A Supplementary Material: Formal Proofs

Proof of Corollary 2.5. The hypothesis $[Y] \mapsto [0]$ implies that for some $\alpha_0 \leq \alpha_2$ and $\varepsilon_0 \leq \varepsilon_2$, there exist chains $Y \in P_{\alpha_1,\varepsilon_1}$ and $Z \in P_{\alpha_2,\varepsilon_2}$ and $W \in P_{\alpha_0,\varepsilon_0}$ such that $\partial Z = Y - W$ in $P_{\alpha_2,\varepsilon_2}$. By Lemma 2.4, these chain can be included to their respective levels in R, giving $Y \in R_{\alpha_1}$, $Z \in R_{\alpha_2}$, and $W \in R_{\alpha_0}$, still satisfying $\partial Z = Y - W$ in R_{α_2}. Hence, applying ι_*, either $\iota_*([Y])$ is trivial in HR_{α_1}, or the transition map $HR_{\alpha_1} \to HR_{\alpha_2}$ takes $\iota_*([Y]) \mapsto [0]$. □

Proof of Lemma 2.6. If $\varepsilon \geq \alpha$, then the second condition in Definition 2.2 becomes $\rho_K(Y) - \rho_V(Y) \leq \alpha$, which is a trivial condition for $0 \leq \rho_V(Y) \leq \rho_K(Y) \leq \alpha$. Thus, the sets $P_{\alpha,\varepsilon}$ are identical for all $\varepsilon \geq \alpha$. □

Proof of Lemma 3.4. Suppose an edge $e = (x_i, x_j)$ lies in $P_{\alpha,\varepsilon}$, meaning that e is represented by a V-geodesic of length $2\rho_V(e)$ satisfying $\rho_V(e) \leq \alpha$ and a K-geodesic of length $2\rho_K(e)$ satisfying $\rho_V(e) \leq \rho_K(e) \leq \min\{\alpha, \rho_V(e) + \varepsilon\}$. By Lemma 3.3, the corresponding edge $e' = f_\sharp(e) = (f(x_i), f(x_j))$ satisfies $\rho_V(e) - \kappa \leq \rho_V(e') \leq \rho_V(e) + \kappa$. Moreover, by the data perturbation lemma on K, we know e' is represented by a K-geodesic with length $2\rho_K(e')$ satisfying $\rho_K(e') \leq \rho_K(e) + \kappa$. Combining these overall, we have the parallax bounds

$$\rho_V(e') \leq \rho_K(e') \leq \min\{\alpha, \rho_V(e) + \varepsilon\} + \kappa$$

which implies both $\rho_K(e') \leq \rho_V(e') + \varepsilon + 2\kappa$ and $\rho_K(e') \leq \alpha + \kappa$. □

Proof of Lemma 4.6. Fix $\lambda' = \frac{1}{2}\lambda_{\text{ball},X}(K) - \kappa$. We aim to show that X' and K are λ'-LSM. Fix any edge $e' = (x', y')$ in R' satisfying $\rho_V(e') \leq \lambda'$. (The set of such edges might be empty, which is allowed by Definition 4.1.) For any such e', there is a corresponding $e = (x, y) \in R$ with $f(x) = x'$, $f(y) = y'$, and $f_\sharp(e) = e'$. The triangle inequality provides $\|y - x\| \leq \|y - y'\| + \|y' - x'\| + \|x' - x\| = 2\lambda' + 2\kappa \leq \lambda_{\text{ball},X}(K)$. We are assuming V is Euclidean, so all balls in V are convex. Since x', y, y' are all within the ball of radius $\lambda_{\text{ball},X}(K)$ around x, all their pairwise geodesics are included in K. Hence, $e' \in P'_{\lambda',0}$. □

Proof of Lemma 5.2. The relationship $c \leq d$ follows from Lemma 2.5. The relationship $d \leq e$ follows from Lemma 2.3 and the observations that $P_{\alpha,\alpha} = P_{\alpha,\infty} = R(X, K)$. Suppose L is parameterized as $L_\alpha = P_{\alpha,\varepsilon(\alpha)}$. Suppose that $c < e$, so either $c < d$ or $d < e$.

