. The kernel we use is $\kappa(x,y) = (1 - \alpha \cdot \langle x, y\rangle)^{-1} = \sum_{k=0}^{\infty} \alpha^k \langle x, y\rangle^k$. To compute the MMD, we use the following formula, which is proven in Supplementary Material SM2.

Theorem 1. *Let $\hat{\mu}_n = \frac{1}{n}(\delta_{x_1} + \cdots + \delta_{x_n})$ be a discrete (nonrandom) measure and let $\mathfrak{u}_d$ be the uniform distribution over the unit d-dimensional disk in $\mathbb{R}^d$. Let κ be a kernel given by $\kappa(x,y) = \sum_{k=0}^{\infty} a_k \langle x, y\rangle^k$, and let Δ be the MMD associated to κ. Then we have*

$$\Delta^2(\hat{\mu}_n, \mathfrak{u}_d) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \kappa(x_i, x_j) + \sum_{k=0}^{\infty} a_{2k} \beta_{d,k} \left(\frac{d}{d+2k} - \frac{2}{n} \sum_{i=1}^{n} \|x_i\|^{2k} \right),$$

where β are the constants $\beta_{d,k} = \pi^{-1/2} \Gamma(\frac{d}{2} + 1) \Gamma(k + \frac{1}{2}) \Gamma(k + \frac{d}{2} + 1)^{-1}$.

To evaluate the p-value arising from the MMD, we use its asymptotic distribution for large sample size n. The MMD is a V-statistic, for which asymptotic convergence under scaling by sample size holds true (section 5 in [42]).

Theorem 2. *Let μ be a Borel measure on $\mathcal{X} \subseteq \mathbb{R}^d$ and let $\hat{\mu}_n$ be the empirical measure of size n drawn from μ. Let $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a function satisfying $\kappa(x,y) = \kappa(y,x)$, and let Δ be the MMD associated to κ. Then there is a convergence in distribution as $n \to \infty$:*

$$n \cdot \Delta^2(\hat{\mu}_n, \mu) \longrightarrow c_\kappa + \sum_{i=1}^{\infty} \lambda_i (Z_i^2 - 1).$$

Here Z_k are independent standard normals and $c_K := \mathbb{E}[\kappa(X,X)] - \mathbb{E}[\kappa(X,Y)]$. λ_k are eigenvalues of the integral operator $L[\phi] = \int \tilde{\kappa}(x, -)\phi(x)\, \mathrm{d}\, \mu(x)$, where $\tilde{\kappa}(x,y) = \kappa(x,y) - \mathbb{E}[\kappa(X,y)] - \mathbb{E}[\kappa(x,Y)] + \mathbb{E}[\kappa(X,Y)]$.

We obtain the asymptotic distributions by Monte Carlo, i.e., by directly sampling the null statistics $n \cdot \Delta^2(\hat{\mu}_n, \mu)$ and using this to construct an empirical cumulative distribution function. To compute p-values for events that lie far outside the Monte Carlo simulation, we use exponential decay to estimate the p-values (see [39]).

Remark. There are many alternative choices for dimension estimation, dimensionality reduction, and the statistical distance used to perform the goodness-of-fit test. These parts can be swapped out in a modular way, and the algorithm can be modified to match the user's needs. For example, a more sophisticated application might use a fractal dimension estimator, use UMAP [35] to perform local dimensionality reduction, and use statistical distances such as the Wasserstein distance. In practice, we found that the kernel MMD is more sensitive to detecting nonuniformity compared to the Wasserstein distance or its regularized Sinkhorn approximation [21]. Future improvements of the algorithm could be made by fine-tuning

each of these steps while using the same conceptual framework. We also remark that the uniformity test can be trivially parallelized since all computations are local. Thus running HADES parallel on multiple cores directly speeds it up.

2.2. Filtering and evaluation. We now explain how to filter the singularity p-values into a binary label and evaluate the quality of the labeling. The labeling quality is ascribed to the hyperparameter set used to run the uniformity tests, and this gives us a way to choose the best set of hyperparameters.

The singularity p-values are filtered by applying a knee detection algorithm to the empirical probability density function of $\log(1/p_\sigma)$. Here, we apply the logarithm to separate very small p-values. We use Gaussian kernel density estimation to produce an empirical probability density function. We use the Kneed [40] algorithm to detect the knee of the probability density, and declare all points appearing after the knee to be singular points. The knee detection effectively identifies smooth points since their singularity p-values are relatively large, so that their distribution of $\log(1/p_\sigma)$ forms a concentrated mass near 0.

The quality of the binary label produced by the filtering step is evaluated using a metric we named *dispersion score*. The dispersion score is defined purely using data points and any binary label on them. The dispersion score is defined using *purity score* and *separation score*.

Definition 3. *Let* $\mathbf{x} = (x_1, \dots x_n), \mathbf{y} = (y_1, \dots y_n)$ *be points and their binary labels,* $x_i \in \mathbb{R}^D, y_i \in \{0,1\}$. *For each* $i = 1, \dots n$, *let* $\mathcal{N}(i) \subseteq \{1, \dots n\}$ *be a set satisfying* $i \in \mathcal{N}(i)$. *Define a partition* $\mathcal{I}_0 \sqcup \mathcal{I}_1 = \{1, \dots n\}$, *where* $\mathcal{I}_a = \{i | y_i = a\}$.

The purity score u_i *is the proportion of indices* $j \in \mathcal{N}(i)$ *with* $y_j = 1$, *and the* separation score s_i *is defined as an AUC (area-under-curve) score:*

$$u_i(\mathbf{y}, \mathcal{N}) = \frac{\#(\mathcal{N}(i) \cap \mathcal{I}_1)}{\#\mathcal{N}(i)}, \quad s_i(\mathbf{x}, \mathbf{y}, \mathcal{N}) = \mathrm{AUC}\left\{ (t_{ij}, y_j) \middle| j \in \mathcal{N}(i) \right\},$$

where $t_{ij} = \langle x_j - x_i, \frac{\tilde{x}_i}{\|\tilde{x}_i\|} \rangle$ *and* $\tilde{x}_i = \sum_{j \in \mathcal{N}(i) \cap \mathcal{I}_1} (x_j - x_i)$.

The dispersion score *is defined as*

$$\mathfrak{D}(\mathbf{x}, \mathbf{y}, \mathcal{N}) = \alpha \cdot \mathcal{D}_1(u_G) + \sum_{i \in \mathcal{I}_1} \mathcal{D}_2(q_i), \quad \textit{where} \quad q_i = 1 - \frac{1}{2}(s_i + u_i),$$

where $u_G = \#(\mathcal{I}_1)/n$ *is the global purity score,* α *is a regularization constant, and* $\mathcal{D}_1, \mathcal{D}_2$ *are* damping functions, *which are bijections* $\mathcal{D}_i : [0,1] \to [0,1]$ *satisfying* $\mathcal{D}_i(x) \leq x$.[1]

Separation score quantifies how well the binary labels are cleanly separated along locally defined axes of direction, $\tilde{x}_i$. Indeed $\tilde{x}_i$ is the sum of displacements $x_j - x_i$ for which $y_j = 1$, and t_{ij} is the projected length of the displacement $x_j - x_i$ onto $\tilde{x}_i$. Thus, s_i measures how well the numbers t_{ij} can classify the binary labels y_j when $j \in \mathcal{N}(i)$.

Dispersion score detects points x_i for which both s_i and u_i are *simultaneously* small, while also penalizing the degenerate case $u_G \approx 1$, when almost all points satisfy $y_i = 1$. The points x_i satisfying $i \in \mathcal{I}_1$ and $s_i + u_i \approx 0$ are far away from other indices $j \in \mathcal{I}_1$ and have poorly

[1] In the code, the default choice of the damping functions is given by $\mathcal{D}_1 = F_{0,2}$ and $\mathcal{D}_2 = F_{0.5,5}$, where $F_{a,b}(t) = (\frac{t-a}{1-a})^b$.

defined local boundaries for separating the label 1 from the label 0. By using the damping functions $\mathcal{D}_1, \mathcal{D}_2$, we ensure that only the points x_i for which q_i is sufficiently large make a meaningful contribution to $\mathfrak{D}$, and also only the degenerate case for which $u_{\mathrm{G}} \approx 1$ makes a meaningful contribution to $\mathfrak{D}$.

Remark. HADES is an *unsupervised learning* algorithm, for which there is no training dataset whose loss value can be minimized over many sets of hyperparameters. Instead, like clustering algorithms, the best set of hyperparameters is chosen by optimizing a qualitatively defined criterion—the dispersion score. The dispersion score differs from the classical clustering quality measures that reward concentration around centroids of clusters. The difference is that it aggregates *local* clustering information gathered from the data points, and thus the dispersion score can still be made small for complex shapes formed by the binary labels. This is adequate since the set of singularities of a stratified space have no reason to be concentrated around their centroid. (See Figure 2; the singular points marked in blue are not point-like clusters sought by the classical clustering quality measures.)

