ISO-DEPTH LEARNING FOR ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

We consider unsupervised anomaly detection for data that lie near a low-dimensional manifold. Our interest is in detecting points that sit far away from the center of the data cloud but remain close to the manifold. Statistical depth in an ambient space may be inadequate for detecting such points, because Euclidean distance can make points that are far apart along the manifold appear close. Local density scores may also fail to detect such points, because they can have sufficiently many nearby neighbors despite being far from the global center.

We propose an Iso-depth framework for anomaly detection that first estimates the intrinsic geometry of the data via ISOMAP and then computes statistical depth in the resulting Euclidean embedding. Iso-depth measures global centrality intrinsically. To facilitate efficient prediction, we train a neural network to approximate the depth function, allowing new observations to be scored with minimal computational cost. We derive an error bound with three components: depth estimation, the ISOMAP embedding, and neural approximation. For halfspace depth, we show how root mean square coordinate error transfers to depth error. We evaluate Iso-depth in controlled synthetic experiments that separate depth estimation, embedding, and out-of-sample prediction errors, and in a real-data application that examines the resulting anomaly rankings.

1 INTRODUCTION

Images, functional observations, time series, and other object-valued data often satisfy geometric constraints that concentrate observations near a low-dimensional manifold (Tenenbaum et al., 2000). When a manifold bends or folds in the ambient space, observations can be far from the center of the distribution along the manifold yet close to other observations in ambient Euclidean distance. We call these observations on-manifold peripheral anomalies. Their detection depends on centrality along the manifold, even when they have nearby neighbors in the original space.

Existing geometric methods measure different kinds of departure from the data. Olson et al. (2018) learn a background manifold with kernel PCA or diffusion maps and score new observations by projection error. Huang et al. (2014) use a heat-diffusion embedding to assess neighborhood density. Herrmann et al. (2023) distinguish structural outliers off the manifold from distributional outliers at its margins and apply outlier scores to manifold embeddings, including multidimensional scaling followed by LOF. In these examples, projection error measures departure from the manifold, while density-based scores compare local neighborhoods. Neither type of score directly measures how far an observation lies from the center of the distribution along the manifold.

Statistical depth provides that center-outward ordering. A depth function $D(\cdot, P)$ assigns low values to points at the periphery of P (Zuo & Serfling, 2000). We define population Iso-depth as depth in embedded coordinates, then estimate those coordinates with ISOMAP. For new observations, a neural network approximates the resulting depth function. Its negative gives an anomaly score that increases as intrinsic depth decreases.

The framework allows several depth functions, but our convergence analysis focuses on halfspace depth. We separate the error of the learned score into depth estimation error, ISOMAP coordinate error, and neural network approximation error. Standard VC theory controls empirical halfspace depth, and existing ISOMAP results control coordinate recovery. We show how root mean square coordinate error transfers to halfspace depth error.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The ISOMAP analysis uses an r_n-neighborhood graph whose radius shrinks with sample size to balance the manifold's curvature in the ambient space and sample coverage. The results require a manifold isometric to a convex Euclidean domain and reliable recovery of its geometry. The out-of-sample bound also requires the stated network approximation condition.

The experiments examine the sources of error separately. A synthetic model with known intrinsic coordinates and population depth separates depth estimation error from the effect of estimating geometry. Paired neural fits trained on estimated and population depth targets are evaluated on test observations. We then compare anomaly detection performance for peripheral observations on a gapped circle and a Swiss roll. A Cardiotocography application examines the resulting rankings and their sensitivity to feature scaling and graph construction.

2 METHODS

2.1 POPULATION-LEVEL DEVELOPMENT

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X : \Omega \to M$ be measurable, where $M \subset \mathbb{R}^p$ is a manifold with Borel σ-algebra $\mathcal{B}_M$ and geodesic metric d_M.

Following Tenenbaum et al. (2000) and Arias-Castro et al. (2020), we assume that M admits an injective map $\phi : M \to S \subset \mathbb{R}^d$, where $d < p$ and $S = \phi(M)$ is convex, such that

$$d_M(x_1, x_2) = \|\phi(x_1) - \phi(x_2)\|$$

for all $x_1, x_2 \in M$. Here, the manifold may bend in the ambient space, as in a Swiss roll, while remaining intrinsically flat. We also require $\phi(M)$ to be convex, a global condition that does not follow from intrinsic flatness alone.

Let $P_X = P \circ X^{-1}$. For a Euclidean depth function D, we measure the on-manifold peripherality of $x \in M$ by $D(\phi(x), P_X \circ \phi^{-1})$. Smaller depth means greater peripherality. Other depth functions are described in Appendix A.1.

For halfspace depth, also called Tukey depth,

$$D_{\mathrm{H}}(x,P) = \inf_{u \in S^{d-1}} P\big(u^\top X \le u^\top x\big), \tag{1}$$

where $X \sim P$ and S^{d-1} is the unit sphere in $\mathbb{R}^d$. Taking the infimum over all directions makes the depth sensitive to globally peripheral observations (Tukey, 1975; Donoho & Gasko, 1992).

Although we have assumed that ϕ preserves the geodesic distance $d_M(x_1, x_2)$, the proposed method remains meaningful when this assumption holds only approximately. In particular, even when ϕ does not preserve geodesic distances exactly, it can still largely retain the relative peripherality, or outlyingness, of points along the manifold M.

To fix ideas, consider a folded surface. A point near the edge of the surface may be far from its central region when distance is measured along the surface, yet be close to that region in the ambient Euclidean space. The map ϕ unfolds the surface into a flat Euclidean representation. Although this representation need not preserve every geodesic distance exactly, a point that is outlying along the surface will typically remain outlying in the embedded space. This contrasts with direct use of ambient Euclidean distance, under which a point that is peripheral along the folded surface may appear close to its center.

2.2 ISOMAP

ISOMAP estimates intrinsic coordinates from graph-based approximations to geodesic distances (Tenenbaum et al., 2000). Given $X_1, \ldots, X_n \in \mathbb{R}^p$, we form a weighted graph G_n by connecting i and j when $\|X_i - X_j\|_2 \le r_n$, with edge weight $\|X_i - X_j\|_2$. We assume G_n is connected. Let $\Gamma = (\gamma_{ij})$ be its shortest-path distance matrix. Under appropriate sampling and graph conditions, these distances approximate manifold geodesic distances (Bernstein et al., 2000).

Classical multidimensional scaling (Torgerson, 1952; Gower, 1966) uses $H = I_n - n^{-1}\mathbf{1}\mathbf{1}^\top$ to form

$$B = -\tfrac{1}{2} H \Gamma^{\circ 2} H, \tag{2}$$

where $\Gamma^{\circ 2}$ denotes entrywise squaring. If (λ_k, v_k) are the d largest positive eigenvalue–eigenvector pairs of B, the embedding is

$$\hat{Z} = [\sqrt{\lambda_1}v_1, \ldots, \sqrt{\lambda_d}v_d]. \tag{3}$$

Its ith row, $\hat{\phi}_n(X_i)$, estimates the intrinsic coordinates of X_i.

The theory treats d as the fixed intrinsic dimension. Synthetic experiments use the known generating dimension. For Cardio, we choose the smallest dimension whose distance residual variance is at most 0.05 (Section 6.6), using training observations for out-of-sample fits. The convergence results do not account for dimension-selection error.

2.3 Sample-level implementation

Let $X_1, \ldots, X_n$ be i.i.d. observations from P_X. ISOMAP gives coordinates $\hat{\phi}(X_1), \ldots, \hat{\phi}(X_n)$, with $\hat{\phi}(X_i)$ the ith row of $\hat{Z}$ in equation 3.

In the second stage, we estimate the population Iso-depth $D(\phi(X_i), P_X \circ \phi^{-1})$. We use the empirical measure

$$P_n \circ \hat{\phi}^{-1} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\hat{\phi}(X_i)}$$

to estimate the distribution $P_X \circ \phi^{-1}$, where $\delta_{\hat{\phi}(X_i)}$ is the Dirac measure at the point $\hat{\phi}(X_i) \in \mathbb{R}^d$. The estimated Iso-depth of an observed point X_i is

$$D\big(\hat{\phi}(X_i), P_n \circ \hat{\phi}^{-1}\big).$$

The sample anomaly score is the negative of this depth. We analyze its error before constructing the out-of-sample estimator. Appendix A.2 summarizes the full procedure, including the out-of-sample extension, in Algorithm 1.

3 Convergence rate

The empirical halfspace-depth bound is a standard consequence of VC theory (van der Vaart & Wellner, 1996). We adapt the deterministic ISOMAP bound of Arias-Castro et al. (2020) to i.i.d. sampling. Our depth-specific step transfers RMS coordinate perturbation to empirical halfspace-depth error (Theorem 3). The out-of-sample analysis combines this step with a Bernstein covering argument and an explicit network approximation condition; Appendix B identifies these components.

Theorem 1 *Suppose $X_1, \ldots, X_n$ are i.i.d. from P_X. Let $S = \phi(M)$ and keep d fixed. If D is the halfspace depth, then*

$$\sup_{z \in S} \big| D\big(z, P_X \circ \phi^{-1}\big) - D\big(z, P_n \circ \phi^{-1}\big) \big| = O_p(n^{-1/2}).$$

Theorem 1 controls empirical-depth sampling error. The additional error from replacing latent with recovered coordinates is

$$T_{2,i} = D\big(\phi(X_i), P_n \circ \phi^{-1}\big) - D\big(\hat{\phi}(X_i), P_n \circ \hat{\phi}^{-1}\big).$$

Let $\mathcal{A}(d)$ denote the group of Euclidean rigid transformations on $\mathbb{R}^d$. Thus, each $T \in \mathcal{A}(d)$ is of the form

$$T(z) = Qz + b,$$

where $Q \in \mathbb{R}^{d \times d}$ is an orthogonal matrix and $b \in \mathbb{R}^d$. We assume that the data depth is invariant under rigid transformations; that is, for every $T \in \mathcal{A}(d)$, $z \in \mathbb{R}^d$, and probability measure P on $\mathbb{R}^d$,

$$D\big(T(z), P \circ T^{-1}\big) = D(z, P). \tag{4}$$

This invariance property is satisfied by the depth functions considered in this paper, including halfspace, projection, spatial, and simplicial depths.

We measure coordinate error after alignment by

$$\delta^2_{\mathrm{ISO},n} := \min_{T\in\mathcal{A}(d)} \frac{1}{n}\sum_{i=1}^{n} \big\|\hat{\phi}(X_i) - (T\circ\phi)(X_i)\big\|^2,$$

and write $\|v\|_{n,2} = (n^{-1}\sum_i v_i^2)^{1/2}$. Let

$$Z = \phi(X), \qquad P_Z = P_X \circ \phi^{-1}.$$

Theorem 2 (ISOMAP rate under i.i.d. sampling) *Suppose*

1. *$M \subset \mathbb{R}^p$ is a compact, connected C^2 embedded manifold with boundary and reach at least $\tau > 0$;*
2. *$\phi : M \to S$ is an isometry, where $S \subset \mathbb{R}^d$ is a compact convex set with nonempty interior;*
3. *P_Z has a density p_Z satisfying*

$$p_Z(z) \geq p_{\min} > 0, \qquad z \in S;$$

4.

$$r_n \asymp \left(\frac{\log n}{n}\right)^{1/(2d)}.$$

If $X_1, \ldots, X_n$ are i.i.d. from P_X and $\hat{\phi}(X_1), \ldots, \hat{\phi}(X_n) \in \mathbb{R}^d$ are the coordinates returned by ISOMAP, then

$$\delta_{\mathrm{ISO},n} = O_p\left\{\left(\frac{\log n}{n}\right)^{1/d}\right\}. \tag{5}$$

We now translate the ISOMAP embedding rate into a convergence rate for $\|T_2\|_{n,2}$, focusing on the halfspace depth, denoted by $D_H(z, P)$.

Theorem 3 *Suppose that the assumptions of Theorem 2 hold. If, for $Z = \phi(X)$, every one-dimensional projection $u^\top Z$, $u \in \mathbb{S}^{d-1}$, has a density f_u satisfying*

$$\sup_{u\in\mathbb{S}^{d-1}} \|f_u\|_\infty \leq M$$

for some $M < \infty$, then

$$\|T_2\|_{n,2} = O_p\Big(n^{-1/2} + \delta^{2/3}_{\mathrm{ISO},n}\Big).$$

Consequently,

$$\|T_2\|_{n,2} = O_p\left\{n^{-1/2} + \left(\frac{\log n}{n}\right)^{2/(3d)}\right\}.$$

The proof aligns the recovered coordinates by a rigid transformation. An ε-slab controls points whose halfspace membership can change, and the fraction of larger coordinate errors is bounded by $\delta^2_{\mathrm{ISO},n}/\varepsilon^2$. Balancing these terms at $\varepsilon = \delta^{2/3}_{\mathrm{ISO},n}$ yields the geometric exponent. Appendix B gives the full argument.

4 ISO-DEPTH FOR DATA NEAR MANIFOLDS

The theoretical development in the preceding section assumes that the observations lie exactly on a d-dimensional manifold M. We model approximate manifold support by allowing small perturbations of these observations.

We formulate "near a manifold" through a local asymptotic model. For each n, let

$$X_i = X_i^\circ + \xi_{i,n}, \qquad i = 1, \ldots, n,$$

where $X_1^\circ, \ldots, X_n^\circ$ are i.i.d. from a distribution P_X° supported on M, and $\xi_{i,n}$ represents the departure from the manifold. We take b_n to be the scale of this departure and assume that, conditionally on X_i°,

$$E(\xi_{i,n} \mid X_i^\circ) = 0, \qquad E\big(\|\xi_{i,n}\|_2^2 \mid X_i^\circ\big) \le C b_n^2, \tag{6}$$

together with the uniform sub-Gaussian tail condition

$$P\left(\|\xi_{i,n}\|_2 > t b_n \mid X_i^\circ\right) \le C_1 \exp(-C_2 t^2), \qquad t > 0, \tag{7}$$

where C, C_1, C_2 do not depend on i or n. Thus, b_n may be interpreted as the scale of the standard deviation of the random departure from the underlying manifold.