If $c < d$, then we can conclude that the edge that killed $[Y]$ in $HR(X, V)$ is not present in $HR(X, K)$, so that edge intersected K^c, which is an open set. Hence, that killing edge intersects a void in the sense of Defn 1.2. Let B_r be an open ball in K^c centered on some point on the V-geodesic of length $2c$. If $B_r = K^c$, then the shortest K-geodesic replaces the diameter $2r$ with the half-circumference πr. Therefore, $\pi r - 2r \leq 2d - 2c$.

If $d < e$, then $[Y] \mapsto [0]$ via $HP_{b,b} \to HP_{d,d}$ and via $HP_{b,b} \to HP_{e,\varepsilon(e)}$, but not for any $HP_{b,b} \to HP_{\alpha,\varepsilon(\alpha)}$ with $\alpha < e$. Therefore, there is a K-geodesic of length $2d$ and a different K-geodesic of length $2e$, either of which could kill $[Y]$. The former edge is not allowed in L because $0 \leq \varepsilon(d) < d - c$. Hence, $c < d$, returning us to the first case. □

Proof of Lemma 5.3. Let L be a Rips-like path given by Definition 5.1. Because X and K are λ-locally simplicially matched, the filtrations levels of R and L are identical up to λ. Therefore, there is a bijection between the persistence diagrams of R and L for all dots born by λ.

Suppose that (b, d) is a dot in the persistence diagram of L with $b \leq \lambda < \omega < d$. This dot represents a class $[Y]$ born in HL_b that dies in HL_d. The class $[Y] \in HR_b$ is born at b must die at some value c, so $[Y] \in \ker HR_{b\to c}$. It cannot be that $c \leq \delta$, because that would imply $[Y] \in \ker HL_{b\to\omega}$, contradicting $\omega < d$. Therefore, $\delta < c$. □

Proof of Theorem 5.4. Let $f_* : HP \to HP'$ denote the map induced by f and similarly, let $f_*^{-1} : HR' \to HR$ denote the map induced by f^{-1}. Consider the following commutative diagram in which

the unlabeled morphisms are transition maps of the respective persistence modules.

$$\begin{array}{ccccccc}
 & & HR'_{\delta-\kappa} & \xrightarrow{f_*^{-1}} & HR_\delta & & \\
 & & \uparrow & & \uparrow & & \\
N' & \xrightarrow{\subset} & HR'_{\lambda-\kappa} & \xrightarrow{f_*^{-1}} & HR_\lambda = HP_{\lambda,\epsilon_1} & \longrightarrow & HP_{\omega,\epsilon_2} \\
 & & \downarrow{=} & & \downarrow{f_*} & & \downarrow{f_*} \\
 & & HP'_{\lambda-\kappa,\epsilon_1+2\kappa} & \longrightarrow & HP'_{\lambda+\kappa,\epsilon_1+2\kappa} & \longrightarrow & HP'_{\omega+\kappa,\epsilon_2+2\kappa}
\end{array}$$

Note that the bi-degrees in the diagram are determined by Lemma 3.3 and Corollary 3.5. Set $N' = \ker HR'_{\lambda-\kappa\to\delta-\kappa} \subset HR'_{\lambda-\kappa}$. Commutativity of the top left square implies that $f_*^{-1}(N') \subset HR_\lambda$ maps vertically to 0 in HR_δ and hence, $f_*^{-1}(N')$ lies in $\ker HR_{\lambda\to\delta}$. The homologically matched assumption on (X, K) implies $f_*^{-1}(N') \subset \ker HP_{(\lambda,\epsilon_1)\to(\omega,\epsilon_2)}$. By commutativity of the bottom right square, $f_*(f_*^{-1}(N'))$ maps horizontally to $0 \in HP'_{\omega+\kappa,\epsilon_2+2\kappa}$. Finally, the locally simplicially matched assumption means we can instead think of N' as a subset of $HP'_{\lambda-\kappa,\epsilon_1+2\kappa}$ which maps to 0 in $HP'_{\omega+\kappa,\epsilon_2+2\kappa}$ by commutativity. Thus,

$$N' = \ker HR'_{\lambda-\kappa\to\delta-\kappa} \subset \ker HR'_{(\lambda-\kappa,\epsilon_1+2\kappa)\to(\omega+\kappa,\epsilon_2+2\kappa)} ,$$

completing the proof. □

B Supplementary Material: Code

The code supporting this article is an initial proof-of-concept. It is available publicly at

https://gitlab.com/geomdata/topological-parallax

It is designed as a simple Python package, following community best-practices for tooling and layout. Documentation is provided there. We recommend that the reader go to the repository for issue tracking, improvements, bug fixes, testing pipelines, etc.