2.3. Hyperparameter selection. HADES uses the following three hyperparameters:

(1) **Local radius r.** Used to isolate neighborhoods.
(2) **PCA threshold η.** Used for dimension estimation.
(3) **Kernel parameter α.** Used in MMD of the uniformity test.

The hyperparameters have the following effects. The radius r and threshold η both need to be at the right range to ignore noise and curvature. (For a thorough mathematical analysis, see [31].) The effect of the kernel parameter $\alpha \in (0, 1)$ is less obvious. Choosing a different kernel parameter causes a different local singular geometry to be penalized. However, we found that the correctness of the output has a low sensitivity to the kernel parameter.

We explain how the sets of hyperparameters to run are automatically chosen by HADES. As explained before, the basic idea is to optimize the dispersion score over multiple sets of hyperparameters. These can be either supplied manually by the user or chosen automatically by HADES. In the automatic hyperparameter selection, we use a grid $r \in [r_{\min}, r_{\max}]$, $\eta \in [\eta_{\min}, \eta_{\max}]$, $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, where we use default values of $\eta \in [0.7, 0.9]$ and $\alpha \in [0.3, 0.7]$ for the PCA threshold and the kernel parameter.

Meanwhile, the range of radius hyperparameter r is chosen using a local scale detection algorithm. The idea here is to slowly enlarge a local neighborhood until the intrinsic dimension of the neighborhood stabilizes. This process is done for multiple data points, and curves of intrinsic dimension estimates are averaged over them. The Kneed algorithm [40] is used to detect the threshold at which intrinsic dimension stabilizes, by going backward from the dimension estimate of the largest neighborhood and shrinking them, and detecting a knee of the curve. The standard intrinsic dimension estimator by Levina and Bickel [29] was used to calculate the intrinsic dimensions of the expanding neighborhoods. After obtaining the knee $\tilde{r}$, we use the range $r \in [1.5\tilde{r}, 5.0\tilde{r}]$.

When the optimal set of hyperparameters is found at the boundary of the grid search, HADES expands the search range toward that direction of the hyperparameter grid. For example, consider the grid search on $(r, \eta, \alpha) \in [0.1, 0.2] \times [0.7, 0.9] \times [0.3, 0.7]$, and suppose the dispersion score was minimized for $(r, \eta, \alpha) = (0.2, 0.8, 0.5)$. Since the optimal choice of r is found at the maximum of the range $[0.1, 0.2]$, HADES will do another grid search on the

range $[0.2, 0.3] \times [0.7, 0.9] \times [0.3, 0.7]$ afterward.[2] This process is repeated until a prespecified end of search bounds is reached.

2.4. Hypothesis tests. HADES performs two types of hypothesis test: local and global. The local hypothesis test is performed at each data point to detect the location of singularities. The global hypothesis test tests the manifold hypothesis, to determine whether the entire dataset was sampled from a manifold or not.

Local hypothesis test. HADES performs the uniformity test at each data point (see section 2.1). The uniformity test is the following (nonparametric) null hypothesis significance test. Denote by $B(x,r) \subset \mathbb{R}^D$ the unit ball of radius r, centered at x. Let $\mathbf{x}_n \subset B(x,r)$ be a dataset. Let Θ be the set of all Borel probability measures on $B(x,r)$. Let Θ_0 be the set of uniform measures over a (zero-curvature) d-dimensional disk of radius r containing x. Let $\Theta_1 = \Theta \backslash \Theta_0$ be the set difference. Then the null hypothesis H_0 and the alternative hypothesis H_1 are defined as follows:

$$H_0 : \mathbf{x}_n \text{ is an i.i.d. sample drawn from some } \mu \in \Theta_0,$$
$$H_1 : \mathbf{x}_n \text{ is an i.i.d. sample drawn from some } \mu \in \Theta_1.$$

The following test statistic T is used to test the null hypothesis (see section 4 for the mathematical definition of the singularity score σ, and see Theorem 2 for the reason for scaling by $\#\mathbf{y}$):

$$T := \mathcal{T}(\mathbf{x}_n \cap B(x,r)),$$
$$\text{where } \mathcal{T}(\mathbf{y}) = (\#\mathbf{y}) \cdot \sigma_\eta^2(\delta_\mathbf{y}).$$

If $T > t_\alpha$, we reject the null hypothesis H_0 in favor of H_1 at significance level α, and if $T \leq t_\alpha$, we fail to reject H_0. Here t_α is the $(1-\alpha)$-quantile of the null distribution of T, which is computationally simulated in the current implementation of HADES.

While Θ_0 only consists of uniform measures over a d-dimensional disk, a future work could consider a nonuniform distribution over a submanifold of nonzero-curvature. Let $c_1, c_2 > 0$. Define Θ_0' to be the set of Borel probability measures μ over a smoothly embedded disk $M \hookrightarrow B(x,r)$, such that the density of μ has Lipschitz constant $\leq c_1$ and the curvature of geodesics over M is $\leq c_2$. Let $\Theta_1' = \Theta \backslash \Theta_0'$. Then an analogous hypothesis test can be formulated by replacing (Θ_0, Θ_1) with (Θ_0', Θ_1'). To quantify the effect of curvature and nonuniformity, one could use the main result from section 4 of [31], where a Wasserstein bound is derived.

A caveat of this local hypothesis test is a false discovery of singularities. If we declare a data point to be singular whenever its associated p-value is below p_0, then we expect that approximately np_0 data points out of the total n data points tested will be false positives declared as singular. This is because p-value is uniformly distributed under the null hypothesis. However, the false discovery is accounted for by the global hypothesis test, as explained below.

[2] In the implementation, we actually use a slightly more sophisticated method for expanding radius range. Observe that the volume of a d-dimensional ball with radius r is $\omega_d r^d$. We expand the radius parameter range such that this estimated local volume expands linearly.

Global hypothesis test. HADES contains a partial implementation of testing the *manifold hypothesis* on the whole dataset. This determines whether the whole dataset (as opposed to individual data points) was sampled from a smooth manifold or not. While more work is needed for a full implementation of the hypothesis test, experimental evidence in Figure 4 shows much promise. Let $\mathbf{x}_n \subset \mathbb{R}^D$ be a given dataset. Let Θ be the set of all Borel probability measures on $\mathbb{R}^D$. Let Θ_0^{G} be the set of all uniform measures over a compact smooth manifold $M \subset \mathbb{R}^D$. Let $\Theta_1^{\mathrm{G}} = \Theta \backslash \Theta_0^{\mathrm{G}}$. The null hypothesis H_0 and the alternative hypothesis H_1 are defined as follows:

$$H_0 : \mathbf{x}_n \text{ is an i.i.d. sample drawn from some } \mu \in \Theta_0^{\mathrm{G}},$$
$$H_1 : \mathbf{x}_n \text{ is an i.i.d. sample drawn from some } \mu \in \Theta_1^{\mathrm{G}}.$$

We propose using the test statistic T^{G} to test the null hypothesis, which is defined as

$$T^{\mathrm{G}} = S_{\mathbf{q}}(\{p_1, \dots p_n\}),$$
$$\text{where } p_i = p\text{-value obtained from the local hypothesis test at } x_i \in \mathbf{x}_n,$$
$$S_{\mathbf{q}}(\{p_1, \dots p_n\}) = \max\{(nq)^{-1} \#\{p_i \le q\} | q \in \mathbf{q}\}$$

It is difficult to access the null distribution of T^{G}, because then we need a computational simulation of all smoothly embedded compact manifolds. However, Figure 4 shows the boxplot of T^{G} for 160 datasets following the null hypothesis (blue) and 120 datasets not following the null hypothesis (orange). While this is insufficient for a full statistical analysis, we attain an AUC score ≈ 1 and the boxplot shows a clear separation between the two types of datasets. A comprehensive simulation of the null distribution of T^{G} is left for future work.

We give a heuristic reasoning for why T^{G} is able to distinguish a manifold from a stratified space. First, note that the number $(nq)^{-1}\#\{p_i \le q\}$ represents the ratio of how many p-values satisfy $p_i < q$ compared to nq. In practice we choose shrinking values $0.5 = q_1 > \cdots > q_m = 0.01$ where $\mathbf{q} = \{q_1, \dots q_m\}$. Since $S_{\mathbf{q}}$ takes the maximum of these ratios, it detects a concentration of small p-values. When the data is sampled from a manifold, $S_{\mathbf{q}} \approx 1$ because a density fluctuation happens rarely for a data sampled from a manifold, causing p-values to be roughly uniformly distributed over $(0, 1)$. When the data is sampled from a stratified space, the geometric singularity causes a clump of data points to return very small p-values, and thus $S_{\mathbf{q}} \gg 1$. Thus when $S_{\mathbf{q}} \approx 1$, we interpret the small p-values to be false positives detected from the local hypothesis tests, and declare the dataset to be sampled from a manifold. Conversely, when $S_{\mathbf{q}} \gg 1$, we declare the dataset to be sampled from a stratified space.