Write $q_n = b_n\sqrt{\log n}$ and require $q_n \to 0$. Appendix C develops the perturbation argument with a noise-dependent radius and additional network-scale conditions. Conditional on the noisy graph-distance and risk bounds, it identifies an added depth-error term $q_n^{4/9}$ under those conditions. The fixed-noise experiments assess finite-sample behavior outside this shrinking-noise regime.

5 Out-of-sample scoring

In the third stage, we learn the Iso-depth function with a neural network. Treating the estimated Iso-depth

$$Y_i := D\big(\hat{\phi}(X_i), P_n \circ \hat{\phi}^{-1}\big)$$

as the pseudo-response for X_i, we then train a neural network to learn the map from the original observations to their corresponding Iso-depth values. This allows us to compute the Iso-depth value for any out-of-sample observation without rerunning the preceding Iso-depth procedure.

For the theoretical development, let $\mathcal{F}(W, L, B)$ denote the class of ReLU networks $f : \mathbb{R}^p \to \mathbb{R}$ with at most L layers, at most W nonzero parameters, and all weights and biases bounded in absolute value by B. For sequences W_n, L_n, and B_n, define

$$\mathcal{F}_n := \mathcal{F}(W_n, L_n, B_n).$$

Let

$$\Pi(t) := \max\{0, \min\{1, t\}\}$$

denote the projection onto $[0, 1]$, and define the corresponding class of projected networks by

$$\mathcal{G}_n := \{\Pi \circ f : f \in \mathcal{F}_n\}.$$

The analysis assumes exact empirical risk minimization over the unprojected class. Specifically, let

$$\tilde{f} \in \operatorname*{arg\,min}_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^{n} \{f(X_i) - Y_i\}^2,$$

and define

$$\hat{f} := \Pi \circ \tilde{f}.$$

Thus $\hat{f} \in \mathcal{G}_n$ and, by construction, $0 \le \hat{f}(x) \le 1$ for every $x \in \mathbb{R}^p$. The function $\hat{f}$ estimates population Iso-depth

$$f_0(x) := D\big(\phi(x), P_X \circ \phi^{-1}\big).$$

For an independent $X \sim P_X$, condition on the training sample and define

$$\mathcal{E}(\hat{f})^2 = \mathbb{E}_X\left[(\hat{f}(X) - f_0(X))^2\right] = \int_M (\hat{f}(x) - f_0(x))^2 \, dP_X(x).$$

The uniform risk bound contributes $\eta_{\mathrm{gen},n} = (W_n \log n/n)^{1/2}$ to root mean square error. This term is retained in the final rate: the chosen W_n balances it with $W_n^{-1/d}$. Appendix B gives the decomposition.

Theorem 4 (Out-of-sample error of Iso-depth) *Suppose Theorem 3 and Lemma 5 apply, and the network class admits uniform approximation error $O(W_n^{-1/d})$ for f_0. If*

$$W_n \asymp \left(\frac{n}{\log n}\right)^{d/(d+2)},$$

then

$$\mathcal{E}(\hat{f}) = O_p\left\{\left(\frac{\log n}{n}\right)^{2/(3d)} + \left(\frac{\log n}{n}\right)^{1/(d+2)}\right\}. \tag{8}$$

6 Numerical experiments

We examine three parts of the proposed procedure: recovery of population depth from an estimated intrinsic geometry, prediction of that depth for independent observations, and anomaly detection at an intrinsic boundary. The first two experiments use a distribution with a known geodesic metric and population halfspace depth. The detection experiment uses a noisy Swiss roll and a noisy gapped circle.

6.1 Experimental settings

Geometry recovery and OOS settings. For geometry recovery, let Z be uniform on the unit ball in $\mathbb{R}^d$. We roll its first coordinate along the unit-speed arc $\gamma(t) = (\cos(\beta t), \sin(\beta t))/\beta$, $\beta = 3\pi/4$, and apply a fixed orthonormal embedding into $\mathbb{R}^{50}$. Figure 1a shows the rolled-disk observations with the outer 3% latent shell highlighted. The map is isometric, so the geodesic metric and population halfspace depth D_0 are known. The main experiment uses $d = 2$, 50 seeds, and $n \in \{250, 400, 630, 1000, 1600, 2500\}$. Appendix D.1 reports the one-dimensional results. We construct a radius graph, compute shortest paths and classical MDS coordinates, and then calculate empirical halfspace depth. The theoretically balanced radius is $r_n = 0.7\{\log(n)/n\}^{1/4}$. With P_n^Z and $P_n^{\widehat{Z}}$ denoting the empirical latent and recovered distributions, respectively, we use

$$T_{1i} = D_0(Z_i) - D_H(Z_i; P_n^Z), \qquad T_{2i} = D_H(Z_i; P_n^Z) - D_H(\widehat{Z}_i; P_n^{\widehat{Z}}).$$

Here $E_{T_1} = \|T_1\|_{n,2}$ is sampling error, $E_{T_2} = \|T_2\|_{n,2}$ is the added geometry error, and $E_D = \|T_1 + T_2\|_{n,2}$. Alternative radius rules assess graph connectivity and sensitivity of depth recovery to the radius (Appendix D.2). All errors are conditional on a connected graph. For out-of-sample prediction, we fit ReLU networks with two hidden layers to the estimated depths at $n \in \{250, 630, 1000, 2500\}$ and evaluate 10,000 independent test observations. Paired networks trained on $D_0(Z_i)$ separate the effect of the estimated training targets. Each network uses the same 1,000-epoch Adam budget.

Detection of peripheral anomaly points. We evaluate how well the methods detect points near the ends of a gapped circle and the edges of a Swiss roll, a 1.5-turn roll of the latent rectangle $[-1, 1] \times [-0.35, 0.35]$.

Their anomaly classes are the outer 3% endpoint or boundary regions in latent coordinates. Both models have $n = 1000$, $p = 50$, Gaussian nuisance coordinates, and 100 paired replications.

For ISOMAP, we set $r_{\text{obs}} = 1.10 r_{\text{conn}}$, where r_{conn} is the smallest radius that connects the observed-distance graph. Baselines include ambient and Euclidean-MDS depth, local outlier factor (LOF) (Breunig et al., 2000) in ambient and ISOMAP coordinates, Isolation Forest (Liu et al., 2008), one-class SVM (Schölkopf et al., 2001), ECOD (Li et al., 2023), and Deep SVDD (Ruff et al., 2018). Latent empirical depth is a diagnostic reference. LOF uses $k = 20$, and the Deep SVDD entry averages three fixed initializations. We report average precision (AP) and area under the ROC curve (AUROC), with AP as the main metric because the anomaly prevalence is about 3%.

6.2 Geometry recovery and depth error

Figure 1b examines the two depth-error terms under the balanced radius. Median total depth error decreased from 0.0195 at $n = 250$ to 0.00647 at $n = 2500$. Over the same range, the geometry-

induced depth perturbation E_{T_2} decreased from 0.00587 to 0.000281, while the empirical-depth sampling term E_{T_1} decreased from 0.0182 to 0.00646.

At $n = 250$, E_{T_2} contributes materially to the total error. By $n = 2500$, it is small relative to E_{T_1}, and E_D is nearly identical to the sampling term. The decline in E_{T_2} is consistent with Theorem 3, which bounds the effect of coordinate recovery error on halfspace depth. Over these sample sizes, sampling error accounts for most of the remaining depth error.

The displayed recovery summaries condition on connected graphs. For the two-dimensional setting, 49 of 50 balanced-radius graphs were connected at $n = 250$, and all 50 were connected at each larger sample size.

6.3 Effect of graph radius on depth recovery

A connected graph does not by itself ensure accurate intrinsic distances. At $n = 2500$, the fixed-radius graph had median aligned coordinate error $\delta_{\mathrm{ISO},n} = 0.002887$ and geometry-induced depth error $E_{T_2} = 0.000267$. Under the nonlocal radius, both errors were larger: 0.124412 and 0.034273, respectively. All graphs in these two settings were connected. This comparison shows how inaccurate geometry can carry through to the depth score. Appendix D.2 reports the other sample sizes and radius rules.

6.4 Prediction on test observations

Figure 1c reports prediction error on test observations. Mean RMSE for the network trained on estimated depths fell from 0.0217 at $n = 250$ to 0.00679 at $n = 2500$. The corresponding RMSE for the network trained on population depths fell from 0.00963 to 0.00210. The population-depth fit had lower RMSE in all 199 available paired replications, and the mean gap decreased from 0.0121 to 0.00470.

The paired fits show the cost of using estimated depths as training targets in these runs. Both test errors decrease with sample size, as does their gap. One training graph was disconnected at $n = 250$, leaving 49 paired fits at that size and 50 at each larger size.

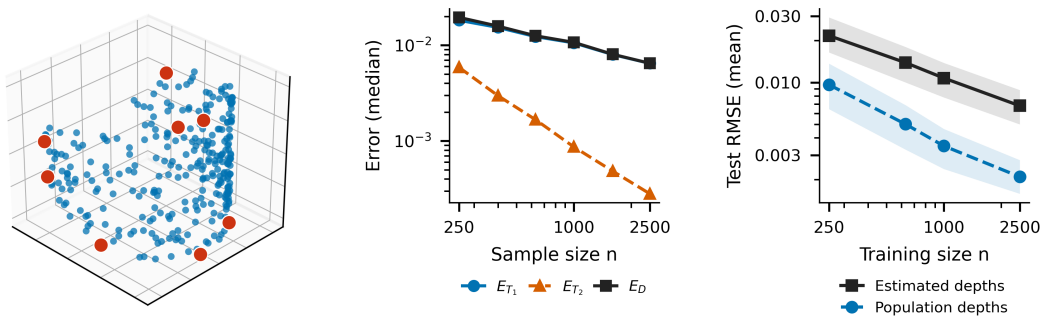
6.5 Detecting peripheral points

Table 1 evaluates points near the ends of the gapped circle and the edges of the Swiss roll. The latent empirical-depth row is a reference for the depth ranking before geometry is estimated. ISOMAP+LOF uses the same recovered coordinates as Iso-depth, so their comparison examines the effect of the score applied to that embedding.

On the gapped circle, Iso-depth had mean AP 0.9815, close to latent empirical depth at 0.9854. ISOMAP+LOF had AP 0.2368. On the Swiss roll, Iso-depth had AP 0.3402, compared with 0.1687 for ISOMAP+LOF. Latent empirical depth reached 0.5786. The gap from that reference shows that geometry recovery remains a substantial source of error in this noisy setting. The full baseline comparison and sensitivity to LOF neighborhood size and graph radius are in Appendix D.4.

Table 1: Detection of peripheral points at $n = 1000$, $p = 50$. Entries are mean (standard deviation) over 100 paired replications. Latent empirical depth is a diagnostic reference.

	Gapped circle		Swiss roll	
Method	AP	AUROC	AP	AUROC
Latent empirical depth (reference)	0.9854 (0.0191)	0.9996 (0.0006)	0.5786 (0.0760)	0.9839 (0.0042)
Iso-depth	0.9815 (0.0198)	0.9995 (0.0006)	0.3402 (0.0941)	0.9415 (0.0214)
ISOMAP+LOF ($k = 20$)	0.2368 (0.1731)	0.7264 (0.1305)	0.1687 (0.0700)	0.7915 (0.0712)
Euclidean MDS+depth	0.0188 (0.0033)	0.1608 (0.0158)	0.2553 (0.0612)	0.8562 (0.0388)
Ambient depth	0.0311 (0.0055)	0.5013 (0.0007)	0.0448 (0.0087)	0.6411 (0.0334)
Ambient LOF ($k = 20$)	0.3362 (0.1909)	0.7560 (0.1331)	0.3157 (0.0814)	0.9140 (0.0318)



(a) (b) (c)

Figure 1: Rolled-disk model, geometry recovery, and out-of-sample prediction ($p = 50$). (a) Three-dimensional PCA view of 300 rolled-disk observations, with the outer 3% latent shell highlighted. (b) Median sampling error E_{T_1}, geometry-induced error E_{T_2}, and total depth error E_D under the balanced radius. (c) Mean independent-test RMSE for networks trained on estimated depths and on population depths; bands show the 10th–90th percentiles across seeds. In (b) and (c), 49 runs are connected at $n = 250$ and 50 at each larger sample size.

6.6 Cardiotocography case study

We use the Cardiotocography dataset to examine Iso-depth on a real dataset with external pathology labels. ADBench contains heterogeneous anomaly-detection tasks, but does not identify which labeled anomalies are peripheral within a low-dimensional intrinsic geometry (Han et al., 2022). Cardio was selected for further study after exploratory comparisons among candidate datasets. This selection limits the scope of the case study and any generalization across ADBench.

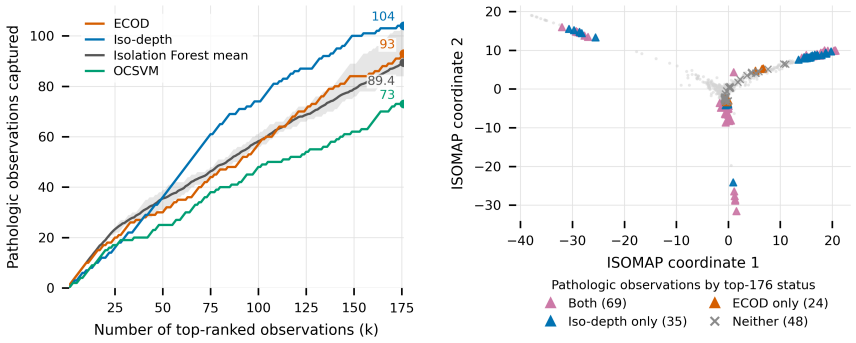
The Cardio benchmark is based on the UCI Cardiotocography records (Ayres-de Campos & Bernardes, 2000). Its 21 variables summarize fetal heart-rate and uterine-contraction recordings processed with SisPorto 2.0 (Ayres-de Campos et al., 2000). The benchmark contains 1,831 observations: 1,655 labeled normal and 176 labeled pathologic. We treat the latter as anomalies. Their prevalence, $176/1831 = 0.096$, provides a reference level for AP.

Feature scaling determines the relative contribution of each variable to the Euclidean distances used to construct the neighborhood graph. Under median centering and IQR scaling, fetal movement (FM) accounted for 94.9% of the total pairwise squared distance. Its small IQR relative to its overall dispersion gave this variable a large weight in the input metric. To limit this amplification, we placed a lower bound on the FM scale while retaining all observations and feature values. The remaining variables retained their IQR scales. This adjustment was computed from covariates alone, without pathology labels. The exact scaling rule and associated distance diagnostics are provided in Appendix E.1.