Because this package relies on GUDHI The GUDHI Project [34], the filtration values are by diameter (not radius). This is important to note, because the theory in the paper is written using radius (not diameter) as the filtration value.

Topological parallax has the same computational complexity as the computation of Rips complexes and their persistence diagrams—albeit with a larger constant. This is because parallax merely inserts a model-evaluation step upon the examination of each edge. The constant therefore is tN^2 for N points and a model that takes time t to evaluate. There are very interesting dimension- and structure-dependendent estimates for the real-life/expected timing of Rips computations, such as Bauer et al. [4]. Distributed persistence can parallelize this process, as in Solomon et al. [31]. See https://gitlab.com/geomdata/dispers.

- Figure 1 was produced with the Jupyter notebook `notebooks/Winky Example.ipynb`. Compare and contrast the Python scripts `runners/run_winky_nn.py` versus `runners/run_winky_tree.py`
- Figure 3 was produced with the Python script `runners/run_cyclooctane_balls.py`. Compare and contrast that script with `runners/run_cyclooctane_nnfar.py` and `run_cyclooctane_nnhull.py`. These examples may take 100-500 GB of RAM to compute the persistence diagrams as currently implemented.
- Figure 4 was produced with the Python script `runners/bond_lengths.py`

C Supplementary Material: Utah Jar Example

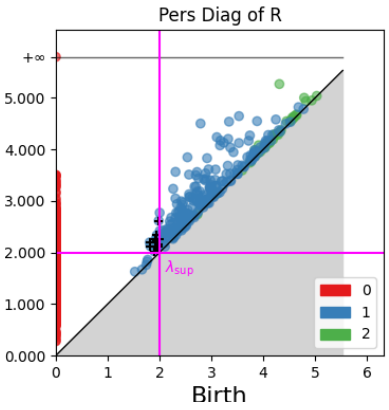
This example demonstrates a situation where a classifier appears to be 100% correct on a semantically meaningful dataset, but the resulting model is too strict for interpolation within a data class, Topological parallax detects this phenomenon, as discussed in Section 6.

We applied parallax to an imagery dataset inspired by the "Utah teapot." The dataset consists of 14000 images—4000 images of a rotated teapot and 10000 images of a rotated teapot with no spout or handle (referred to as the "Utah jar"). See Figure 5a. All images were rendered with PoV-Ray.

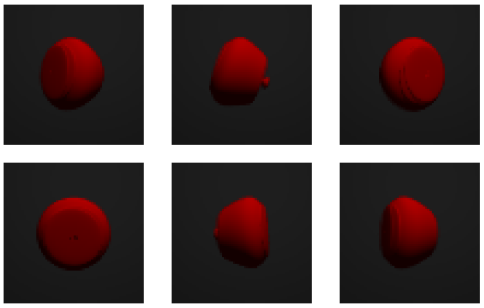
We constructed a bespoke classifier on this dataset of images which accepts any image within a distance of 1.0 to the set of jar images using the Euclidean metric on flattened images. Our classifier rejects all other inputs. We applied parallax to this model and found very small values for both λ_{lo} and λ_{sup}, with $\lambda_{\text{lo}} \sim \lambda_{\text{sup}} < 2$. See Figure 5b.

Given the hard cut-off distance of 1.0 in the classifier and these λ values, we expect the model to prohibit reasonable interpolations of images. Indeed, we found numerous interpolations rejected by the model which seem valid to the human eye. See Figure 5c.

(a) Examples from the modified Utah Teapot dataset.