3. Comparison with topological methods. We demonstrate significantly improved time complexity and statistical foundation of HADES in the singularity detection task, compared to the previous topological methods. Topological methods of singularity detection are based on persistent homology, a prominent tool from *topological data analysis* [50, 45, 10, 8, 9, 33, 13, 14]. Persistent homology computes topological features at varying scales of data, and the main idea behind topological methods for singularity detection is to compute persistent homology on local neighborhoods of data. In particular, the recent algorithms in [50, 45] use the fact that a small annular neighborhood of a point on a manifold has the topology of a sphere, whose topology is well-understood.

Time complexity. A major advantage of HADES over singularity detection algorithms based on persistent homology is that HADES scales much better to high-dimensional data. We first have the following.

Theorem 4. *The time complexity of the uniformity test on n points in $\mathbb{R}^D$ is $O(n^2D)$.*

Proof. The dimensionality reduction step is an application of PCA. SVD is performed on a rectangular matrix of dimension $k \times D$, which has the time complexity of $O(k^2D)$ [49]. With the estimated dimension of $\hat{d}$, a matrix multiplication between a rectangular matrix of size $(k \times \hat{d})$ and a diagonal matrix of size $(\hat{d} \times \hat{d})$ is performed, for which the time complexity is $O(k\hat{d})$, which is less than $O(k^2D)$. This step thus amounts to the time complexity of

$$T_1 = O(k^2D).$$

The goodness-of-fit step computes the MMD of a $\hat{d}$-dimensional point set of size k. Following the expression computed in Theorem 1, the time complexity for this step is

$$T_2 = O(k^2d + k + kd) = O(k^2d). \qquad \blacksquare$$

In comparison, the time complexity of persistent homology increases exponentially in the intrinsic dimension of data. The computational complexity of Ripser [6], a highly optimized Python package for computing persistent homology, is $O(s^3)$, where s is the number of simplices constructed. However, a dataset of k points has a total of $s = \binom{k}{d+1} = O(k^{d+1})$ simplices of dimension d. A small annular local neighborhood of a d-dimensional manifold is topologically a $(d-1)$-sphere and requires computationally constructing d-simplices. Therefore, the computational complexity of the $(d-1)$th persistent homology group is $O(k^{3d+3})$. Persistent homology computation corresponds to the uniformity test, where in our algorithm we instead use PCA and kernel MMD. Using the computational complexity of the uniformity test given above, we have the following comparison of computational complexity incurred by local shape analysis:

$$\begin{aligned} \text{Uniformity test:} \quad & O(k^2D), \\ \text{Persistent homology:} \quad & O(k^{3d+3}). \end{aligned}$$

Thus we observe an exponential dependence of persistent homology computation on the intrinsic dimension d of data, whereas the uniformity test has a linear dependence on the ambient dimension D. When $D \gg d$, preprocessing data by dimensionality reduction allows us to circumvent the dependence of HADES on the ambient dimension D.

In Figure 1, we compare computation times of HADES (blue curve) and Ripser (orange curve). Each of the four plots shows computation times for a fixed dimension,[3] but varying sample size. We observe that while HADES shows poorer performance than Ripser in low-dimensional data, the situation is quickly reversed in high-dimensional data.

[3]For d-dimensional data, we use samples of the unit d-dimensional ball for HADES and samples of the unit $(d-1)$-dimensional sphere for Ripser.

Diminishing persistence. We observe from computational experiments that the topological signature of a high-dimensional sphere has a small persistence. This appears to present problems in applying the standard practice in topological data analysis, which declares a point on the persistent diagram as a genuine signal only if the point has a high persistence. In the case of the d-dimensional sphere, one seeks one highly persistent point on the d-dimensional persistence diagram, since the d-dimensional sphere has a 1-dimensional dth homology group, and all other kth homology groups are zero for $k > 0$.

As such we define the STS at (d, n) to be the most persistent point of $\mathrm{PD}_d(\mathbf{X}_n)$, where $\mathrm{PD}_d(\mathbf{X}_n)$ is the dth persistence diagram of the Rips filtration on $\mathbf{X}_n$, and $\mathbf{X}_n$ is an i.i.d. sample of size n from the d-dimensional sphere. Figure 1 tabulates birth times and persistence (y-axis) of the STS at (d, n) for varying sample size n (x-axis) and the dimension d (different curves, color-coded). The STS is significant because it is the main signal sought by the standard practice of topological data analysis.

Figure 1 indicates that the STS of a high-dimensional sphere has a small persistence and a large birth time. The small persistence tells us that STS becomes increasingly unreliable in

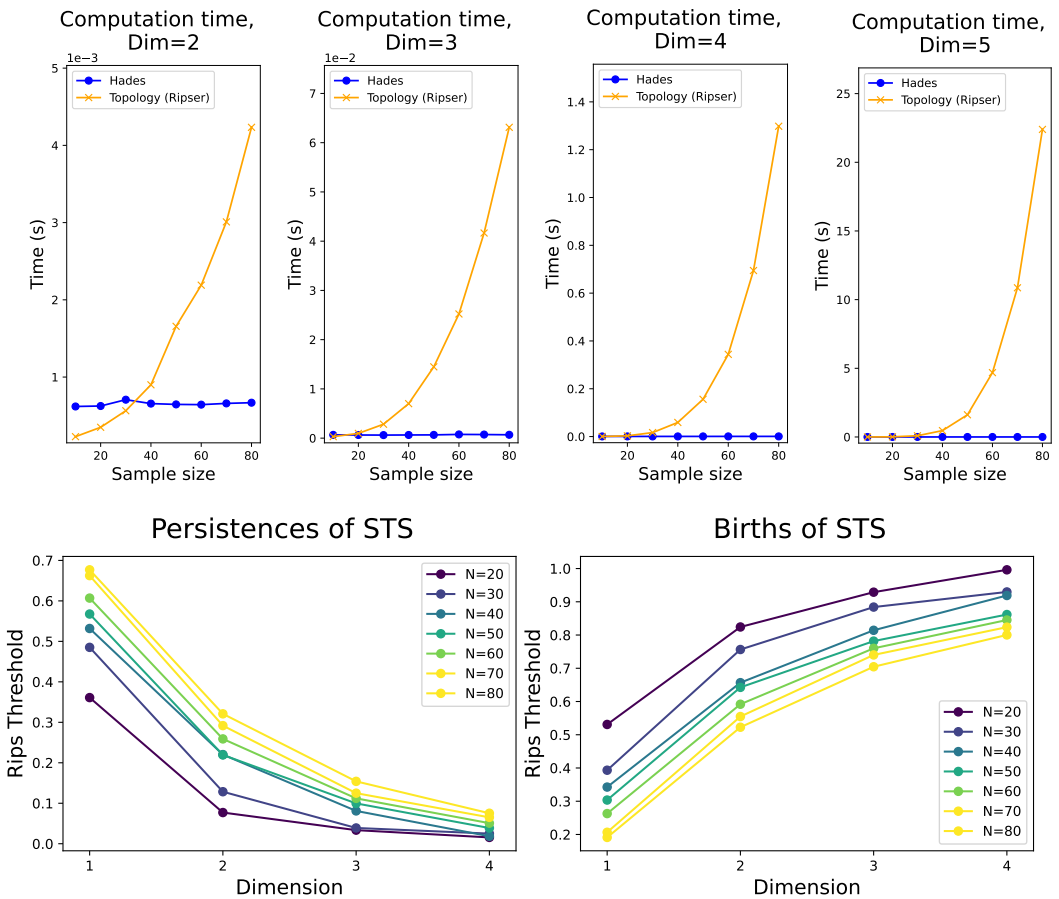

Figure 1. *Top row: Comparison of computation time of local shape analysis in* HADES *(blue) versus Ripser (orange), a highly optimized library for computing persistent homology. Bottom row: Persistence of the significant topological signature (STS) in high-dimensional spheres decays significantly across dimensions.*

high dimensions, due to it resembling "topological noise." This appears to defy the current paradigm of topological data analysis where highly persistent topological features are to be seen as genuine signal and other topological features are to be seen as noise. The large birth time tells us that one cannot use a small connectivity threshold to detect STS, and therefore that it is difficult to reduce the number of high-dimensional simplices appearing in the full filtration of a point cloud.