The intrinsic dimension, d, was selected without pathology labels. Among $d = 1, \ldots, 16$, we took the smallest d for which the residual variance between graph shortest-path distances and distances reconstructed from the d-dimensional MDS coordinates was at most 0.05. This quantity was 0.1312 at $d = 1$ and 0.0390 at $d = 2$. We therefore use $d = 2$ in the Cardio analysis.

We compute halfspace depth exactly in the selected two-dimensional representation. Iso-depth attained AP 0.596, AUROC 0.920, and precision 0.591 among the top 176 ranks. PCA-Depth attained AP 0.245, while LOF applied to the same ISOMAP coordinates attained AP 0.167.

Figure 2a shows the cumulative number of pathologic observations among the first k ranks. We show the curve through $k = 176$, the number of observations labeled pathologic in Cardio. After about the first 50 ranks, Iso-depth captures more pathologic observations than the baseline methods. At $k = 176$, Iso-depth identified 104, compared with 93 for ECOD. Figure 2b shows where the observations selected by these two methods fall in the learned two-dimensional coordinates. The methods select 69 of the same pathologic observations. Iso-depth alone selects 35, many near the ends of the displayed branches, while ECOD alone selects 24, mostly closer to the central junction. Comparisons with the other baselines are reported in Appendix E.1.



(a) (b)

Figure 2: Cardio dataset results. (a) Cumulative number of pathologic observations among the first k ranks. (b) Two-dimensional ISOMAP representation, with pathologic observations classified by their top-176 status under Iso-depth and ECOD.

To evaluate scores for new observations, we fitted Iso-depth-NN on the training data in 20 stratified 70/30 splits and scored the held-out observations. Its mean AP was 0.577, compared with 0.598 for ECOD and 0.589 for Isolation Forest. Among the top 53 test ranks, its mean precision was 0.566, compared with 0.547 and 0.511, respectively, where 53 is the number of pathologic observations in each test set. Further split-level results are reported in Appendix E.1.

7 Conclusion and Limitations

Iso-depth defines global peripherality through statistical depth in intrinsic coordinates and estimates it using an ISOMAP embedding. Bending or folding in the ambient space can make peripheral observations appear close to the center, motivating the use of distances along the manifold.

Our theoretical contribution is a bound that transfers RMS coordinate error to halfspace-depth error. Combined with empirical-depth and ISOMAP recovery bounds, it yields an out-of-sample error bound for the neural estimator under the stated assumptions. The near-manifold analysis extends this bound to shrinking perturbations under additional graph and network conditions.

The experiments separate the sources of error. On the rolled disk, geometry-induced depth error became small relative to sampling error over the sample sizes studied. Radius comparisons show how graph construction affects this error, while paired neural fits assess the effect of estimated depth targets on test prediction. On the gapped circle and Swiss roll, depth achieved higher average precision than LOF on the same ISOMAP embedding for detecting points near the ends and edges of the latent distribution. The Cardio study examines the resulting rankings and their sensitivity to the input metric and graph.

The error bound assumes that the manifold is globally isometric to a convex Euclidean domain. The out-of-sample result also requires exact empirical risk minimization and the stated neural approximation condition. The near-manifold guarantee concerns shrinking noise. Fixed noise remains outside its scope. In practice, errors in the neighborhood graph can distort the embedding and affect the detection of peripheral observations.

Further work is needed on selecting the intrinsic dimension and neighborhood size from data, assessing the recovered geometry, and constructing inductive representations that preserve geodesic structure. Extending the convergence analysis to projection, spatial, and simplicial depths is another direction.

AI use statement

We used generative AI tools to brainstorm ideas for the synthetic data generation process and to improve the clarity and readability of the manuscript. The authors made the final decisions about

the data-generating process, reviewed the AI-assisted edits, and take responsibility for the content of the paper.

REFERENCES

Ery Arias-Castro, Adel Javanmard, and Bruno Pelletier. Perturbation bounds for Procrustes, classical scaling, and trilateration, with applications to manifold learning. *Journal of Machine Learning Research*, 21(15):1–37, 2020. URL http://jmlr.org/papers/v21/18-720.html.

Diogo Ayres-de Campos and Joao Bernardes. Cardiotocography. UCI Machine Learning Repository, dataset, 2000. URL https://archive.ics.uci.edu/dataset/193/cardiotocography.

Diogo Ayres-de Campos, Joao Bernardes, Antonio Garrido, Joaquim Marques-de Sa, and Luis Pereira-Leite. SisPorto 2.0: A program for automated analysis of cardiotocograms. *The Journal of Maternal-Fetal Medicine*, 9(5):311–318, 2000. doi: 10.1002/1520-6661(200009/10)9:5⟨311::AID-MFM12⟩3.0.CO;2-9.

Mira Bernstein, Vin de Silva, John C. Langford, and Joshua B. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Department of Psychology, Stanford University, 2000.

Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. LOF: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pp. 93–104, 2000. doi: 10.1145/335191.335388.

Michael A. Burr and Robert J. Fabrizio. Uniform convergence rates for halfspace depth. *Statistics & Probability Letters*, 124:33–40, 2017. doi: 10.1016/j.spl.2017.01.002.

Probal Chaudhuri. On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872, 1996. doi: 10.1080/01621459.1996.10476954.

David L. Donoho and Miriam Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20(4):1803–1827, 1992. doi: 10.1214/aos/1176348890.

Lutz Dümbgen. Limit theorems for the simplicial depth. *Statistics & Probability Letters*, 14(2):119–128, 1992. doi: 10.1016/0167-7152(92)90075-G.

J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3/4):325–338, 1966. doi: 10.1093/biomet/53.3-4.325.

Songqiao Han, Xiyang Hu, Hailiang Huang, Minqi Jiang, and Yue Zhao. ADBench: Anomaly detection benchmark. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/cf93972b116ca5268827d575f2cc226b-Abstract.html.

Moritz Herrmann, Florian Pfisterer, and Fabian Scheipl. A geometric framework for outlier detection in high-dimensional data. *WIREs Data Mining and Knowledge Discovery*, 13(3):e1491, 2023. doi: 10.1002/widm.1491.

Hao Huang, Hong Qin, Shinjae Yoo, and Dantong Yu. Physics-based anomaly detection defined on manifold space. *ACM Transactions on Knowledge Discovery from Data*, 9(2):1–39, 2014. doi: 10.1145/2641574.

Zheng Li, Yue Zhao, Xiyang Hu, Nicola Botta, Cezar Ionescu, and George H. Chen. ECOD: Unsupervised outlier detection using empirical cumulative distribution functions. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12181–12193, 2023. doi: 10.1109/TKDE.2022.3159580.

Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422, 2008. doi: 10.1109/ICDM.2008.17.

Regina Y. Liu. On a notion of data depth based on random simplices. *The Annals of Statistics*, 18 (1):405–414, 1990. doi: 10.1214/aos/1176347507.

Ryumei Nakada and Masaaki Imaizumi. Adaptive approximation and generalization of deep neural network with intrinsic dimensionality. *Journal of Machine Learning Research*, 21(174):1–38, 2020. URL https://jmlr.org/papers/v21/20-002.html.

C. C. Olson, K. P. Judd, and J. M. Nichols. Manifold learning techniques for unsupervised anomaly detection. *Expert Systems with Applications*, 91:374–385, 2018. doi: 10.1016/j.eswa.2017.08.005.

Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 4393–4402. PMLR, 2018. URL https://proceedings.mlr.press/v80/ruff18a.html.

Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001. doi: 10.1162/089976601750264965.

Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000. doi: 10.1126/science.290.5500.2319.

Warren S. Torgerson. Multidimensional scaling: I. Theory and method. *Psychometrika*, 17(4):401–419, 1952. doi: 10.1007/BF02288916.

John W Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians*, volume 2, pp. 523–531, Vancouver, 1975.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York, 1996. doi: 10.1007/978-1-4757-2545-2.

Yijun Zuo. Projection-based depth functions and associated medians. *The Annals of Statistics*, 31 (5):1460–1490, 2003. doi: 10.1214/aos/1065705115.

Yijun Zuo and Robert Serfling. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482, 2000. doi: 10.1214/aos/1016218226.

A BACKGROUND

A.1 STATISTICAL DEPTH

Statistical depth measures how central a point is relative to an underlying distribution. For a probability measure P on $\mathbb{R}^d$, a depth function $D(\cdot; P)$ assigns larger values to points near the center of P and smaller values to more peripheral points, thereby inducing a center-outward ordering of the data. The general properties of depth functions, including affine invariance, maximality at a center, monotonicity away from a deepest point, and vanishing at infinity, were formulated by Zuo & Serfling (2000).

Projection depth first measures the largest standardized projected outlyingness of x and then transforms it into a depth value (Zuo, 2003). One common form is

$$D_{\mathrm{proj}}(x, P) = \big(1 + O(x, P)\big)^{-1}, \qquad O(x, P) = \sup_{u \in S^{d-1}} \frac{|u^\top x - \mathrm{med}(u^\top X)|}{\mathrm{MAD}(u^\top X)}, \tag{9}$$

where med is the median and MAD the median absolute deviation. Spatial depth is based on the average unit direction from x to the distribution (Chaudhuri, 1996),

$$D_{\mathrm{sp}}(x, P) = 1 - \left\| \mathbb{E}\left[\frac{x - X}{\|x - X\|_2} \right] \right\|_2, \tag{10}$$

and simplicial depth is the probability that x lies inside a random simplex generated by $d + 1$ independent observations (Liu, 1990),

$$D_{\mathrm{sim}}(x, P) = P\big(x \in \mathrm{conv}(X_1, \ldots, X_{d+1})\big). \tag{11}$$

For observed data $X_1, \ldots, X_n$, the sample version replaces P by the empirical measure $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$.

Low depth describes global peripherality. A point may have many nearby neighbors and still have low depth relative to the full distribution. Computing depth after recovering the intrinsic coordinates makes this centrality depend on the estimated manifold geometry.

A.2 SAMPLE-LEVEL IMPLEMENTATION

For later use in the convergence analysis, we also define the centered function class

$$\mathcal{H}_n := \left\{ D\big(\phi(\,\cdot\,), P_X \circ \phi^{-1}\big) - f : f \in \mathcal{G}_n \right\}.$$

The classes $\mathcal{F}_n$, $\mathcal{G}_n$, and $\mathcal{H}_n$ will be used in Appendix B to control, respectively, neural-network approximation, the projected Iso-depth estimator, and the associated empirical-process fluctuations.

Finally, we define the inductive anomaly score by

$$S_{\mathrm{NN}}(x) := -\hat{f}(x),$$

so that a large value of $S_{\mathrm{NN}}(x)$ corresponds to high on-manifold peripherality, or a high degree of anomaly.

The full procedure is summarized in Algorithm 1, assuming that the neighborhood graph is connected. The neural network provides the out-of-sample extension of the sample depth scores. The convergence results in Sections 3 and 4 use the intrinsic dimension d and the halfspace depth $D = D_H$.

Algorithm 1 Iso-depth with an out-of-sample extension

Require: Data matrix $X = (X_1, \dots, X_n)^\top \in \mathbb{R}^{n \times p}$, neighborhood radius $r_n > 0$, embedding dimension d, rigid-motion-invariant depth function D, and neural-network class $\mathcal{F}_n$.

Ensure: Sample anomaly scores $S_1, \dots, S_n$ and an out-of-sample scoring function S_{NN}.

In-sample Iso-depth

1: Construct the undirected r_n-neighborhood graph G_n: connect i and j if $\|X_i - X_j\|_2 \le r_n$, and assign the edge weight $\|X_i - X_j\|_2$.
2: Compute the all-pairs shortest-path distance matrix $\Gamma = (\gamma_{ij}) \in \mathbb{R}^{n \times n}$ on G_n.
3: Apply d-dimensional classical MDS to Γ to obtain $\hat{z}_1, \dots, \hat{z}_n \in \mathbb{R}^d$.
4: Form the empirical embedded distribution $\widehat{P}_n = n^{-1} \sum_{i=1}^n \delta_{\hat{z}_i}$.
5: Compute pseudo-depth labels $Y_i = D(\hat{z}_i, \widehat{P}_n)$ for $i = 1, \dots, n$.
6: Define sample anomaly scores $S_i = -Y_i$ for $i = 1, \dots, n$.

Out-of-sample extension

7: Fit

$$\tilde{f} \in \operatorname*{arg\,min}_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^n \{f(X_i) - Y_i\}^2$$

and set $\hat{f} = \Pi \circ \tilde{f}$.
8: For a new point $x \in \mathbb{R}^p$, output the inductive score $S_{\mathrm{NN}}(x) = -\hat{f}(x)$.

B Proofs and error decomposition

B.1 Notation and error decomposition

We use the network classes and centered class $\mathcal{H}_n$ defined in Appendix A.2.

The empirical-process tools below are standard: a VC maximal inequality controls empirical half-spaces, and Bernstein's inequality with a covering argument relates empirical and population risk. The ISOMAP argument applies Corollary 4 of Arias-Castro et al. (2020) with an i.i.d. covering-radius bound. We use these results to derive the depth perturbation bound and combine it with a network approximation condition. The endpoint approximation property is stated in Assumption 1; the near-manifold extension also requires the graph and risk conditions in Appendix C.

For any $f \in \mathcal{G}_n$, define population and empirical root mean square error by

$$\mathcal{E}(f) = \left\{ \mathbb{E}_X \left[(f_0(X) - f(X))^2 \right] \right\}^{1/2}, \qquad \mathcal{E}_n(f) = \left\{ n^{-1} \sum_{i=1}^n [f_0(X_i) - f(X_i)]^2 \right\}^{1/2}.$$

The expectation is over an independent $X \sim P_X$, with $\mathcal{E}(f)^2 = \int_M (f_0(x) - f(x))^2 \, dP_X(x)$. For a fitted f, the training sample is held fixed. Evaluating this functional at $\hat{f}$ gives a random conditional risk. The empirical error is used to bound this risk.