(b) The persistence diagram of the dataset marked by parallax.



(c) Interpolative images that are visually similar to true dataset samples but are rejected by the overly sensitive model.

Figure 5: Using parallax to detect a dataset-model geometry mismatch. (a) The images consist of the Utah teapot (rendered with PoV-Ray) with spouts and handles removed for simplification. (b) We construct a simple classifier which accepts any images within a Euclidean distance of 1.0 of the (flattened) image and rejects all others, and apply parallax to it. (c) Six interpolated images are shown, all of which look very similar to the original dataset but are rejected by the model (corresponding to the X marks in (b)).

D Supplementary Material: Understanding the Bi-Complex

The parallax complex $P(X, K, V)$ is bi-filtered by the parameters α (the length of a geodesic) and ε (the distortion of geodesic length between K and V). Figure 6 provides some visualizations that may help the reader interpret traditional barcodes of Rips-like paths, and how these bi-filtration parameters may be estimated. These parameters are related to the size of voids in the model, as seen in Lemma 5.2.

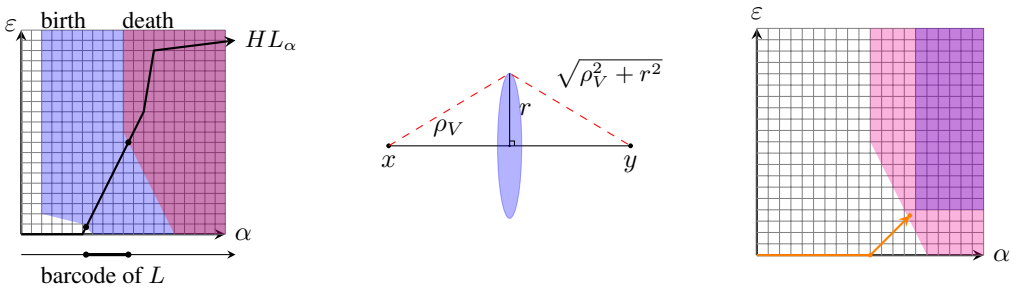


Figure 6: Left: Births and deaths are bi-filtered in HP, and are observed by barcodes on Rips-like paths. Cor 3.5 means this picture is stable within $\pm(\kappa, 2\kappa)$. Center: An edge $(x, y) \in P_{\alpha,\varepsilon}$ can be estimated by sampling tubular neighborhoods. Right: Algorithm 7.5 underestimates α, ε.

E Supplementary Material: Table of Symbols

Notation	Plain Meaning	First Appearance	Term
V	A geodesic space, such as $\mathbb{R}^n$	p.2	Ambient Space
X	A finite set in V	p.2	Dataset
k	A perception function on V	p.2	Model (as function)
K	Support set of k	p.2	Model (as set)
$\mathcal{M}(X)$	Models compatible with dataset X	p.2	
$\mathcal{M}^*(K)$	Datasets compatible with model K	p.2	
K°	Interior of set K in topological space V	p.2	
$\overline{K}$	Closure of set K in topological V	p.2	
K^c	Complement of set K in V	p.2	
Ω	Bounded open set in K^c	p.2	Void
$R(X, K)$	Rips complex of X in geodesic space K	p.4	
α	a filtration level or radius	p.4	
$B_\alpha(x)$	Geodesic ball of radius α about x	p.4	
Y	Chain in a Rips complex	p.4	
$\rho_K(Y)$	Minimal filtration radius for Y in $R(X, K)$	p.4	
ε	Gap in filtration radius between V and K	p.4	
P	Parallax bi-complex for X, K, V	p.4	Parallax
HP	Homology of P	p.4	
L	A 1-parameter path through P	p.4	Rips-like Path
HL	Homology of L	p.4	
$\overset{\kappa}{\approx}$	Pointwise perturbation of X in V	p.5	Perturbation
$\overset{\kappa}{\approx}_K$	Pointwise perturbation of X in K	p.5	K-Perturbation
$f_\sharp$	Induced map on a simplicial complex	p.5	
f_*	Push-forward map on homology	p.5	
$\lambda_\bullet$	A meaningful filtration value. See subscript.	p.6	