This situation may be improved by using low-dimensional topological signal of high-dimensional spheres, which runs on smaller time complexity. In fact, even the 1-dimensional sphere (circle) exhibits systematic high-dimensional topological signals in large connectivity thresholds [30, 1], and high-dimensional spheres exhibit systematic low-dimensional topological signals.

4. Theoretical guarantee.

In this section, we will define a mathematically precise version of the singularity detection algorithm, and state a theorem guaranteeing that the algorithm detects singularities correctly. In the following, we fix the ambient dimension D and threshold parameter $\eta \in (0,1)$. We first define the PCA dimension estimator and projector precisely.

Definition 5. *Given μ, a Borel probability measure on $\mathbb{R}^D$, its* estimated dimension $\hat{d}(\mu)$, *and* linear regression $\mathcal{L}(\mu)$ *are defined as*

$$\hat{d}(\mu) = \min\left\{k \,\middle|\, \frac{\lambda_{k+1}+\cdots+\lambda_D}{\lambda_1+\cdots+\lambda_D} \le \eta\right\},$$
$$\mathcal{L}(\mu) = \operatorname{span}\left(\mathcal{E}(\mu,\lambda_1),\ldots,\mathcal{E}(\mu,\lambda_{\hat{d}(\mu)})\right),$$

where $(\lambda_1,\ldots\lambda_D)$ are eigenvalues of $\Sigma[\mu]$, $\mathcal{E}(\mu,\lambda)$ is the λ-eigenspace of $\Sigma[\mu]$, and $\Sigma[\mu]$ is the covariance matrix of μ. The parameter η is implicit from the notation $\hat{d}(\mu), \mathcal{L}(\mu)$.

Using $\hat{d}$ and $\mathcal{L}$, we define the mathematically precise version of the *singularity score.* In the following let $\Delta(\mu,\nu)$ denote the kernel MMD associated to the Gaussian kernel $\kappa(x,y) = \exp(-\gamma\cdot\|x-y\|^2)$ for some fixed $\gamma > 0$. Also denote by $\mathfrak{u}_d$ the uniform measure over the unit d-dimensional disk centered at the origin. We first define the *abstract singularity score,* and use this for empirical measures to define the *empirical singularity score.*

Definition 6. *The* abstract singularity score *is defined as*

$$\sigma(\mu) = \Delta(\mu_\perp, \mathfrak{u}_{\hat{d}}),$$

where $\hat{d} = \hat{d}(\mu)$ and $\mu_\perp = \Pi(\mu, \mathcal{L}\mu)$ is the pushforward of μ along the projection to $\mathcal{L}\mu$.

Let $\mathbf{x} = \{x_1,\ldots x_n\} \subset \mathbb{R}^D$ and let $r > 0$. Denote $\mathbf{x}(z) = \mathbf{x} \cap \mathcal{B}(z,r)\backslash\{z\}$, where $\mathcal{B}(z,r) \subseteq \mathbb{R}^D$ is the open ball of radius r, centered at z[4]*. The* local empirical measure *of $\mathbf{x}$ at z is*

$$\hat{\mu}(z) = g_{z,r}\left(\frac{1}{\#\mathbf{x}(z)}\sum_{y\in\mathbf{x}(z)}\delta_y\right),$$

[4]The point z is excluded for a technical reason concerning Wasserstein concentration inequalities, although the proofs can be modified to be included if necessary.

where $g_{z,r}(\nu)$ *is the pushforward of a measure* ν *by the affine map* $x \mapsto r^{-1}(x-z)$. *The* i*th* empirical singularity score *of* $\mathbf{x}$ *is defined as*

$$\hat{\sigma}_i(\mathbf{x}, r, \eta) = \sigma(\hat{\mu}(x_i)).$$

Note that all of $\hat{d}, \mathcal{L}, \sigma$ depend on the choice of dimension estimation threshold η. We now state the setup and the main theorem.

Setup. Let $M = M_1 \cup M_2$, where $M_1, M_2 \subseteq \mathbb{R}^D$ are smooth compact d-dimensional manifolds embedded in $\mathbb{R}^D$. Suppose there exist $d_0, \phi > 0$ such that the following holds for every $x \in M_1 \cap M_2$: the tangent spaces $T_x M_1$ and $T_x M_2$ intersect at a d_0-dimensional subspace, and all principal angles of the pair are $\geq \phi$. Let μ be the uniform measure over M, and let $\mathbf{X}_n = (X_1, \dots X_n)$ be an i.i.d. sample of size n drawn from μ.

Theorem 7 (theoretical guarantee). *There exist constants* $\xi, \eta_-, \eta_+, c_A, c_B, r_0 > 0$ *depending only on* M *such that the following holds. Given* $\eta \in [\eta_-, \eta_+]$, $r \leq r_0$, *and* $q \in (0,1)$, *the following implications both hold for all* i *with probability at least* q, *when* n *is large enough:*

(1) *when the distance of* X_i *from* $M_1 \cap M_2$ *is less than* $c_A r$, *then* $\hat{\sigma}_i > 2\xi$;

(2) *when the distance of* X_i *from* $M_1 \cap M_2$ *is greater than* $c_B r$, *then* $\hat{\sigma}_i < \xi$,

where $\hat{\sigma}_i = \hat{\sigma}_i(\mathbf{X}, r, \eta)$.

The proof of the theorem requires much work, and it is presented in Supplementary Material SM3. One main tool for the theorem is the Wasserstein distance, instead of the kernel MMD, which is possible since $\Delta(\mu, \nu) \leq \sqrt{2\gamma} \cdot \mathrm{W}(\mu, \nu)$ for the Gaussian kernel $\kappa(x, y) = e^{-\gamma \|x-y\|^2}$ (Lemma SM20). The advantage of the Wasserstein distance is that it is intuitively easy to prove geometric claims.

It has the following key ingredients:

(1) For a fixed $z \in M$ and as $r \to 0, n \to \infty$, the empirical measure $\hat{\mu}(z)$ converges to the uniform distribution over $T_z M^\circ := T_z M \cap \mathcal{B}(0,1)$, where $\mathcal{B}(0,1) \subseteq \mathbb{R}^D$ is the unit ball of radius 1. Convergence is quantified using the Wasserstein distance (Proposition SM13).

(2) The singularity score function $\mu \mapsto \sigma(\mu)$ is a Lipschitz continuous function in μ, where Lipschitz continuity is quantified using the Wasserstein distance (Proposition SM39).

(3) The singularity score of the limiting measure at each point as $r \to 0, n \to \infty$ is zero at smooth points and positive at singular points (Propositions SM41, SM42).

(4) By moving sufficiently far away from the singularities, the local neighborhood of a point only isolates one manifold M_i at a time, instead of cutting through both M_1 and M_2 (Proposition SM19).

To understand the proof, the reader is advised to start from the last part, subsection SM3.5, and work backward to identify the components used in the proof.

We remark that the constants c_A, c_B appearing in the theorem are unfortunately intrinsic features of the singularity detection. Suppose that $x \in M$, the ball of radius r is used to isolate local neighborhood of x, and the distance of x to the singularities of M is $c \cdot r$ where $c \in \mathbb{R}^+$. Then there is an inherent ambiguity in choosing c_0 such that whenever $c > c_0$, x is declared nonsingular, and whenever $c < c_0$, x is declared singular.

5. Experiments. We implemented HADES in Python and performed various computational experiments. Singularity detection lacks a ground truth for most real-world datasets and is an *unsupervised learning* algorithm. We follow the standard two-step approach to assess the performance of a singularity detection algorithm:

(i) **Synthetic data.** We plot singularities detected from 2- and 3-dimensional datasets and visually inspect that the singularities are detected correctly. Then we detect singularities from families of high-dimensional synthetic datasets whose singularities are completely understood by construction, and use a receiver-operating-characteristic (ROC) curve to quantitatively assess accuracy of the algorithm.

(ii) **Real data.** We study datasets of road networks, cyclo-octane conformation, images of handwritten digits, and images of clothing items. For the road network and cyclo-octane conformation datasets, we recover the already-known locations of the singularities. For the image datasets whose geometry is not well-understood, we observe that images with a high singularity score are anomalous from visual inspection.

For details of the experiments, see Supplementary Material SM1.

5.1. Synthetic data: Visualization and ROC curves. We first apply HADES to the 2- and 3-dimensional point clouds in Figure 2, where singular points detected by the algorithm are marked blue. These synthetic datasets are generated from known data distributions of various geometric shapes, and uniform noise has been added to the datasets. They demonstrate that the algorithm is robust to noise and curvature. The algorithm simultaneously detects multiple types of singularities such as intersections, branching points, sharp corners, and cones. We also observe that no singularities are detected for the first row, which consist entirely of manifolds. This is enabled by a heuristic test of the manifold described in subsection 2.4. The sizes of datasets range from 5,000 to 10,000. The time taken to extract singularities was about 3 minutes per dataset.