Write $Z_i = \phi(X_i)$, $\widehat{Z}_i = \hat{\phi}(X_i)$, $P_Z = P_X \circ \phi^{-1}$, $P_n^Z = n^{-1} \sum_i \delta_{Z_i}$, and $P_n^{\widehat{Z}} = n^{-1} \sum_i \delta_{\widehat{Z}_i}$. With $\|v\|_{n,2} = (n^{-1} \sum_i v_i^2)^{1/2}$, define

$$\begin{aligned} T_{1,i} &= D(Z_i, P_Z) - D(Z_i, P_n^Z), \\ T_{2,i} &= D(Z_i, P_n^Z) - D(\widehat{Z}_i, P_n^{\widehat{Z}}), \\ T_{3,i} &= D(\widehat{Z}_i, P_n^{\widehat{Z}}) - \hat{f}(X_i). \end{aligned}$$

Their sum equals $f_0(X_i) - \hat{f}(X_i)$, so the triangle inequality gives

$$\mathcal{E}_n(\hat{f}) \le \|T_1\|_{n,2} + \|T_2\|_{n,2} + \|T_3\|_{n,2}. \tag{12}$$

Step 1 bounds population error using empirical error. Step 2 controls the three terms, and Step 3 combines the bounds.

B.2 STEP 1: RELATION BETWEEN IN-SAMPLE AND OUT-OF-SAMPLE ERRORS

Let

$$h_f(x) := D(\phi(x), P_X \circ \phi^{-1}) - f(x). \tag{13}$$

We first establish a probabilistic inequality relating the in-sample and out-of-sample errors.

Lemma 5 *Suppose*

1. $L_n = O(1)$, $B_n = O(W_n^s)$, $W_n = O(n^r)$ *for some* $s, r > 0$;
2. $L_n \geq 1$, $B_n \geq 1$ *and* $W_n \geq 1$.

Let

$$\Lambda_n = W_n \log\bigl(2L_n B_n^{L_n}(W_n+1)^{L_n} n\bigr).$$

Then, there is a constant $C > 0$ *such that, for every* $u > 0$,

$$P\left(\mathbb{E}[h_f^2(X)] \leq \frac{2}{n}\sum_{i=1}^n h_f^2(X_i) + C\frac{\Lambda_n + u}{n} \quad \forall\, h_f \in \mathcal{H}_n\right) \geq 1 - e^{-u}. \tag{14}$$

$\Lambda_n = O(W_n \log n)$.

Proof Since both terms on the right-hand side of (13) take values in $[0,1]$, we have $|h_f| \leq 1$, and therefore

$$\mathrm{Var}\bigl\{h_f^2(X)\bigr\} \leq \mathbb{E}h_f^4(X) \leq \mathbb{E}h_f^2(X).$$

By Bernstein's inequality and the inequality $\sqrt{2ab} \leq a/2 + b$, we have, for each fixed $f \in \mathcal{G}_n$,

$$P\left(\mathbb{E}h_f^2(X) \leq \frac{2}{n}\sum_{i=1}^n h_f^2(X_i) + \frac{8v}{3n}\right) \geq 1 - e^{-v}.$$

We next extend this bound uniformly over $\mathcal{G}_n$. Since the truncation map Π is 1-Lipschitz,

$$N(\varepsilon, \mathcal{G}_n, \|\cdot\|_\infty) \leq N(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty).$$

Since the map $f \mapsto h_f$ is isometric under the sup norm, we have

$$N(\varepsilon, \mathcal{H}_n, \|\cdot\|_\infty) = N(\varepsilon, \mathcal{G}_n, \|\cdot\|_\infty).$$

by Lemma 21 of Nakada & Imaizumi (2020),

$$\log N(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) \leq W_n \log\bigl(2L_n B_n^{L_n}(W_n+1)^{L_n}/\varepsilon\bigr).$$

Thus we have

$$\log N(\varepsilon, \mathcal{H}_n, \|\cdot\|_\infty) \leq W_n \log\bigl(2L_n B_n^{L_n}(W_n+1)^{L_n}/\varepsilon\bigr). \tag{15}$$

Let $\{h_1, \ldots, h_N\}$ be an ε-net of $\mathcal{H}_n$, where $N = N(\varepsilon, \mathcal{H}_n, \|\cdot\|_\infty)$. Applying the Bernstein bound to each h_j with $v = u + \log N$ and taking a union bound, we obtain

$$P\left(\mathbb{E}h_j^2(X) \leq \frac{2}{n}\sum_{i=1}^n h_j^2(X_i) + \frac{8(u+\log N)}{3n},\ j = 1, \ldots, N\right) \geq 1 - e^{-u}.$$

For any $h_f \in \mathcal{H}_n$, choose h_j in the net such that $\|h_f - h_j\|_\infty \leq \varepsilon$. Since $|h_f|, |h_j| \leq 1$,

$$|h_f^2 - h_j^2| = |h_f - h_j|\,|h_f + h_j| \leq 2\varepsilon.$$

Hence

$$\mathbb{E}h_f^2(X) \leq \mathbb{E}h_j^2(X) + 2\varepsilon$$

and

$$\frac{2}{n}\sum_{i=1}^n h_j^2(X_i) \leq \frac{2}{n}\sum_{i=1}^n h_f^2(X_i) + 4\varepsilon.$$

Combining these inequalities gives

$$P\left(\mathbb{E}h_f^2(X) \le \frac{2}{n}\sum_{i=1}^{n} h_f^2(X_i) + 6\varepsilon + \frac{8(u+\log N)}{3n},\ \forall h_f \in \mathcal{H}_n\right) \ge 1 - e^{-u}.$$

Take $\varepsilon = 1/n$. By (15),

$$\log N \le W_n \log\bigl(2L_n B_n^{L_n}(W_n+1)^{L_n} n\bigr) = \Lambda_n.$$

Therefore,

$$P\left(\mathbb{E}h_f^2(X) \le \frac{2}{n}\sum_{i=1}^{n} h_f^2(X_i) + 6\varepsilon + \frac{8(u+\Lambda_n)}{3n},\ \forall h_f \in \mathcal{H}_n\right) \ge 1 - e^{-u}. \tag{16}$$

Since $L_n \ge 1$, $B_n \ge 1$, and $W_n \ge 1$, we have

$$2L_n B_n^{L_n}(W_n+1)^{L_n} n \ge 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 = 4.$$

Therefore,

$$\Lambda_n = W_n \log\bigl(2L_n B_n^{L_n}(W_n+1)^{L_n} n\bigr) \ge \log 4 > 0.$$

Consequently,

$$\begin{aligned}\frac{6}{n} + \frac{8(\Lambda_n+u)}{3n} &\le \frac{6}{\log 4}\frac{\Lambda_n}{n} + \frac{8}{3}\frac{\Lambda_n+u}{n} \\ &\le \left(\frac{6}{\log 4} + \frac{8}{3}\right)\frac{\Lambda_n+u}{n} \\ &= C\frac{\Lambda_n+u}{n},\end{aligned}$$

where $C = 6/\log 4 + 8/3$ is a positive constant independent of n and u. The above inequalities, combined with (16), imply the desired inequality (14).

To show $\Lambda_n = O(W_n \log n)$, we note that

$$\begin{aligned}\Lambda_n &= W_n \log\bigl(2L_n B_n^{L_n}(W_n+1)^{L_n} n\bigr) \\ &= W_n \left\{\log(2L_n) + L_n \log B_n + L_n \log(W_n+1) + \log n\right\}.\end{aligned}$$

Since $L_n = O(1)$ and $B_n = O(W_n^s)$ for some fixed $s > 0$,

$$\log B_n = O(\log W_n).$$

since $W_n = O(n^r)$ for some $r > 0$, we have $\log W_n = O(\log n)$. It follows that

$$\log\bigl(2L_n B_n^{L_n}(W_n+1)^{L_n} n\bigr) = O(\log n),$$

and therefore $\Lambda_n = O(W_n \log n)$. ■

Relation equation 14 is a bound on squared risks:

$$P\left(\mathcal{E}(f)^2 \le 2\mathcal{E}_n(f)^2 + C\frac{\Lambda_n+u}{n} \quad \forall f \in \mathcal{G}_n\right) \ge 1 - e^{-u}. \tag{17}$$

Taking square roots gives $\mathcal{E}(\hat{f}) \le \sqrt{2}\mathcal{E}_n(\hat{f}) + O_p\{(W_n \log n/n)^{1/2}\}$. We denote this stochastic remainder by $\eta_{\mathrm{gen},n}$. There is no assumption that it is smaller than the realized empirical error. For the chosen W_n, it has the same order as $W_n^{-1/d}$ and yields $(\log n/n)^{1/(d+2)}$ in the final bound.

We first record a regularity property of halfspace depth that will be used throughout the subsequent analysis. Let

$$Z = \phi(X), \qquad P_Z = P_X \circ \phi^{-1},$$

and, for $u \in \mathbb{S}^{d-1}$, let f_u and F_u denote, respectively, the density and distribution function of $u^\top Z$.

Lemma 6 (Lipschitz continuity of halfspace depth) *Suppose that every one-dimensional projection $u^\top Z$, $u \in \mathbb{S}^{d-1}$, has a density f_u satisfying*

$$\sup_{u \in \mathbb{S}^{d-1}} \|f_u\|_\infty \le M$$

for some $M < \infty$. Then

$$|D_H(z, P_Z) - D_H(z', P_Z)| \le M\|z - z'\|_2, \qquad z, z' \in \mathbb{R}^d.$$

In particular, $D_H(\cdot, P_Z)$ is Lipschitz continuous.

Proof Recall that

$$D_H(z, P_Z) = \inf_{u \in \mathbb{S}^{d-1}} F_u(u^\top z).$$

Since $\|f_u\|_\infty \le M$ uniformly in u,

$$|F_u(s) - F_u(t)| \le M|s - t|, \qquad s, t \in \mathbb{R}.$$

Using the elementary inequality $|\inf_u a_u - \inf_u b_u| \le \sup_u |a_u - b_u|$, we obtain

$$\begin{aligned} |D_H(z, P_Z) - D_H(z', P_Z)| &\le \sup_{u \in \mathbb{S}^{d-1}} |F_u(u^\top z) - F_u(u^\top z')| \\ &\le M \sup_{u \in \mathbb{S}^{d-1}} |u^\top (z - z')| \\ &= M\|z - z'\|_2. \end{aligned}$$

This proves the result. ∎

Theorem 7 *Suppose the conditions of Lemma 5 and Lemma 6 hold. If D is the halfspace depth and*

$$W_n \asymp \left(\frac{n}{\log n}\right)^{d/(d+2)},$$

then

$$\mathcal{E}(\hat{f}) = O_p\left\{\mathcal{E}_n(\hat{f}) + \left(\frac{\log n}{n}\right)^{1/(d+2)}\right\}.$$

Proof By Lemma 5, since $\hat{f} \in \mathcal{G}_n$, for every $u > 0$, with probability at least $1 - e^{-u}$,

$$\mathcal{E}(\hat{f})^2 \le 2\mathcal{E}_n(\hat{f})^2 + C\frac{\Lambda_n + u}{n}.$$

Since $\Lambda_n = O(W_n \log n)$, it follows that

$$\mathcal{E}(\hat{f})^2 = O_p\left\{\mathcal{E}_n(\hat{f})^2 + \frac{W_n \log n}{n}\right\}.$$

Substituting the prescribed $W_n \asymp (n/\log n)^{d/(d+2)}$ gives the claimed stochastic term. Under this choice,

$$W_n^{-2/d} \asymp \frac{W_n \log n}{n} \asymp \left(\frac{\log n}{n}\right)^{2/(d+2)}.$$

Hence

$$\mathcal{E}(\hat{f})^2 = O_p\left\{\mathcal{E}_n(\hat{f})^2 + \left(\frac{\log n}{n}\right)^{2/(d+2)}\right\}.$$

Taking square roots yields

$$\mathcal{E}(\hat{f}) = O_p\left\{\mathcal{E}_n(\hat{f}) + \left(\frac{\log n}{n}\right)^{1/(d+2)}\right\},$$

which proves the result. ∎

B.3 STEP 2(A): EMPIRICAL DISTRIBUTION ERROR

The first term, T_1, arises from replacing the population distribution P_X with the empirical distribution P_n, while keeping the embedding ϕ fixed at true embedding. Since this error is solely due to the discrepancy between P_X and P_n, its convergence rate is governed by Glivenko-Cantelli-type asymptotic properties of the empirical measure. Under suitable regularity conditions, this rate is typically $n^{-1/2}$. Such results are available in the literature for a variety of depth functions under different sets of assumptions. In what follows, we focus on four representative depth functions: the halfspace (Tukey) depth (Burr & Fabrizio, 2017), the projection depth (Zuo, 2003), the spatial depth (Chaudhuri, 1996), and the simplicial depth (Dümbgen, 1992).

Since

$$\|T_1\|_{n,2} \le \sup_{z\in S}\big|D(z,\, P_X\circ\phi^{-1}) - D(z,\, P_n\circ\phi^{-1})\big|,$$

it suffices to establish the convergence rate of the right-hand side. In the following, we derive the convergence result for halfspace depth, while stating without proof the corresponding rates for the projection, spatial and simplicial depths.

Let $Z_1,\dots,Z_n$ be an i.i.d. sample from P_Z, a distribution on $\mathbb{R}^d$, and let $P_n^Z = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$ be the empirical distribution based on $Z_1,\dots,Z_n$. For $u\in\mathbb{S}^{d-1}$ and $t\in\mathbb{R}$, define

$$F_u(t) = P_Z(u^\top Z \le t), \qquad F_{n,u}(t) = P_n^Z(u^\top Z \le t),$$

and let

$$R_n := \sup_{(u,t)\in\mathbb{S}^{d-1}\times\mathbb{R}} |F_{n,u}(t) - F_u(t)|\,.$$

We first prove a lemma concerning the rate of R_n.