Going beyond visual inspection, we quantify accuracy of HADES on three families of spaces:

(1) One solid d-dimensional ball (singularity at boundary).

(2) Two d-dimensional spheres intersecting at a $(d-1)$-dimensional sphere (singularity at intersection).

(3) Two $2d$-dimensional disks intersecting orthogonally at a d-dimensional disk (singularity at intersection and boundary).

Visual inspection is inadequate for inspecting high-dimensional singularities, so we use a ROC curve and its AUC to assess the performance. The AUC scores were all ≥ 0.89. The ROC curves and the AUC are shown in Figure 3.

5.2. Synthetic data: Manifold hypothesis. We perform heuristic tests of the manifold hypothesis with HADES using the methods described in the second half of subsection 2.4. Unlike the *local* tests of detecting singularities, we perform a *global* test, on one dataset at a time. For this, datasets consisting of synthetically generated point clouds were created, with a *binary* label on whether each point cloud was a stratified space (with singularities) or a manifold (without singularities). Figure 4 shows the results.

In the figure, we use three different scores to test the manifold hypothesis. The default option explained in subsection 2.4 is termed SUPC (small uniformity p-value concentration), and

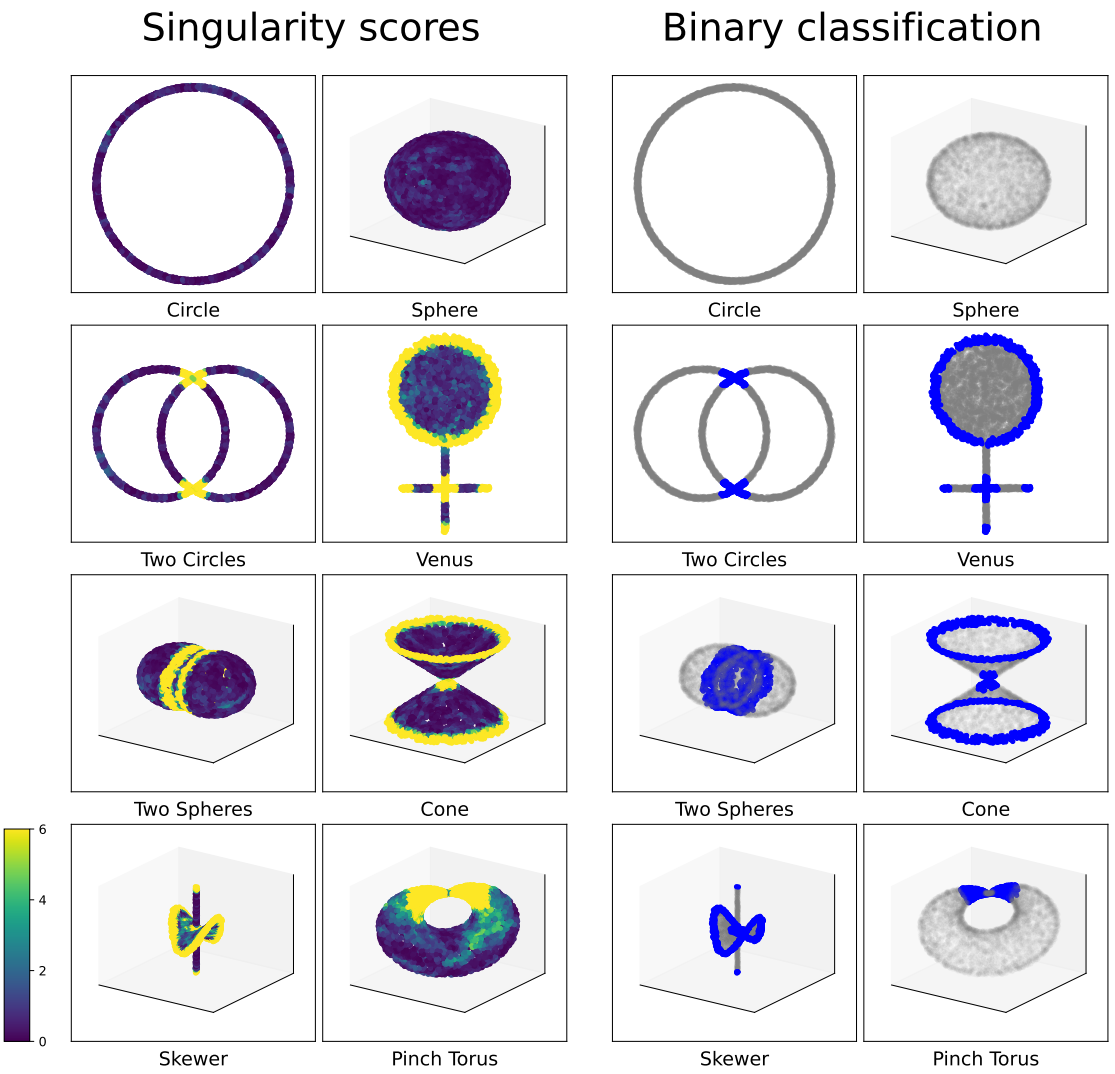

Figure 2. *Singularities discovered by* HADES *marked blue in synthetic datasets.*

we also use two alternative scores: UPUP and KS. Both UPUP and KS compare the empirical distribution of p-values to the uniform distribution over $[0, 1]$; here UPUP (uniformity p-value uniformity p-value) is a kernel MMD test, and KS is a one-sample Kolmogorov–Smirnov test.

As sample sizes increase from 1000 to 8000, AUC values for all of SUPC, UPUP, KS increase, with UPUP and KS reaching just about 0.7 and SUPC reaching the AUC score 1.00. The boxplot on the right shows the distribution of SUPC scores for the manifolds and stratified spaces at sample size 8000, demonstrating a clean separation between the two types of data. This indicates that the manifold hypothesis can be effectively tested with SUPC.

5.3. Real data: Road network. We apply HADES to the Massachusetts Roads Dataset [36], a dataset consisting of pixelized images of road networks in Massachusetts. Each road network is mathematically a planar embedding of a graph. Intersections and sharp corners of

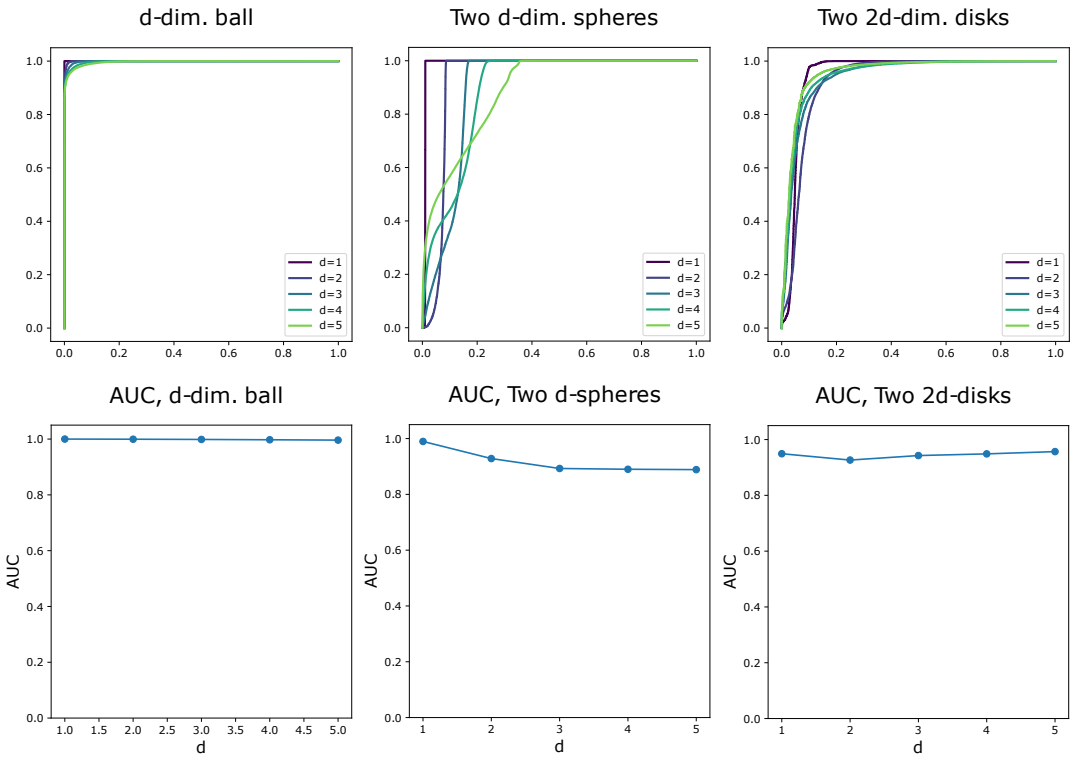

Figure 3. *ROC curve and AUC scores of singularities discovered by* HADES *in synthetic datasets. The Viridis colour bar (left) applies to the first two columns. It indicates the value* $\log_{10}(1/p_i)$ *for the singularity p-value* p_x *evaluated at each data point* x *which ranges over the entire data set per each plot.*

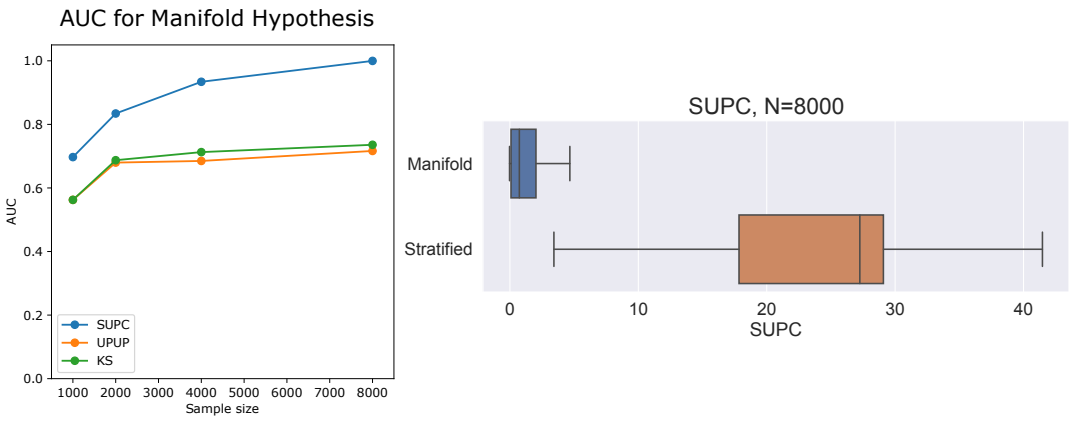

Figure 4. *Testing the manifold hypothesis with* HADES*. Left: AUC of manifold hypothesis performed with three different scores: SUPC, UPUP, KS with synthetic datasets. Right: Boxplots of SUPC for manifolds versus stratified spaces at sample size* 8000*.*

the road are singular points, and everything else is locally a straight line, and thus are smooth points. From Figure 5, visual inspection reveals that singularities are accurately detected. Each image had 1500×1500 resolution, containing 45,000 to 200,000 pixels with nonzero

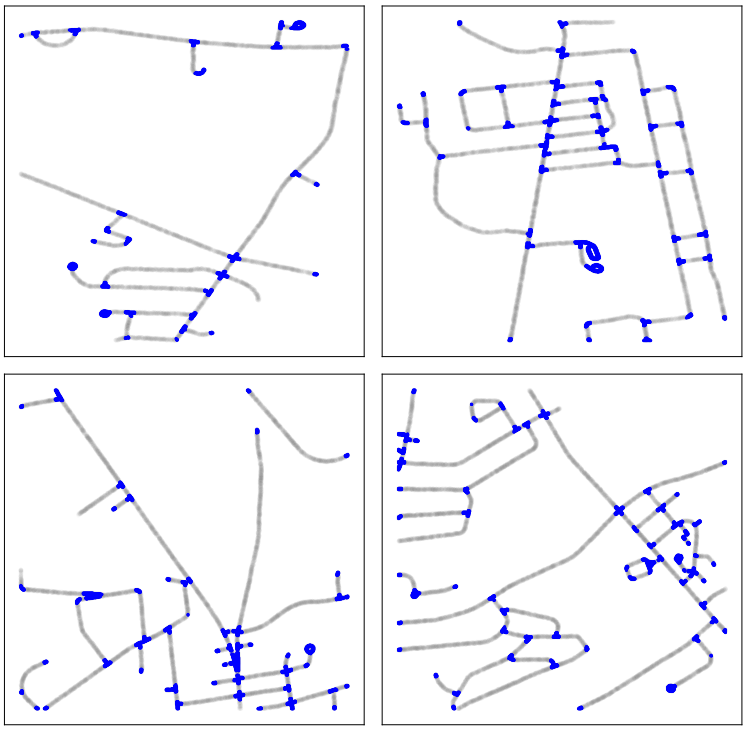

Figure 5. *Singularities discovered by* HADES *marked blue in the Massachusetts Roads Dataset.*

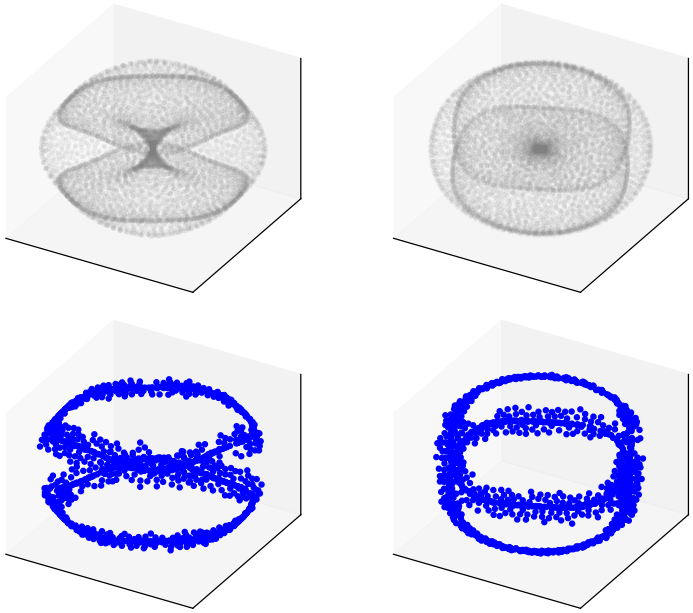

Figure 6. *Singularities discovered by* HADES *marked blue in a cyclo-octane conformation dataset, which are union of two circles. Each row shows rotations of the same* 3-dimensional *Isomap projection of the* 24-*dimensional dataset. The first row shows the whole dataset and the second row shows singularities.*

Figure 7. *Images with the lowest singularity scores (left half) and the highest singularity scores (right half), upon applying HADES to the MNIST handwritten digits dataset.*

brightness values. The time taken to run each dataset ranges from 6 to 31 s. Expanding this analysis, the same computational experiment can be performed to other datasets that can be modeled as (1-dimensional) graphs, including images of neurons, and filamentary structures formed by galaxies.

5.4. Real data: Cyclo-octane conformation. We apply HADES to the dataset of cyclo-octane conformations. This dataset, introduced in [34], consists of 6040 points on the 24-dimensional space $\mathbb{R}^{24}$ that parametrizes 3-dimensional positions of 8 carbon molecules in the cyclo-octane C_8H_{16}. The space of cyclo-octane was previously identified to be the union of a Klein bottle and a sphere, intersecting at two circles [34]. These two circles are singularities of the space of conformations, and indeed they are correctly detected by HADES, as seen in Figure 6. The 3-dimensional projections of the conformation dataset, obtained using the dimensionality reduction algorithm Isomap [46], is displayed in Figure 6; we emphasize that the computation wasn't done on the 3-dimensional projection, and instead was done directly on the original 24-dimensional data.

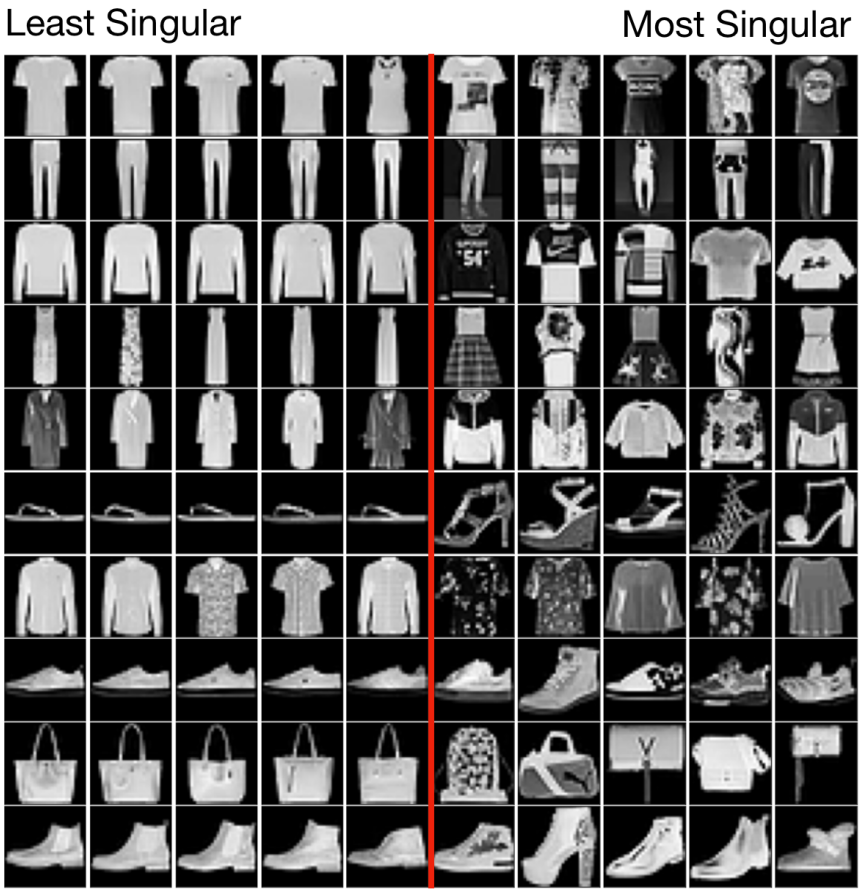

Figure 8. *Images with the lowest singularity scores (left half) and the highest singularity scores (right half), upon applying HADES to the Fashion-MNIST dataset.*