Lemma 8 (Uniform empirical halfspace approximation) *If $Z_1,\dots,Z_n$ be i.i.d. random vectors in $\mathbb{R}^d$ with distribution P_Z, where d is fixed, then*

$$R_n = O_p(n^{-1/2}).$$

Proof Let

$$\mathcal{H}_d = \big\{H_{u,t} = \{z\in\mathbb{R}^d : u^\top z \le t\} : u\in\mathbb{S}^{d-1},\ t\in\mathbb{R}\big\}$$

denote the class of affine halfspaces in $\mathbb{R}^d$. Then

$$R_n = \sup_{H\in\mathcal{H}_d} \big|P_n^Z(H) - P_Z(H)\big|\,.$$

Since $\mathcal{H}_d$ is a VC class with VC-dimension $d+1$, by the maximal inequality for VC classes (van der Vaart & Wellner (1996)) we have

$$\mathbb{E}\left[\sup_{H\in\mathcal{H}_d} \big|\sqrt{n}\{P_n^Z(H) - P_Z(H)\}\big|\right] \le C\sqrt{d+1},$$

where $C>0$ is a universal constant, which implies $\mathbb{E}R_n = O(n^{-1/2})$. Hence, by Markov's inequality, $R_n = O_p(n^{-1/2})$. ∎

Proof Let $Z=\phi(X)$ and $Z_i=\phi(X_i)$, $i=1,\dots,n$. Recall that the population and empirical halfspace depths can be written as

$$D\big(z, P_X\circ\phi^{-1}\big) = \inf_{u\in\mathbb{S}^{d-1}} F_u(u^\top z), \qquad D\big(z, P_n\circ\phi^{-1}\big) = \inf_{u\in\mathbb{S}^{d-1}} F_{n,u}(u^\top z).$$

Using the elementary inequality $|\inf_u a_u - \inf_u b_u| \le \sup_u |a_u - b_u|$, we obtain, for every $z\in S$,

$$\begin{aligned}
&\big|D\big(z, P_X\circ\phi^{-1}\big) - D\big(z, P_n\circ\phi^{-1}\big)\big| \\
&\quad\le \sup_{u\in\mathbb{S}^{d-1}} \big|F_u(u^\top z) - F_{n,u}(u^\top z)\big| \\
&\quad\le \sup_{\substack{u\in\mathbb{S}^{d-1}\\ t\in\mathbb{R}}} |F_u(t) - F_{n,u}(t)| = R_n.
\end{aligned}$$

Since the bound holds for every $z \in S$,

$$\sup_{z \in S} \left| D\big(z, P_X \circ \phi^{-1}\big) - D\big(z, P_n \circ \phi^{-1}\big) \right| \le R_n.$$

By Lemma 8, $R_n = O_p(n^{-1/2})$. Therefore,

$$\sup_{z \in S} \left| D\big(z, P_X \circ \phi^{-1}\big) - D\big(z, P_n \circ \phi^{-1}\big) \right| = O_p(n^{-1/2}),$$

as desired. ■

B.4 Step 2(B): error due to ISOMAP

Proof We first treat $x_1, \dots, x_n$ as fixed and describe the deterministic bound established by Arias-Castro et al. (2020). Let

$$Y_i = \phi(x_i) - \bar{\phi}_n, \qquad \bar{\phi}_n = \frac{1}{n} \sum_{i=1}^{n} \phi(x_i),$$

be the centered true embedding of the sample, and let

$$\rho_n = \max_{1 \le i \le n} \|Y_i\|_2, \qquad \omega_n = \left\{ \lambda_{\min} \left(n^{-1} \sum_{i=1}^{n} Y_i Y_i^\top \right) \right\}^{1/2}, \qquad a_n = \sup_{x \in M} \min_{1 \le i \le n} d_M(x, x_i).$$

The quantities ρ_n, ω_n, and a_n are called the max-radius, half-width, and geodesic covering radius, respectively. Here a_n denotes sample coverage. The generalization remainder is $\eta_{\mathrm{gen},n}$, defined after equation 17. Coverage enters through $(a_n/r_n)^2$ and contributes to the ISOMAP and depth terms of the final rate. Let r_n be the neighborhood radius used to construct the r_n-neighborhood graph and let

$$\xi_n = c_0 \left(\frac{r_n}{\tau} \right)^2 \vee c_1 \left(\frac{a_n}{r_n} \right)^2,$$

where c_0 and c_1 are the universal constants in the graph-distance approximation bounds of Arias-Castro et al. (2020). By Corollary 4 of that paper, if

$$\xi_n \le \frac{1}{24} \left(\frac{\rho_n}{\omega_n} \right)^{-2}, \tag{18}$$

then

$$\min_{Q \in O(d)} \left(\frac{1}{n} \sum_{i=1}^{n} \|\hat{\phi}(x_i) - Q Y_i\|_2^2 \right)^{1/2} \le 36 \sqrt{d}\, \frac{\rho_n^3}{\omega_n^2}\, \xi_n. \tag{19}$$

Next, we treat $X_1, \dots, X_n$ as i.i.d. random variables and develop a stochastic counterpart of (19). We first show that (19) holds with probability tending to one.

To this end, we derive the stochastic order of the geodesic covering radius a_n, which is now random. Since ϕ is an isometry,

$$a_n = \sup_{z \in S} \min_{1 \le i \le n} \|z - \phi(X_i)\|_2.$$

Thus, a_n is the covering radius of the embedded sample $\{\phi(X_1), \dots, \phi(X_n)\}$ in S. Since S is compact, there exists a constant $C > 0$ such that, for every sufficiently small $t > 0$, S admits a t-net $\{z_1, \dots, z_{N_t}\}$ satisfying $N_t \le C t^{-d}$. because S is convex, compact, and has nonempty interior, there exists $c > 0$ such that, for every $z \in S$ and all sufficiently small $t > 0$,

$$\mathrm{Leb}\{S \cap B(z,t)\} \ge c t^d,$$

where $\mathrm{Leb}(\cdots)$ is the Lebesgue measure. Since $p_Z(z) \ge p_{\min} > 0$ on S, there exists $c_1 > 0$ such that

$$P_Z\{B(z,t)\} \ge c_1 t^d$$

uniformly in $z \in S$ for all sufficiently small $t > 0$. If every ball $B(z_j, t)$ contains at least one embedded sample point, then every $z \in S$ is within distance $2t$ of some $\phi(X_i)$, and hence $a_n \le 2t$. Therefore,

$$P(a_n > 2t) \le Ct^{-d}(1 - c_1 t^d)^n \le Ct^{-d} e^{-c_1 n t^d}.$$

Taking $t_n = A(\log n/n)^{1/d}$ with A sufficiently large gives

$$P(a_n > 2t_n) \le \frac{Cn}{A^d \log n}\, n^{-c_1 A^d} \longrightarrow 0,$$

and hence

$$a_n = O_p\left\{ \left(\frac{\log n}{n} \right)^{1/d} \right\}.$$

Next, since $r_n \asymp (\log n/n)^{1/(2d)}$, we have

$$\left(\frac{r_n}{\tau}\right)^2 = O\left\{ \left(\frac{\log n}{n} \right)^{1/d} \right\}, \qquad \left(\frac{a_n}{r_n}\right)^2 = O_p\left\{ \left(\frac{\log n}{n} \right)^{1/d} \right\},$$

and therefore

$$\xi_n = O_p\left\{ \left(\frac{\log n}{n} \right)^{1/d} \right\} = o_p(1).$$

It remains to control ρ_n and ω_n. Since S is compact, $\rho_n = O_p(1)$. To bound ω_n away from zero, write $\Sigma_Z = \mathrm{Cov}(Z)$. The positive density on a set with nonempty interior implies $\lambda_{\min}(\Sigma_Z) > 0$: a zero variance direction would confine the distribution to a hyperplane. Compact support and fixed d give $n^{-1}\sum_i Y_i Y_i^\top \to_p \Sigma_Z$ entrywise and in operator norm. Hence $\omega_n \to_p \sqrt{\lambda_{\min}(\Sigma_Z)} > 0$, so $\omega_n^{-1} = O_p(1)$. Therefore, $\rho_n/\omega_n = O_p(1)$, and since $\xi_n = o_p(1)$,

$$\xi_n \left(\frac{\rho_n}{\omega_n}\right)^2 = o_p(1).$$

Thus, (18) holds with probability tending to one, and consequently so does (19).

Finally, since $\rho_n^3/\omega_n^2 = O_p(1)$,

$$36\sqrt{d}\,\frac{\rho_n^3}{\omega_n^2}\,\xi_n = O_p\left\{ \left(\frac{\log n}{n} \right)^{1/d} \right\}.$$

Thus,

$$\min_{Q \in O(d)} \left(\frac{1}{n} \sum_{i=1}^n \|\hat{\phi}(X_i) - Q\{\phi(X_i) - \bar{\phi}_n\}\|_2^2 \right)^{1/2} = O_p\left\{ \left(\frac{\log n}{n} \right)^{1/d} \right\}.$$

Since the centering $\bar{\phi}_n$ can be absorbed into the translation component of a rigid transformation, this is equivalent to (5). ∎

Proof Let $T_n \in \mathcal{A}(d)$ attain the minimum in the definition of $\delta_{\mathrm{ISO},n}$, and set

$$Z_i = \phi(X_i), \qquad \tilde{Z}_i = T_n^{-1}\{\hat{\phi}(X_i)\}, \qquad e_i = \|\tilde{Z}_i - Z_i\|_2, \qquad i = 1, \ldots, n.$$

Since T_n is a rigid transformation, $e_i = \|\hat{\phi}(X_i) - (T_n \circ \phi)(X_i)\|_2$, and hence

$$\delta_{\mathrm{ISO},n} = \left(\frac{1}{n} \sum_{i=1}^n e_i^2 \right)^{1/2}.$$

Let

$$P_n^Z = P_n \circ \phi^{-1} = \frac{1}{n}\sum_{j=1}^n \delta_{Z_j}, \qquad P_n^{\tilde{Z}} = \frac{1}{n}\sum_{j=1}^n \delta_{\tilde{Z}_j}.$$

By invariance of the halfspace depth under rigid transformations,

$$D_H\left(\hat{\phi}(X_i), P_n \circ \hat{\phi}^{-1}\right) = D_H(\tilde{Z}_i, P_n^{\tilde{Z}}),$$

so

$$T_{2,i} = D_H(Z_i, P_n^Z) - D_H(\tilde{Z}_i, P_n^{\tilde{Z}}).$$

For each i, decompose $|T_{2,i}| \le A_i + B_i + C_i$, where

$$A_i := \left|D_H(Z_i, P_n^Z) - D_H(Z_i, P_Z)\right|,$$
$$B_i := \left|D_H(Z_i, P_Z) - D_H(\tilde{Z}_i, P_Z)\right|,$$
$$C_i := \left|D_H(\tilde{Z}_i, P_Z) - D_H(\tilde{Z}_i, P_n^{\tilde{Z}})\right|.$$

Writing $A = (A_1, \dots, A_n)^\top$, and similarly for B and C, the triangle inequality gives

$$\|T_2\|_{n,2} \le \|A\|_{n,2} + \|B\|_{n,2} + \|C\|_{n,2}.$$

We first bound $\|A\|_{n,2}$. By Theorem 1 and Lemma 8,

$$\sup_z \left|D_H(z, P_n^Z) - D_H(z, P_Z)\right| \le R_n = O_p(n^{-1/2}),$$

and therefore $\|A\|_{n,2} \le R_n = O_p(n^{-1/2})$.

We next bound $\|B\|_{n,2}$. By Lemma 6,

$$B_i = \left|D_H(Z_i, P_Z) - D_H(\tilde{Z}_i, P_Z)\right| \le M\|Z_i - \tilde{Z}_i\|_2 = Me_i.$$

Consequently,

$$\|B\|_{n,2} \le M\left(\frac{1}{n}\sum_{i=1}^n e_i^2\right)^{1/2} = M\delta_{\mathrm{ISO},n} = O_p(\delta_{\mathrm{ISO},n}).$$

It remains to bound $\|C\|_{n,2}$. Let $F_{n,u}$ be as defined in Lemma 8, and define

$$\tilde{F}_{n,u}(t) = \frac{1}{n}\sum_{j=1}^n \mathbf{1}\{u^\top \tilde{Z}_j \le t\}.$$

For any $\varepsilon > 0$, if $\mathbf{1}\{u^\top Z_j \le t\} \ne \mathbf{1}\{u^\top \tilde{Z}_j \le t\}$, then either $|u^\top Z_j - t| \le \varepsilon$ or $e_j > \varepsilon$. Indeed, if $e_j \le \varepsilon$, then

$$|u^\top(\tilde{Z}_j - Z_j)| \le \|\tilde{Z}_j - Z_j\|_2 \le \varepsilon,$$

so the halfspace indicator can change only when Z_j lies within an ε-slab around the boundary hyperplane $\{z : u^\top z = t\}$. Therefore,

$$|F_{n,u}(t) - \tilde{F}_{n,u}(t)| \le P_n^Z\{|u^\top Z - t| \le \varepsilon\} + \frac{1}{n}\sum_{j=1}^n \mathbf{1}\{e_j > \varepsilon\}.$$

For the first term, letting F_u denote the distribution function of $u^\top Z$, the definition of R_n and the density bound give

$$P_n^Z\{|u^\top Z - t| \le \varepsilon\} \le F_u(t+\varepsilon) - F_u(t-\varepsilon) + 2R_n \le 2M\varepsilon + 2R_n.$$

For the second term,

$$\frac{1}{n}\sum_{j=1}^n \mathbf{1}\{e_j > \varepsilon\} \le \frac{1}{n\varepsilon^2}\sum_{j=1}^n e_j^2 = \frac{\delta_{\mathrm{ISO},n}^2}{\varepsilon^2}.$$

Hence

$$\sup_{u,t} |F_{n,u}(t) - \tilde{F}_{n,u}(t)| \le 2M\varepsilon + 2R_n + \frac{\delta_{\mathrm{ISO},n}^2}{\varepsilon^2}.$$

Taking $\varepsilon = \delta_{\mathrm{ISO},n}^{2/3}$ and using $R_n = O_p(n^{-1/2})$ gives

$$\sup_{u,t} |F_{n,u}(t) - \tilde{F}_{n,u}(t)| = O_p\Big(n^{-1/2} + \delta_{\mathrm{ISO},n}^{2/3}\Big).$$

Therefore, using $|\inf_u a_u - \inf_u b_u| \le \sup_u |a_u - b_u|$,

$$\sup_z \Big|D_H(z, P_n^Z) - D_H(z, P_n^{\tilde{Z}})\Big| = O_p\Big(n^{-1/2} + \delta_{\mathrm{ISO},n}^{2/3}\Big).$$

Now, by the triangle inequality,

$$\begin{aligned}\Big|D_H(z, P_Z) - D_H(z, P_n^{\tilde{Z}})\Big| \le \big|D_H(z, P_Z) - D_H(z, P_n^Z)\big| \\ + \Big|D_H(z, P_n^Z) - D_H(z, P_n^{\tilde{Z}})\Big|.\end{aligned}$$