Running HADES on the entire conformation dataset took 5 s on a standard laptop. This shows great improvement from the previous benchmark for this dataset, in [45], in which their singularity detection algorithm *Geometric Anomaly Detection* took at least several hours on parallel processing, as informed by the first author in a private communication.

5.5. Real data: Images of handwritten digits and clothing. We apply HADES to image datasets, of handwritten digits (MNIST) and clothing items (Fashion-MNIST), and find that images with high singularity scores are visibly more anomalous. MNIST is a standard dataset of images of handwritten digits [28] consisting of 60,000 data points, where there are 6,000 data points for each digit from $0, 1, \ldots 9$. Each data point is a $28 \times 28 = 784$-dimensional vector of brightness values between 0 and 1, where each entry of the vector indicates the brightness value of each pixel in the image. Similarly, the Fashion-MNIST dataset consists of 28×28 images of 10 classes of clothing items,[5] where there are 6,000 data points per class.

[5]T-shirt, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, Ankle boot.

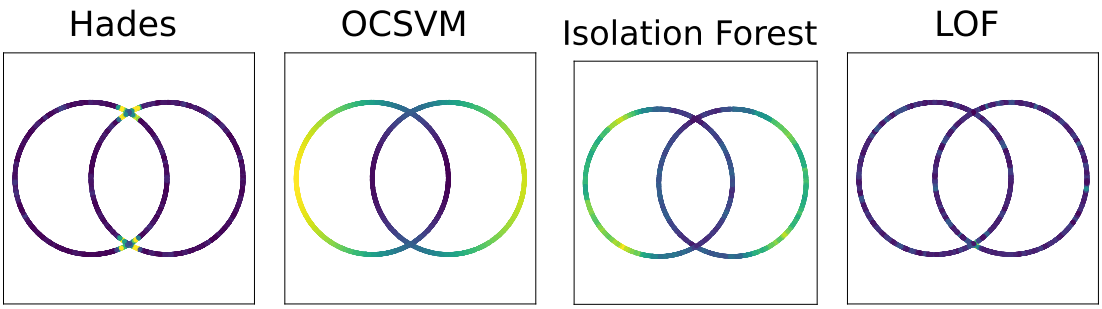

Figure 9. *HADES is different from the* anomaly detection *algorithms; points marked as highly anomalous are marked in yellow.*

We applied HADES on MNIST and Fashion-MNIST datasets on each class of 6,000 images[6] and sorted the images according to their singularity scores. Prior to applying HADES, each 784-dimensional image vector was reduced to a 100-dimensional vector by applying discrete cosine transform. Figure 7 (MNIST) and Figure 8 (Fashion-MNIST) show the results, where the left half of each figure displays images with the lowest singularity scores and the right half displays images with the highest singularity scores.

Images on the right half have irregular characteristics when compared to images on the left. This is explained by the fact that HADES assesses *local uniformity*. Indeed, images on the left look similar to each other, indicating that there are a lot more similar images of small, subtle variations, thus locally constituting a more uniform distribution with a well-behaved variation. On the other hand, images on the right arise from irregular handwriting and clothing items. This means that there wouldn't be a uniform distribution of similar variations of the images, and thus picked up by HADES as highly singular. The computation time for running HADES on 6,000 images corresponding to each digit spanned 30 s to 45 s.

5.6. Anomaly detection. We remark that HADES has a different objective to existing anomaly detection algorithms. Whereas HADES detects anomalies in local geometry, existing anomaly detection algorithms detect outliers. Along with HADES, three anomaly detection algorithms were tested in Figure 9 (One-Class SVM [41], Isolation Forest [32], Local Outlier Factor [15]). The points with a high anomaly score are marked in yellow (viridis colormap).

6. Conclusion. We introduced and studied HADES, an unsupervised learning algorithm that assigns a singularity score to data points. This is done by measuring how much the local geometry deviates from a manifold using a goodness-of-fit. The strengths of the algorithm are first its speed, in particular compared with recent topological approaches, and second that it can be seen as a first step toward learning the full stratified space. The main disadvantage is that the goodness-of-fit algorithm simply detects what is *not* like a disk, and doesn't give further details about the local geometry. This is where future research may blossom by using the richer information of local geometry provided by topological methods; for example, one may compute persistent homology only at points declared to be singular by HADES. These research works together aim to create a computational toolbox for modeling general data using stratified spaces.

[6]Similar results were obtained from running HADES on the entire dataset of 60,000 images.

REFERENCES

[1] M. Adamaszek and H. Adams, *The Vietoris–Rips complexes of a circle*, Pacific J. Math., 290 (2017), pp. 1–40, https://doi.org/10.2140/pjm.2017.290.1.

[2] E. Arias-Castro, *Clustering based on pairwise distances when the data is of mixed dimensions*, IEEE Trans. Inform. Theory, 57 (2011), pp. 1692–1706, https://doi.org/10.1109/TIT.2011.2104630.

[3] E. Arias-Castro, G. Lerman, and T. Zhang, *Spectral clustering based on local PCA*, J. Mach. Learn. Res., 18 (2017), pp. 253–309, https://jmlr.org/papers/v18/14-318.html.

[4] A. Babaeian, A. Bayestehtashk, and M. Bandarabadi, *Multiple manifold clustering using curvature constrained path*, PLOS One, 10 (2015), e0137986, https://doi.org/10.1371/journal.pone.0137986.

[5] P. Bartlett, *Theoretical statistics; stat* 210*b*, 2013, https://www.stat.berkeley.edu/∼bartlett/courses/2013spring-stat210b/.

[6] U. Bauer, *Ripser: Efficient computation of Vietoris–Rips persistence barcodes*, J. Appl. Comput. Topol., 5 (2021), pp. 391–423, https://doi.org/10.1007/s41468-021-00071-5.

[7] M. Belkin and P. Niyogi, *Laplacian eigenmaps for dimensionality reduction and data representation*, Neural Comput., 15 (2003), pp. 1373–1396, https://doi.org/10.1162/089976603321780317.

[8] P. Bendich, D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and D. Morozov, *Inferring local homology from sampled stratified spaces*, in Proceedings of the 48th Annual Symposium on Foundations of Computer Science (FOCS'07), IEEE, 2007, pp. 536–546.

[9] P. Bendich and J. Harer, *Persistent intersection homology*, Found. Comput. Math., 11 (2011), pp. 305–336, https://doi.org/10.1007/s10208-010-9081-1.

[10] P. Bendich, B. Wang, and S. Mukherjee, *Local homology transfer and stratification learning*, in Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms, 2012, pp. 1355–1370.

[11] T. Birdal, A. Lou, L. J. Guibas, and U. Simsekli, *Intrinsic dimension, persistent homology and generalization in neural networks*, Adv. Neural Inf. Process. Syst., 34 (2021), pp. 6776–6789, https://proceedings.neurips.cc/paper/2021/hash/35a12c43227f217207d4e06ffefe39d3-Abstract.html.

[12] E. Boissard and T. Le Gouic, *On the mean speed of convergence of empirical and occupation measures in Wasserstein distance*, Ann. Inst. Henri Poincaré Probab. Stat., 50 (2014), pp. 539–563.

[13] Y. Bokor, K. Turner, and C. Williams, *Reconstructing linearly embedded graphs: A first step to stratified space learning*, Found. Data Sci., 4 (2022), pp. 537–561, https://doi.org/10.3934/fods.2021026.

[14] Y. Bokor Bleile, *Towards Stratified Space Learning:* 2*-Complexes*, https://arxiv.org/abs/2305.02724, 2023.