Using Theorem 1 and the preceding bound,

$$\sup_z \Big|D_H(z, P_Z) - D_H(z, P_n^{\tilde{Z}})\Big| = O_p\Big(n^{-1/2} + \delta_{\mathrm{ISO},n}^{2/3}\Big).$$

Since C_i is this difference evaluated at $z = \tilde{Z}_i$, it follows that

$$\|C\|_{n,2} = O_p\Big(n^{-1/2} + \delta_{\mathrm{ISO},n}^{2/3}\Big).$$

Combining the three bounds,

$$\|T_2\|_{n,2} = O_p\Big(n^{-1/2} + \delta_{\mathrm{ISO},n} + \delta_{\mathrm{ISO},n}^{2/3}\Big).$$

Since $\delta_{\mathrm{ISO},n} = o_p(1)$, we have $\delta_{\mathrm{ISO},n} = O_p(\delta_{\mathrm{ISO},n}^{2/3})$, and therefore

$$\|T_2\|_{n,2} = O_p\Big(n^{-1/2} + \delta_{\mathrm{ISO},n}^{2/3}\Big).$$

Finally, Theorem 2 yields $\delta_{\mathrm{ISO},n} = O_p\{(\log n/n)^{1/d}\}$, which implies

$$\|T_2\|_{n,2} = O_p\left\{n^{-1/2} + \left(\frac{\log n}{n}\right)^{2/(3d)}\right\},$$

as desired. ∎

B.5 Step 2(c): In-sample error due to deep learning approximation

We next control the neural-network approximation error. Since $Y_i \in [0,1]$ and projection onto $[0,1]$ cannot increase the distance to Y_i,

$$\|T_3\|_{n,2} = \|Y - \hat{f}\|_{n,2} \le \|Y - \tilde{f}\|_{n,2} = \inf_{f \in \mathcal{F}_n} \|Y - f\|_{n,2}.$$

Let $P_Z = P_X \circ \phi^{-1}$ and define

$$\delta_{\mathrm{NN},n} := \inf_{f \in \mathcal{F}_n} \left[\frac{1}{n}\sum_{i=1}^n \big\{D_H\big(\phi(X_i), P_Z\big) - f(X_i)\big\}^2\right]^{1/2}.$$

By the triangle inequality,

$$\|T_3\|_{n,2} \le \delta_{\mathrm{NN},n} + \|T_1\|_{n,2} + \|T_2\|_{n,2}. \tag{20}$$

It therefore remains to control $\delta_{\mathrm{NN},n}$.

We first record the geometric covering property needed for the neural-network approximation. For $A \subset \mathbb{R}^p$, let $N(A, \rho, \|\cdot\|_\infty)$ denote the minimum number of $\|\cdot\|_\infty$-balls of radius ρ required to cover A.

Lemma 9 *Suppose that $M \subset \mathbb{R}^p$ is a compact d-dimensional C^2 embedded submanifold with positive reach. Then there exists a constant $C_M > 0$ such that*

$$N\big(\mathrm{Supp}(P_X), \rho, \|\cdot\|_\infty\big) \le C_M \rho^{-d}, \qquad 0 < \rho \le 1.$$

Proof Since $\mathrm{Supp}(P_X) \subseteq M$, it suffices to cover M. By the standard covering-number bound for compact smooth manifolds with positive reach,

$$N\big(M, \rho, \|\cdot\|_2\big) \le C_M \rho^{-d}.$$

Since $\|x\|_\infty \le \|x\|_2$, every Euclidean ball of radius ρ is contained in an $\|\cdot\|_\infty$-ball of the same radius. Hence

$$N\big(\mathrm{Supp}(P_X), \rho, \|\cdot\|_\infty\big) \le N\big(M, \rho, \|\cdot\|_\infty\big) \le N\big(M, \rho, \|\cdot\|_2\big) \le C_M \rho^{-d}.$$

■

After a fixed coordinate-wise affine rescaling of the network input, we may assume without loss of generality that $M \subset [0,1]^p$. We state the network approximation property needed for the rate.

Assumption 1 (Network approximation) *The chosen network classes satisfy*

$$\inf_{f \in \mathcal{F}_n} \sup_{x \in \mathrm{Supp}(P_X)} |f_0(x) - f(x)| = O(W_n^{-1/d}).$$

Their depths and weight bounds also satisfy the complexity conditions in Lemma 5.

Theorem 5 of Nakada & Imaizumi (2020) motivates this approximation condition, but its stated dimension is strictly larger than the Minkowski dimension. Applying its construction at the endpoint d, with the same depth and weight constraints used here, requires a separate verification. We therefore state the approximation property explicitly as a condition of the out-of-sample rate. Under this condition, $\delta_{\mathrm{NN},n} = O(W_n^{-1/d})$.

Lemma 6 gives Lipschitz continuity in latent coordinates. In particular, $D_H(\cdot, P_Z)$ is Lipschitz on $\mathbb{R}^d$. Consequently, provided that $x \mapsto \phi(x)$ is Lipschitz on M with respect to the ambient Euclidean metric,

$$x \longmapsto D_H\big(\phi(x), P_Z\big)$$

is Lipschitz on M and therefore admits a Lipschitz extension to $[0,1]^p$.

Corollary 10 *Under the conditions of Assumption 1 and the conditions of Steps 2(A) and 2(B) for halfspace depth,*

$$\|T_3\|_{n,2} = O_p\left(W_n^{-1/d} + n^{-1/2} + \delta_{\mathrm{ISO},n}^{2/3}\right).$$

Consequently, if

$$\delta_{\mathrm{ISO},n} = O_p\left\{\left(\frac{\log n}{n}\right)^{1/d}\right\},$$

then

$$\|T_3\|_{n,2} = O_p\left\{W_n^{-1/d} + n^{-1/2} + \left(\frac{\log n}{n}\right)^{2/(3d)}\right\}.$$

Proof By equation 20 and Assumption 1,

$$\|T_3\|_{n,2} \le \delta_{\mathrm{NN},n} + \|T_1\|_{n,2} + \|T_2\|_{n,2}, \qquad \delta_{\mathrm{NN},n} = O(W_n^{-1/d}).$$

For halfspace depth, Steps 2(A) and 2(B) give

$$\|T_1\|_{n,2} = O_p(n^{-1/2}), \qquad \|T_2\|_{n,2} = O_p\left(n^{-1/2} + \delta_{\mathrm{ISO},n}^{2/3}\right).$$

Combining these bounds proves the first assertion. The second follows immediately from

$$\delta_{\mathrm{ISO},n}^{2/3} = O_p\left\{\left(\frac{\log n}{n}\right)^{2/(3d)}\right\}.$$

■

B.6 Step 3: Overall out-of-sample error bound

Proof By the decomposition in Step 2,

$$\mathcal{E}_n(\hat{f}) \le \|T_1\|_{n,2} + \|T_2\|_{n,2} + \|T_3\|_{n,2}.$$

For halfspace depth, Steps 2(A)–2(C) give

$$\|T_1\|_{n,2} = O_p(n^{-1/2}),$$
$$\|T_2\|_{n,2} = O_p\left\{n^{-1/2} + \left(\frac{\log n}{n}\right)^{2/(3d)}\right\},$$
$$\|T_3\|_{n,2} = O_p\left\{W_n^{-1/d} + n^{-1/2} + \left(\frac{\log n}{n}\right)^{2/(3d)}\right\}.$$

Hence

$$\mathcal{E}_n(\hat{f}) = O_p\left\{W_n^{-1/d} + n^{-1/2} + \left(\frac{\log n}{n}\right)^{2/(3d)}\right\}.$$

Taking square roots in Lemma 5 gives

$$\mathcal{E}(\hat{f}) = O_p\left\{W_n^{-1/d} + \left(\frac{W_n \log n}{n}\right)^{1/2} + n^{-1/2} + \left(\frac{\log n}{n}\right)^{2/(3d)}\right\}.$$

Balancing the neural-network approximation and stochastic complexity terms,

$$W_n^{-1/d} \asymp \left(\frac{W_n \log n}{n}\right)^{1/2},$$

gives

$$W_n \asymp \left(\frac{n}{\log n}\right)^{d/(d+2)}.$$

Under this choice,

$$W_n^{-1/d} \asymp \left(\frac{W_n \log n}{n}\right)^{1/2} \asymp \left(\frac{\log n}{n}\right)^{1/(d+2)}.$$

Therefore,

$$\mathcal{E}(\hat{f}) = O_p\left\{n^{-1/2} + \left(\frac{\log n}{n}\right)^{2/(3d)} + \left(\frac{\log n}{n}\right)^{1/(d+2)}\right\}.$$

Since $d \ge 1$,

$$n^{-1/2} = O\left\{\left(\frac{\log n}{n}\right)^{1/(d+2)}\right\},$$

and hence

$$\mathcal{E}(\hat{f}) = O_p\left\{\left(\frac{\log n}{n}\right)^{2/(3d)} + \left(\frac{\log n}{n}\right)^{1/(d+2)}\right\},$$

which proves the result. ■

C Near-manifold perturbations

This section gives the perturbation analysis for the local model in Section 4. We require $q_n = b_n\sqrt{\log n} \to 0$; the graph and network conditions below impose additional restrictions.

We first translate the standard-deviation scale b_n into a uniform perturbation bound. Define

$$q_n = b_n\sqrt{\log n}.$$

By equation 7 and the union bound, for any sufficiently large constant $K > 0$,

$$P\left(\max_{1\le i\le n}\|\xi_{i,n}\|_2 > Kq_n\right) \le \sum_{i=1}^{n} P\left(\|\xi_{i,n}\|_2 > Kb_n\sqrt{\log n}\right)$$
$$\le C_1 n^{1-C_2K^2} \longrightarrow 0.$$

Hence

$$\max_{1\le i\le n}\|\xi_{i,n}\|_2 = O_p(q_n) = O_p\left(b_n\sqrt{\log n}\right). \tag{21}$$

The logarithmic factor arises because ISOMAP constructs a graph from all observations simultaneously and therefore requires uniform control of the perturbations.

We next derive the effect of this perturbation on the ISOMAP embedding. Let

$$a_n = \sup_{x\in M}\min_{1\le i\le n} d_M(x, X_i^\circ)$$

be the geodesic covering radius of the latent manifold sample. Under the assumptions of Theorem 2,

$$a_n = O_p\left\{\left(\frac{\log n}{n}\right)^{1/d}\right\}. \tag{22}$$

The following result extends the graph-geodesic argument used in Theorem 2 to the near-manifold setting.

Theorem 11 (ISOMAP under near-manifold perturbation) *Suppose that the geometric and distributional assumptions of Theorem 2 hold for the latent observations X_i°, and suppose equation 6–equation 7 hold. Assume in addition that the observed graph is connected with probability tending to one and its shortest-path distances satisfy*

$$\max_{i,j}|\gamma_{ij} - d_M(X_i^\circ, X_j^\circ)| = O_p\big((r_n/\tau)^2 + (a_n/r_n)^2 + q_n/r_n\big). \tag{23}$$

Let r_n denote the radius used to construct the neighborhood graph from $X_1, \dots, X_n$. If

$$r_n \to 0, \qquad \frac{a_n}{r_n} = o_p(1), \qquad \frac{q_n}{r_n} = o(1),$$

then the ISOMAP embedding based on the noisy observations satisfies

$$\delta_{\mathrm{ISO},n} = O_p\left\{\left(\frac{r_n}{\tau}\right)^2 + \left(\frac{a_n}{r_n}\right)^2 + \frac{q_n}{r_n}\right\}, \tag{24}$$

where $\delta_{\mathrm{ISO},n}$ is measured relative to the true embedding of the latent points X_i°.

In particular, letting

$$s_n = \left(\frac{\log n}{n}\right)^{1/d}$$

and choosing

$$r_n \asymp s_n^{1/2} \vee q_n^{1/3}, \tag{25}$$

we obtain

$$\delta_{\mathrm{ISO},n} = O_p\left(s_n + q_n^{2/3}\right) = O_p\left\{\left(\frac{\log n}{n}\right)^{1/d} + b_n^{2/3}(\log n)^{1/3}\right\}. \tag{26}$$

Proof Let $\eta_n = (r_n/\tau)^2 + (a_n/r_n)^2 + q_n/r_n = o_p(1)$. Latent distances are bounded by compactness. Condition equation 23 therefore bounds the maximum error of their squares by $O_p(\eta_n)$. Double centering gives normalized Frobenius Gram error $O_p(\eta_n)$. The smallest eigenvalue of the empirical latent covariance converges to a positive constant, as in the proof of Theorem 2. The classical MDS perturbation bound of Arias-Castro et al. (2020) consequently gives $\delta_{\mathrm{ISO},n} = O_p(\eta_n)$. For $r_n \asymp s_n^{1/2} \vee q_n^{1/3}$, the three terms in η_n are bounded by $O_p(s_n + q_n^{2/3})$, which proves the stated rate. ∎

Condition equation 23 is an additional geometric requirement. An edgewise perturbation bound alone does not control the number of edges in an observed shortest path. A proof of this condition for the noisy radius graph must control both path directions; connectivity and a uniform noise bound alone do not supply that argument.

We next translate the perturbed ISOMAP rate into a depth approximation rate. Let

$$Z_i = \phi(X_i^\circ), \qquad P_Z = P_X^\circ \circ \phi^{-1}.$$

Assume, as in Theorem 3, that every projection $u^\top Z$ has a density f_u satisfying

$$\sup_{u \in \mathbb{S}^{d-1}} \|f_u\|_\infty \le M.$$

The proof of that theorem depends on the ISOMAP coordinates only through the Procrustes discrepancy $\delta_{\mathrm{ISO},n}$ and therefore remains valid in the present setting. Hence

$$\|T_2\|_{n,2} = O_p\left\{n^{-1/2} + \delta_{\mathrm{ISO},n}^{2/3}\right\}.$$

Combining this with equation 26 gives

$$\begin{aligned}\|T_2\|_{n,2} &= O_p\left\{n^{-1/2} + s_n^{2/3} + q_n^{4/9}\right\} \\ &= O_p\left\{n^{-1/2} + \left(\frac{\log n}{n}\right)^{2/(3d)} + b_n^{4/9}(\log n)^{2/9}\right\}.\end{aligned} \tag{27}$$

The exponent $4/9$ in the last term has two sources. The off-manifold perturbation first contributes $q_n^{2/3}$ to the ISOMAP embedding error through optimization of the neighborhood radius. The stability bound for halfspace depth then converts an embedding error δ into an error of order $\delta^{2/3}$, producing $(q_n^{2/3})^{2/3} = q_n^{4/9}$.

To obtain the corresponding out-of-sample result, we also need the neural-network approximation argument to retain the intrinsic dimension d. Let

$$M_{Cq_n} = \{x \in \mathbb{R}^p : d(x, M) \le Cq_n\}.$$

For q_n smaller than the reach of M, the nearest-point projection $\pi_M : M_{Cq_n} \to M$ is well defined. We take the population target for a noisy observation x to be

$$f_0(x) = D_H\left(\phi\{\pi_M(x)\}, P_Z\right).$$

Assume that f_0 admits a uniformly Lipschitz extension on a fixed tubular neighborhood of M. if the approximation scale ρ satisfies $q_n = O(\rho)$, the tube M_{Cq_n} has the same d-dimensional covering order at scale ρ as M:

$$N(M_{Cq_n}, \rho, \|\cdot\|_\infty) \le C_M \rho^{-d}.$$

Consequently, the neural-network approximation argument in Assumption 1 continues to give the intrinsic approximation rate $O(W_n^{-1/d})$ provided

$$q_n = O(W_n^{-1/d}). \tag{28}$$

On the tube, $\|\pi_M(X_i) - X_i^\circ\| \le 2\|\xi_{i,n}\|$. The assumed Lipschitz extension bounds the target change by $O_p(q_n)$, which is absorbed by $O_p(q_n^{4/9})$ when $q_n \to 0$. The tail requirement in the following theorem controls the population risk outside this tube.

We can now state the overall result.

Theorem 12 (Iso-depth rate for data near a manifold) *Suppose the assumptions of Theorem 11, Theorem 3, and the out-of-sample theorem hold, together with equation 28. Assume the approximation and uniform risk bounds hold for the noisy population with constants uniform in n. Extend the projected target to a bounded function outside the tube so that its squared-risk contribution from that region is $o(A_n^2)$, where $A_n = (\log n/n)^{2/(3d)} + (\log n/n)^{1/(d+2)}$. If*

$$W_n \asymp \left(\frac{n}{\log n}\right)^{d/(d+2)},$$

then

$$\mathcal{E}(\hat{f}) = O_p\left\{\left(\frac{\log n}{n}\right)^{2/(3d)} + \left(\frac{\log n}{n}\right)^{1/(d+2)} + b_n^{4/9}(\log n)^{2/9}\right\}. \tag{29}$$

The first two terms in equation 29 are precisely the convergence rate under the exact-manifold model, whereas the last term quantifies the additional error due to departure from the manifold. To make this interpretation explicit, let

$$A_n := \left(\frac{\log n}{n}\right)^{2/(3d)} + \left(\frac{\log n}{n}\right)^{1/(d+2)}$$

denote the exact-manifold rate. If

$$b_n^{4/9}(\log n)^{2/9} = O(A_n), \tag{30}$$

or equivalently

$$b_n = O\left\{\frac{A_n^{9/4}}{\sqrt{\log n}}\right\}, \tag{31}$$

then

$$\mathcal{E}(\hat{f}) = O_p(A_n).$$

Thus the exact-manifold convergence rate is preserved whenever the standard-deviation scale of the off-manifold noise is sufficiently small that its induced error is no larger than the statistical error already present under the ideal manifold model.

Since

$$\frac{2}{3d} - \frac{1}{d+2} = \frac{4-d}{3d(d+2)},$$

condition equation 31 can be expressed more explicitly as

$$b_n = \begin{cases} O\left[(\log n)^{-1/2}\left(\frac{\log n}{n}\right)^{9/\{4(d+2)\}}\right], & d < 4, \\ O\left[(\log n)^{-1/2}\left(\frac{\log n}{n}\right)^{3/8}\right], & d = 4, \\ O\left[(\log n)^{-1/2}\left(\frac{\log n}{n}\right)^{3/(2d)}\right], & d > 4. \end{cases}$$

Under these regimes, departure from the manifold is asymptotically indistinguishable, at the level of the convergence rate, from the ordinary statistical error of Iso-depth.

D Numerical experiment details

The settings below correspond to the experiments in Section 6.

D.1 Rolled-ball model and numerical implementation

Let Z be uniform on the unit ball in $\mathbb{R}^d$, $d \in \{1, 2\}$, and let $\beta = 3\pi/4$. The native coordinates are $\psi(Z) = (\cos(\beta Z_1)/\beta, \sin(\beta Z_1)/\beta, Z_2, \ldots, Z_d)$. A fixed orthonormal map embeds these coordinates in $\mathbb{R}^{50}$. The chart preserves lengths, and the convex latent ball gives $d_M\{\psi(z), \psi(z')\} = \|z - z'\|_2$. The population halfspace depth is $D_0(z) = \frac{1}{2} I_{1-\|z\|_2^2}((d+1)/2, 1/2)$, where I is the regularized incomplete beta function; for $d = 1$, this is $(1 - |z|)/2$.

The geometry-recovery experiment uses 50 seeds, 41000–41049, at each of $n = 250, 400, 630, 1000, 1600, 2500$. The radius is $r_n = c_d\{\log(n)/n\}^{1/(2d)}$. Using pilot seeds 51000–51019, we chose the smallest $c_d \in \{0.25, 0.35, 0.5, 0.7, 1, 1.4, 2, 2.8, 4\}$ for which at least 19 of 20 pilot graphs were connected at every n and no edge joined points with latent distance above $1.5r_n$. This gave $c_1 = 0.5$ and $c_2 = 0.7$. The audit uses the known latent geometry but no labels or depth errors.

The radius graph has Euclidean edge lengths. We compute shortest paths and classical MDS coordinates, then exact depth in one or two dimensions, with angular tolerance 10^{-12} in two dimensions.

Disconnected graphs count toward the connectivity rate but have no recovery errors. The graph error is

$$E_{G,\mathrm{RMS}} = \left[\binom{n}{2}^{-1} \sum_{i<j} \{(\gamma_{ij} - d_{ij})/2\}^2\right]^{1/2},$$

where $d_{ij} = \|Z_i - Z_j\|_2$. The coordinate error δ_{ISO} is the RMSE after rigid Procrustes alignment without scaling. We use the definitions of T_1, T_2, E_D from the main text.

The balanced graph was connected in 48 of 50 runs at $d = 1, n = 250$ and 49 of 50 at $d = 2, n = 250$. Every larger-sample cell had 50 connected runs. In one dimension, $E_{T_2} = 0$ in every valid run because the recovered coordinates preserve or reverse the latent order. Thus the coordinate error bound is not attained by the depth error in these runs.

Table 2: Median recovery errors under the balanced radius, conditional on connected graphs.

d	n	$E_{G,\mathrm{RMS}}$	E_{T_1}	E_{T_2}	E_D
1	250	0.000457	0.021589	0	0.021589
1	400	0.000318	0.015990	0	0.015990
1	630	0.000222	0.012922	0	0.012922
1	1000	0.000153	0.011861	0	0.011861
1	1600	0.000104	0.009556	0	0.009556
1	2500	0.000071	0.007266	0	0.007266
2	250	0.018498	0.018196	0.005873	0.019516
2	400	0.009409	0.015325	0.002972	0.015807
2	630	0.005134	0.012289	0.001666	0.012523
2	1000	0.002697	0.010563	0.000866	0.010700
2	1600	0.001581	0.007980	0.000483	0.008028
2	2500	0.001027	0.006458	0.000281	0.006473

Across the six sample sizes, the fitted sampling-error exponents were 0.486 for $d = 1$ and 0.470 for $d = 2$. The two-dimensional geometry-depth exponent was 1.575 against $\log(n/\log n)$. Zero one-dimensional geometry-depth errors were excluded from that fit. These slopes describe the observed sample-size range, not the sharpness of the theoretical bounds.

D.2 Graph-radius sensitivity in depth recovery

For $d = 2$, write $s(n) = \{\log(n)/n\}^{1/2}$ and $r_* = 0.7s(1000)^{1/2}$. The radius comparison uses the sparse rule $r_* s(n)/s(1000)$, the fixed rule r_*, and the nonlocal rule $\max\{r_*, 1.1(2\sin\beta/\beta)\}$. Each rule has 50 seeds at each sample size. All sparse and nonlocal graphs were connected; the fixed rule had 37 of 50 connected graphs at $n = 250$ and 50 at every larger size.

The nonlocal radius connects parts of the roll that are far apart along the manifold. Under the map ψ in Appendix D.1, the two opposite latent endpoints $z_- = (-1, 0)$ and $z_+ = (1, 0)$ satisfy

$$\|\psi(z_+) - \psi(z_-)\|_2 = \frac{2\sin\beta}{\beta}, \qquad d_M\{\psi(z_+), \psi(z_-)\} = \|z_+ - z_-\|_2 = 2.$$

The orthonormal map into $\mathbb{R}^{50}$ preserves the chord length. At $\beta = 3\pi/4$, it is about 0.600, while the nonlocal radius is about 0.660. The radius therefore permits direct edges between sampled points near opposite ends of the roll. These edges can replace a long manifold path with a short ambient chord and underestimate geodesic distance. The fixed and nonlocal radii are both constant in n; the nonlocal radius exceeds the endpoint chord by construction.

Median depth error for the nonlocal graphs rose from 0.030067 at $n = 250$ to 0.035590 at $n = 2500$. At $n = 2500$, errors under the fixed and sparse rules were 0.006462 and 0.006506. All runs were connected at that size, but the nonlocal edges distorted the recovered geometry.

D.3 Independent-test prediction

The out-of-sample experiment uses the same rolled-ball model with $d = 2$ and 50 seeds at $n = 250, 630, 1000, 2500$. Each seed has 10,000 independent test observations. Oracle and pseudo-depth fits share their test sample, initialization stream, and minibatch stream. The network has

Table 3: Radius sensitivity: medians over connected runs. Stress is $\{\sum_{i,j}(\gamma_{ij} - \|\hat{z}_i - \hat{z}_j\|)^2 / \sum_{i,j}\gamma_{ij}^2\}^{1/2}$.

Rule	n	$E_{G,\mathrm{RMS}}$	δ_{ISO}	E_{T_1}	E_{T_2}	E_D	Stress
fixed	250	0.063602	0.099438	0.017838	0.018264	0.026127	0.053318
fixed	400	0.021647	0.031689	0.015325	0.005768	0.017279	0.022066
fixed	630	0.007801	0.010565	0.012289	0.002219	0.012746	0.009632
fixed	1000	0.002697	0.003728	0.010563	0.000866	0.010700	0.003923
fixed	1600	0.001469	0.002610	0.007980	0.000402	0.008014	0.001793
fixed	2500	0.001562	0.002887	0.006458	0.000267	0.006462	0.001148
sparse	250	0.006321	0.010749	0.018298	0.002611	0.018936	0.009593
sparse	400	0.004546	0.007168	0.015325	0.001811	0.015430	0.006971
sparse	630	0.003642	0.005304	0.012289	0.001311	0.012478	0.005426
sparse	1000	0.002697	0.003728	0.010563	0.000866	0.010700	0.003923
sparse	1600	0.002235	0.002913	0.007980	0.000632	0.008066	0.002992
sparse	2500	0.001867	0.002348	0.006458	0.000453	0.006506	0.002265
nonlocal	250	0.040558	0.091368	0.018298	0.021598	0.030067	0.067477
nonlocal	400	0.044426	0.099020	0.015325	0.024744	0.032690	0.074543
nonlocal	630	0.048926	0.110227	0.012289	0.028071	0.032898	0.082839
nonlocal	1000	0.052152	0.117119	0.010563	0.031372	0.034115	0.087512
nonlocal	1600	0.053732	0.121002	0.007980	0.032921	0.034536	0.090161
nonlocal	2500	0.055502	0.124412	0.006458	0.034273	0.035590	0.092550

two ReLU hidden layers, with widths 10, 14, 16, and 22, respectively, and a linear output. Adam uses learning rate 0.001, batch size 128, zero weight decay, and 1,000 epochs. After each optimizer update, parameters are clipped to $[-W, W]$, where W is the parameter count; predictions are clipped to $[0, 1]$ at evaluation. All available fits completed the training budget and no parameter clipping was active. This optimization diagnostic does not certify exact ERM.

One disconnected graph at $n = 250$ left 199 paired fits; all 200 oracle fits completed.

Table 4: Out-of-sample prediction. RMSE of the network prediction against the population depth D_0 on 10,000 independent test observations. Mean (SD), median, and 10th–90th percentiles are across seeds.

n	Training target	Runs	Mean (SD)	Median	10th–90th percentile
250	Estimated depths	49	0.021697 (0.004880)	0.020678	0.016450–0.029368
250	Population depths	50	0.009632 (0.002965)	0.009217	0.006436–0.013665
630	Estimated depths	50	0.013869 (0.003244)	0.013069	0.010006–0.018001
630	Population depths	50	0.005009 (0.001363)	0.004938	0.003291–0.006869
1000	Estimated depths	50	0.010747 (0.002273)	0.010554	0.007807–0.013823
1000	Population depths	50	0.003494 (0.000860)	0.003546	0.002387–0.004631
2500	Estimated depths	50	0.006793 (0.001588)	0.006660	0.004998–0.008821
2500	Population depths	50	0.002097 (0.000460)	0.002045	0.001584–0.002765

D.4 Intrinsic-boundary detection models and baselines

Data-generating mechanisms. For the Swiss roll, let $Z = (Z_1, Z_2)$ be uniform on $[-1, 1] \times [-0.35, 0.35]$. A unit-speed parameterization maps Z_1 to the spiral interval $[1.5\pi, 4.5\pi]$, producing a 1.5-turn roll, and uses Z_2 as the vertical coordinate. In these three coordinates, the anomaly label is

$$Y = \mathbf{1}\{\min(1 - |Z_1|,\, 0.35 - |Z_2|) < \delta_h\}\,,$$

where δ_h is chosen so that the boundary band has probability 0.03. In particular, with $h = 0.35$,

$$\delta_h = \frac{(1+h) - \sqrt{(1+h)^2 - 0.12h}}{2}.$$

Edge and corner points belong to the same anomaly class.