[15] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, *LOF: Identifying density-based local outliers*, in Proceedings of the ACM SIGMOD International Conference on Management of Data, 2000, pp. 93–104.

[16] B. C. Brown, A. L. Caterini, B. L. Ross, J. C. Cresswell, and G. Loaiza-Ganem, *The union of manifolds hypothesis*, in NeurIPS Workshop on Symmetry and Geometry in Neural Representations, 2022.

[17] B. C. Brown, A. L. Caterini, B. L. Ross, J. C. Cresswell, and G. Loaiza-Ganem, *Verifying the union of manifolds hypothesis for image data*, in Proceedings of the 11th International Conference on Learning Representations, 2022.

[18] F. Camastra and A. Staiano, *Intrinsic dimension estimation: Advances and open problems*, Inform. Sci., 328 (2016), pp. 26–41, https://doi.org/10.1016/j.ins.2015.08.029.

[19] F. Camastra and A. Vinciarelli, *Estimating the intrinsic dimension of data with a fractal-based method*, IEEE Trans. Pattern Anal. Mach. Intell., 24 (2002), pp. 1404–1407, https://doi.org/10.1109/TPAMI.2002.1039212.

[20] C. Ceruti, S. Bassis, A. Rozza, G. Lombardi, E. Casiraghi, and P. Campadelli, *DANCo: An intrinsic dimensionality estimator exploiting angle and norm concentration*, Pattern Recognit., 47 (2014), pp. 2569–2581, https://doi.org/10.1016/j.patcog.2014.02.013.

[21] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport*, Adv. Neural Inf. Process. Syst., 26 (2013), https://papers.nips.cc/paper_files/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html.

[22] C. Davis and W. M. Kahan, *The rotation of eigenvectors by a perturbation.* III, SIAM J. Numer. Anal., 7 (1970), pp. 1–46, https://doi.org/10.1137/0707001.
[23] A. M. Farahmand, C. Szepesvári, and J.-Y. Audibert, *Manifold-adaptive dimension estimation*, in Proceedings of the 24th International Conference on Machine Learning, 2007, pp. 265–272.
[24] G. Haro, G. Randall, and G. Sapiro, *Stratification learning: Detecting mixed density and dimensionality in high dimensional point clouds*, Adv. Neural Inf. Process. Syst., 19 (2006), https://proceedings.neurips.cc/paper_files/paper/2006/hash/51be2fed6c55f5aa0c16ff14c140b187-Abstract.html.
[25] M. Helmer and V. Nanda, *Conormal spaces and Whitney stratifications*, Found. Comput. Math., 23 (2023), pp. 1745–1780, https://link.springer.com/article/10.1007/s10208-022-09574-8.
[26] A. J. Hoffman and H. W. Wielandt, *The variation of the spectrum of a normal matrix*, in Selected Papers of Alan J. Hoffman: With Commentary, World Scientific, Singapore, 2003, pp. 118–120.
[27] K. Johnsson, C. Soneson, and M. Fontes, *Low bias local intrinsic dimension estimation from expected simplex skewness*, IEEE Trans. Pattern Anal. Mach. Intell., 37 (2014), pp. 196–202, https://doi.org/10.1109/TPAMI.2014.2343220.
[28] Y. LeCun, *The MNIST Database of Handwritten Digits*, 1998, http://yann. lecun. com/exdb/mnist/.
[29] E. Levina and P. Bickel, *Maximum likelihood estimation of intrinsic dimension*, Adv. Neural Inf. Process. Syst., 17 (2004), https://papers.nips.cc/paper_files/paper/2004/hash/74934548253bcab8490ebd74afed7031-Abstract.html.
[30] U. Lim, *Strange random topology of the circle*, in 40th International Symposium on Computational Geometry (SoCG 2024), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024, https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.SoCG.2024.70.
[31] U. Lim, H. Oberhauser, and V. Nanda, *Tangent space and dimension estimation with the wasserstein distance*, SIAM J Appl. Algebr. Geom., 8 (2024), pp. 650–685, https://doi.org/10.1137/22M1522711.
[32] F. T. Liu, K. M. Ting, and Z.-H. Zhou, *Isolation forest*, in Proceedings of the 8th International Conference on Data Mining, IEEE, 2008, pp. 413–422.
[33] T. Mäder and L. Waas, *From samples to persistent stratified homotopy types*, J. Appl. Comput. Topol., 8 (2024), pp. 761–838.
[34] S. Martin, A. Thompson, E. A. Coutsias, and J.-P. Watson, *Topology of cyclo-octane energy landscape*, J. Chem. Phys., 132 (2010), 234115, https://doi.org/10.1063/1.3445267.
[35] L. McInnes, J. Healy, and J. Melville, *UMAP: Uniform Manifold Approximation and Projection*, J. Open Source Softw., 3 (2018).
[36] V. Mnih, *Machine Learning for Aerial Image Labeling*, Ph.D. thesis, University of Toronto, 2013.
[37] V. Nanda, *Local cohomology and stratification*, Found. Comput. Math., 20 (2020), pp. 195–222, https://doi.org/10.1007/s10208-019-09424-0.
[38] P. Niyogi, S. Smale, and S. Weinberger, *Finding the homology of submanifolds with high confidence from random samples*, Discrete Comput. Geom., 39 (2008), pp. 419–441, https://doi.org/10.1007/s00454-008-9053-2.
[39] H. Rootzen, *Extreme value theory for moving average processes*, Ann. Probab., 14 (1986), pp. 612–652.
[40] V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, *Finding a "kneedle" in a haystack: Detecting knee points in system behavior*, in Proceedings of the 31st International Conference on Distributed Computing Systems Workshops, IEEE, 2011, pp. 166–171.
[41] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, *Estimating the support of a high-dimensional distribution*, Neural Comput., 13 (2001), pp. 1443–1471, https://doi.org/10.1162/089976601750264965.
[42] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York, 2009.
[43] K. Slavakis, S. Salsabilian, D. S. Wack, S. F. Muldoon, H. E. Baidoo-Williams, J. M. Vettel, M. Cieslak, and S. T. Grafton, *Riemannian multi-manifold modeling and clustering in brain networks*, in Wavelets and Sparsity XVII, Proc. SPIE 10394, SPIE, 2017, pp. 9–24, https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10394/103940B/Riemannian-multi-manifold-modeling-and-clustering-in-brain-networks/10.1117/12.2274405.short.
[44] M. Soltanolkotabi and E. J. Candes, *A geometric analysis of subspace clustering with outliers*, Ann. Statist., 40 (2012), pp. 2195–2238, https://doi.org/10.1214/12-AOS1034.

[45] B. J. Stolz, J. Tanner, H. A. Harrington, and V. Nanda, *Geometric anomaly detection in data*, Proc. Natl. Acad. Sci. USA, 117 (2020), pp. 19664–19669, https://doi.org/10.1073/pnas.2001741117.

[46] J. B. Tenenbaum, V. de Silva, and J. C. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science, 290 (2000), pp. 2319–2323, https://doi.org/10.1126/science.290.5500.2319.

[47] L. Van der Maaten and G. Hinton, *Visualizing data using t-SNE*, J. Mach. Learn. Res., 9 (2008), https://www.jmlr.org/papers/v9/vandermaaten08a.html.

[48] L. J. P. Van Der Maaten, E. O. Postma, and H. J. van den Herik, *Dimensionality reduction: A comparative review*, Tilburg University Technical Report, TiCC-TR 2009-005, 2009, https://www.researchgate.net/profile/Eric-Postma/publication/228657549_Dimensionality_Reduction_A_Comparative_Review/links/0046353a3047fc2863000000/Dimensionality-Reduction-A-Comparative-Review.pdf.

[49] V. Vasudevan and M. Ramakrishna, *A Hierarchical Singular Value Decomposition Algorithm for Low Rank Matrices*, preprint, arXiv:1710.02812, 2017.

[50] J. Von Rohrscheidt and B. Rieck, *Topological singularity detection at multiple scales*, in International Conference on Machine Learning, PMLR, 2023, https://proceedings.mlr.press/v202/von-rohrscheidt23a.html.

[51] X. Wang, K. Slavakis, and G. Lerman, *Multi-manifold modeling in non-euclidean spaces*, in Artificial Intelligence and Statistics, PMLR, 2015, pp. 1023–1032.

[52] Y. Yu, T. Wang, and R. J. Samworth, *A useful variant of the Davis–Kahan theorem for statisticians*, Biometrika, 102 (2015), pp. 315–323, https://doi.org/10.1093/biomet/asv008.

[53] Z. Zhang and H. Zha, *Principal manifolds and nonlinear dimensionality reduction via tangent space alignment*, SIAM J. Sci. Comput., 26 (2004), pp. 313–338, https://doi.org/10.1137/S1064827502419154.