For the gapped circle, let $U \sim \mathrm{Unif}[-1, 1]$ and set

$$\theta(U) = \frac{\pi}{8} + \frac{2\pi - \pi/4}{2}(U + 1), \qquad \psi_C(U) = \frac{2}{2\pi - \pi/4}\{\cos\theta(U), \sin\theta(U)\}.$$

In these two coordinates, the label is $Y = \mathbf{1}\{|U| > 0.97\}$. The two endpoint bands form one anomaly class.

For either dataset, let W be the native coordinate vector, let m be its dimension, and set $q = p - m$. We append q independent $N(0, \sigma_\xi^2)$ coordinates, where

$$D_M^2 = 2\,\mathrm{tr}\{\mathrm{Cov}(W \mid Y = 0)\}, \qquad \sigma_\xi = \frac{0.05 D_M}{\sqrt{2q}}.$$

The covariance trace is evaluated once by deterministic quadrature over the normal latent region and does not depend on the realized sample. Under this calibration,

$$\frac{\mathbb{E}\|\xi - \xi'\|^2}{\mathbb{E}\|W - W'\|^2} = 0.05^2$$

for independent draws from the normal population. The detection experiment uses seeds 1000–1099, $n = 1000$, $p = 50$, and radius $1.10 r_{\mathrm{conn}}$, with 100 runs per method and dataset. We also test multipliers 1.00 and 1.25. Iso-depth and ISOMAP+LOF use the same coordinates. Ambient depth uses 10,000 random directions; latent and recovered one- or two-dimensional depths are exact. LOF uses $k = 20$ primarily and $k \in \{10, 20, 50, 100\}$ for sensitivity. Isolation Forest uses 500 trees and subsamples of size 256. OCSVM uses an RBF kernel, $\nu = 0.05$, and inverse median squared distance as its bandwidth parameter. ECOD uses its empirical-tail score.

Deep SVDD uses the DeepOD 0.4.1 implementation with three initializations per data seed. Before fitting, we subtract the column means of the fitting sample and divide by one scalar, the root mean square of the centered observation norms. The network has hidden widths 100 and 50, a 128-dimensional representation, ReLU activations, and no bias terms. The center, loss, and score are those of the implementation. Adam uses learning rate 0.001, batch size 64, and weight decay 10^{-5}, for 100 epochs. For each data seed, AP and AUROC are averaged over the three fits. Labels are not used to choose an initialization or epoch. All fits completed.

Table 5 includes all baselines. Tables 6 and 8 vary the LOF neighborhood size and graph radius.

Table 5: Full baseline comparison for peripheral-point detection at $n = 1000$, $p = 50$, and radius multiplier 1.10. Entries are mean (standard deviation) over 100 paired replications.

	Gapped circle		Swiss roll	
Method	AP	AUROC	AP	AUROC
Iso-depth	0.9815 (0.0198)	0.9995 (0.0006)	0.3402 (0.0941)	0.9415 (0.0214)
Latent empirical depth (reference)	0.9854 (0.0191)	0.9996 (0.0006)	0.5786 (0.0760)	0.9839 (0.0042)
ISOMAP+LOF ($k = 20$)	0.2368 (0.1731)	0.7264 (0.1305)	0.1687 (0.0700)	0.7915 (0.0712)
Ambient LOF ($k = 20$)	0.3362 (0.1909)	0.7560 (0.1331)	0.3157 (0.0814)	0.9140 (0.0318)
Ambient depth	0.0311 (0.0055)	0.5013 (0.0007)	0.0448 (0.0087)	0.6411 (0.0334)
Euclidean MDS+depth	0.0188 (0.0033)	0.1608 (0.0158)	0.2553 (0.0612)	0.8562 (0.0388)
Isolation Forest	0.0449 (0.0152)	0.5589 (0.0552)	0.0471 (0.0162)	0.5647 (0.0573)
OCSVM	0.1393 (0.0390)	0.7882 (0.0496)	0.3183 (0.0765)	0.7116 (0.0575)
ECOD	0.0507 (0.0178)	0.5864 (0.0482)	0.0616 (0.0203)	0.6254 (0.0529)
Deep SVDD (100 epochs)	0.2614 (0.1254)	0.7782 (0.1005)	0.0588 (0.0176)	0.6073 (0.0456)

On the Swiss roll, Iso-depth exceeded ambient LOF by 0.0245 AP (SE 0.0097) and same-embedding LOF by 0.1715 AP (SE 0.0070). When the best ambient LOF neighborhood size was selected using evaluation labels, its mean AP exceeded Iso-depth's. Iso-depth retained a positive difference from same-embedding LOF across both datasets and metrics.

Table 6: LOF sensitivity for the same samples and recovered coordinates as the main detection experiment. Entries are AP / AUROC.

Dataset	Representation	k=10 AP / AUROC	k=20 AP / AUROC	k=50 AP / AUROC	k=100 AP / AUROC
Gapped circle	Ambient LOF	0.0771 / 0.5402	0.3362 / 0.7560	0.8531 / 0.9809	0.9377 / 0.9975
Gapped circle	ISOMAP+LOF	0.0653 / 0.5323	0.2368 / 0.7264	0.7340 / 0.9628	0.9195 / 0.9965
Swiss roll	Ambient LOF	0.2313 / 0.8288	0.3157 / 0.9140	0.3198 / 0.9223	0.2905 / 0.8701
Swiss roll	ISOMAP+LOF	0.1203 / 0.7123	0.1687 / 0.7915	0.2040 / 0.8567	0.1993 / 0.8768

Table 7: Label-assisted best-k diagnostic. Differences are Iso-depth minus the per-seed, per-metric best LOF score. These choices use evaluation labels.

Dataset	Oracle-best LOF family	Metric	R	Difference (SE)	98.75% CI
Gapped circle	Ambient LOF	AP	100	+0.0267 (0.0063)	[+0.0117, +0.0441]
Gapped circle	Ambient LOF	AUROC	100	+0.0014 (0.0003)	[+0.0006, +0.0024]
Gapped circle	ISOMAP+LOF	AP	100	+0.0464 (0.0075)	[+0.0281, +0.0664]
Gapped circle	ISOMAP+LOF	AUROC	100	+0.0024 (0.0005)	[+0.0013, +0.0039]
Swiss roll	Ambient LOF	AP	100	-0.0229 (0.0109)	[-0.0497, +0.0044]
Swiss roll	Ambient LOF	AUROC	100	+0.0112 (0.0028)	[+0.0040, +0.0187]
Swiss roll	ISOMAP+LOF	AP	100	+0.1135 (0.0065)	[+0.0974, +0.1300]
Swiss roll	ISOMAP+LOF	AUROC	100	+0.0586 (0.0028)	[+0.0522, +0.0664]

Table 8: Observed-radius sensitivity. Entries are means with standard deviations over 100 seeds.

Dataset	Multiplier	AP mean (SD)	AUROC mean (SD)
Gapped circle	1.00	0.9778 (0.0212)	0.9993 (0.0007)
Gapped circle	1.10	0.9815 (0.0198)	0.9995 (0.0006)
Gapped circle	1.25	0.9832 (0.0195)	0.9995 (0.0006)
Swiss roll	1.00	0.2896 (0.1001)	0.9126 (0.0487)
Swiss roll	1.10	0.3402 (0.0941)	0.9415 (0.0214)
Swiss roll	1.25	0.3985 (0.0985)	0.9572 (0.0158)

E DATA APPLICATION DETAILS

E.1 CARDIO PREPROCESSING AND DIAGNOSTICS

Preprocessing and feature-scale sensitivity. The ADBench Cardio file contains standardized values for 21 cardiotocography variables and 1,831 records: 1,655 normal and 176 pathologic. It excludes the suspect category. Before analysis, we median-centered each variable and scaled it by its IQR, with an adjustment to the scale of fetal movement (FM), as described below. The Cardio results reported in the main text use this preprocessing.

Feature scaling determines the relative contributions of the variables to the Euclidean distances used to construct the neighborhood graph. In the original measurement scale, FM has many zero values and a long upper tail. In the standardized ADBench file, its standard deviation is approximately 1, whereas its IQR is only 0.043. Consequently, under IQR scaling, FM accounted for 94.9% of the total pairwise squared Euclidean distance. This concentration reflects the weight assigned to FM by the input metric.

To limit the amplification associated with its small IQR, we used

$$s_{\mathrm{FM}} = \max\left\{\frac{\mathrm{IQR}(\mathrm{FM})}{1.349},\ 0.25\ \mathrm{SD}_{\mathrm{pop}}(\mathrm{FM})\right\}$$

as the FM scale. All other variables retained their IQR scales. The factor 1.349 expresses the IQR on a standard-deviation scale under a normal reference distribution, and the second term places a lower bound on the denominator. All observations and feature values were retained, without clipping or removal.

Here, $\mathrm{SD}_{\mathrm{pop}}$ denotes the empirical standard deviation calculated with divisor n_{fit}. Preprocessing quantities were computed from all 1,831 observations in the transductive analysis and from training observations only in each out-of-sample split. Pathology labels were not used to calculate these quantities.

For comparison, unmodified IQR scaling yielded a one-dimensional embedding under the dimension-selection rule below. We also examined applying

$$s_j = \max\left\{\frac{\mathrm{IQR}(X_j)}{1.349},\ 0.25\ \mathrm{SD}_{\mathrm{pop}}(X_j)\right\}$$

to every variable. Under this alternative, five sparse variables jointly accounted for approximately 82% of the total pairwise squared distance. These comparisons document the sensitivity of the input metric to feature scaling. The FM adjustment addresses the concentration observed in Cardio and is not proposed as a general preprocessing rule.

Embedding dimension. The graph radius is r_{conn}, the smallest radius that connects the graph; each out-of-sample split recomputes it from its training observations. For the preprocessed Cardio data, we selected the embedding dimension using residual variance,

$$\mathrm{RV}(d) = 1 - \mathrm{Corr}^2\left((\gamma_{ij})_{i<j}, \left(\|\widehat{Z}_i^{(d)} - \widehat{Z}_j^{(d)}\|_2\right)_{i<j}\right),$$

where γ_{ij} denotes the fitted graph shortest-path distance. We selected the smallest $d \in \{1, \ldots, 16\}$ satisfying $\mathrm{RV}(d) \le 0.05$. Residual variance was 0.1312 at $d = 1$ and 0.0390 at $d = 2$, giving a two-dimensional representation. This criterion measures agreement with the fitted graph distances and does not establish the manifold's intrinsic dimension.

Baseline settings. LOF uses $k = 20$, on the preprocessed covariates (Ambient LOF) and on the ISOMAP coordinates. OCSVM uses an RBF kernel, $\nu = 0.1$, and the scikit-learn `"scale"` bandwidth. Isolation Forest uses 500 trees and subsamples of size 256, with 10 seeds in the transductive analysis and one seed per split in the out-of-sample analysis. ECOD uses the PyOD implementation. Deep SVDD uses the settings of Appendix D.4, with three seeds in the transductive analysis and one per split out of sample. PCA-Depth computes exact halfspace depth after projecting onto the first $d = 2$ principal components. Ambient depth in $\mathbb{R}^{21}$ is approximated with 32,000 random directions. We checked it against the result with 8,000 directions, requiring a rank correlation of

at least 0.995 and a Jaccard index of at least 0.95 for the shallowest 10% of observations. Out of sample, LOF, OCSVM, ECOD, PCA-Depth, and ambient depth are fit on the training observations, and test observations are scored against them.

Transductive rankings. We computed halfspace depth exactly in the two-dimensional Cardio representation. Iso-depth attained AP 0.596, AUROC 0.920, and precision 0.591 at $k = 176$. PCA-Depth attained AP 0.245, and LOF on the same ISOMAP coordinates attained AP 0.167.

At $k = 176$, ECOD selected 93 pathologic observations, OCSVM selected 73, and Isolation Forest selected an average of 89.4 across 10 seeds. Their precisions were 0.528, 0.415, and 0.508, respectively. Iso-depth and ECOD selected 69 pathologic observations in common; 35 were selected only by Iso-depth, 24 only by ECOD, and 48 by neither. The cutoff $k = 176$ equals the number of pathologic records and is used only for evaluation. These counts do not give the ISOMAP coordinates a clinical interpretation.

Connectivity-radius diagnostic. At 0.7, 0.8, and 0.9 times the connectivity radius r_{conn}, the largest component contained 1,824 observations. The same seven observations formed the other component at each radius. This group was identified from covariates; six members had pathologic labels.

The components merged at r_{conn}. Severe deceleration (DS) accounted for 96.5% of the squared length of the minimum-spanning-tree edge joining them. Thus, the connection of seven observations, about 0.4% of the sample, determined the radius for the whole graph. At that radius, the graph density was 0.890, so most pairs were connected directly by their Euclidean distances. Connectivity alone therefore does not establish that the graph shortest paths recover latent geodesic distances.

Out-of-sample evaluation. We used 20 stratified 70/30 splits. In each split, training observations alone determined the scaler, graph, dimension, pseudo-depths, and neural network. The network has two ReLU hidden layers of width 16 and a linear output, with 641 parameters. It is trained by Adam with learning rate 0.001, batch size 128, and no weight decay, for 1,000 epochs. After each update, parameters are clipped to $[-641, 641]$. We use the final-epoch network and clip its predictions to $[0, 1]$. Pathology labels were used for stratification and evaluation. Every split selected two dimensions and passed the numerical checks for exact depth.

Across splits, Iso-depth-NN had AP 0.577 ± 0.076, AUROC 0.916 ± 0.019, and precision at $k = 53$ of 0.566 ± 0.063. The corresponding values were 0.598 ± 0.045, 0.938 ± 0.009, and 0.547 ± 0.045 for ECOD, and 0.589 ± 0.046, 0.932 ± 0.009, and 0.511 ± 0.043 for Isolation Forest.

At $k = 53$, Iso-depth-NN had higher precision than ECOD in 12 splits and Isolation Forest in 15, with one tie in each comparison. Its AP exceeded PCA-Depth, Ambient-Depth, OCSVM, and Deep SVDD in every split. ECOD and Isolation Forest had higher mean AP and AUROC. The cutoff $k = 53$ uses the test-set anomaly count and is therefore an evaluation cutoff.