

APPENDIX

A EXTENDED OVERVIEW OF TOPOLOGICAL DATA ANALYSIS

In this section, we give a brief introduction of algebraic topology for topological data analysis to supplement Section 2.1 with formal definitions of ingredients that are used throughout this paper. Here, we use the notation $\mathbb{B}_{\mathbb{R}^d}(x, r)$ (or simply $\mathbb{B}(x, r)$ when the space is clear from context) for the open ball in \mathbb{R}^d centered at x with radius r . We will let \mathbb{X} and \mathcal{X} denote a subset of \mathbb{R}^d and a finite collection of points from W respectively, unless otherwise stated.

A.1 DISTANCE BETWEEN SETS ON METRIC SPACES

When topological information of the underlying space is approximated by the observed points, it is often needed to compare two sets with respect to their metric structures. Here we present two distances on metric spaces, Hausdorff distance and Gromov-Hausdorff distance.

The *Hausdorff distance* is on sets embedded in the same metric spaces. This distance measures how two sets are close to each other in the embedded metric space. When $S \subset \mathbb{X}$, we denote by $U_r(S)$ the r -neighborhood of a set S in a metric space, i.e. $U_r(S) = \bigcup_{x \in S} \mathbb{B}_{\mathbb{X}}(x, r)$.

Definition A.1 (Hausdorff distance). *Let \mathbb{X} be a metric space, and $X, Y \subset \mathbb{X}$ be a subset. The Hausdorff distance between X and Y , denoted by $d_H(X, Y)$, is defined as*

$$d_H(X, Y) = \inf\{r > 0 : X \subset U_r(Y) \text{ and } Y \subset U_r(X)\}.$$

The *Gromov-Hausdorff distance* measures how two sets are far from being isometric to each other. To define the distance, we first define a relation between two sets called *correspondence*.

Definition A.2. *Let X and Y be two sets. A correspondence between X and Y is a set $C \subset X \times Y$ whose projections to both X and Y are both surjective, i.e. for every $x \in X$, there exists $y \in Y$ such that $(x, y) \in C$, and for every $y \in Y$, there exists $x \in X$ with $(x, y) \in C$.*

For a correspondence, we define its *distortion* by how the metric structures of two sets differ by the correspondence.

Definition A.3. *Let X and Y be two metric spaces, and C be a correspondence between X and Y . The distortion of C is defined by*

$$dis(C) = \sup\{|d_X(x, x') - d_Y(y, y')| : (x, y), (x', y') \in C\}.$$

Now the Gromov-Hausdorff distance is defined as the smallest possible distortion between two sets.

Definition A.4 (Gromov-Hausdorff distance). *Let X and Y be two metric spaces. The Gromov-Hausdorff distance between X and Y , denoted as $d_{GH}(X, Y)$, is defined as*

$$d_{GH}(X, Y) = \frac{1}{2} \inf_C dis(C),$$

where the infimum is over all correspondences between X and Y .

A.2 SIMPLICIAL COMPLEX AND NERVE THEOREM

One of the most commonly used simplicial complexes is the Vietoris-Rips complex, in which simplices are constructed based on pairwise distances between vertices.

Definition A.5 (Vietoris-Rips complex). *Let $\mathcal{X} \subset \mathbb{X}$ be finite and $r > 0$. The Vietoris-Rips complex $R_{\mathcal{X}}(r)$ is the simplicial complex defined as*

$$R_{\mathcal{X}}(r) := \{\sigma \subset \mathcal{X} : d(x_i, x_j) < 2r, \forall x_i, x_j \in \sigma\}. \quad (8)$$

Another common choice is the *Čech complex*, defined as below.

Definition A.6 (Čech complex). *Let $\mathcal{X} \subset \mathbb{X}$ be finite and $r > 0$. The (weighted) Čech complex is the simplicial complex*

$$\check{C}ech_{\mathcal{X}}^{\mathbb{X}}(r) := \{\sigma \subset \mathcal{X} : \bigcap_{x \in \sigma} \mathbb{B}_{\mathbb{X}}(x, r) \neq \emptyset\}, \quad (9)$$

The superscript \mathbb{X} will be dropped when understood from the context.

From equation 8 and equation 9, the Čech complex and Vietoris-Rips complex have the following interleaving inclusion relationship

$$\check{\text{Cech}}_{\mathcal{X}}(r) \subset R_{\mathcal{X}}(r) \subset \check{\text{Cech}}_{\mathcal{X}}(2r). \quad (10)$$

In particular, when \mathbb{X} is a Euclidean space, then the constant 2 can be tightened to $\sqrt{2}$:

$$\check{\text{Cech}}_{\mathcal{X}}(r) \subset R_{\mathcal{X}}(r) \subset \check{\text{Cech}}_{\mathcal{X}}(\sqrt{2}r). \quad (11)$$

Definition A.7 (Alpha Complex). *Let X be a finite set of points in \mathbb{R}^d . For each $u_i \in X$, the Voronoi cell of u_i is the set of points that are closest to u_i : $V_{u_i} = \{x \in \mathbb{R}^d | d(u_i, x) \leq d(u_j, x), \forall u_j \in X, u_j \neq u_i\}$. For $r > 0$ and each $u_i \in X$, let us denote the closed r -ball with center u_i and radius r as $B_{u_i}(r)$. Then, we define $R_{u_i}(r) = B_{u_i}(r) \cap V_{u_i}$, which is the intersection of each r -ball with its corresponding Voronoi cell. The Alpha complex is a collection of simplices such that all $R_{u_i}(r)$ of the vertices in the simplex have an intersection:*

$$\text{Alpha}(r) = \{\sigma \subset X | \cap_{u_i \in \sigma} R_{u_i}(r) \neq \emptyset\}.$$

Similar to the Vietoris-Rips complex, we can build a filtration on the Alpha complex by monotonically increasing r .

The topology of the Čech complex is linked to underlying continuous spaces via Nerve Theorem. Let $r > 0$ and consider the union of balls $\cup_{x \in \mathcal{X}} \mathbb{B}_{\mathbb{X}}(x, r)$. Then the union of balls is homotopic equivalent to the Čech complex by the following Nerve Theorem.

Theorem A.1 (Nerve Theorem). *Let $\mathcal{X} \subset \mathbb{X}$ be a finite set and $r > 0$. Suppose for any finite subset $\{x_1, \dots, x_k\} \subset \mathcal{X}$, the intersection $\bigcap_{j=1}^k \mathbb{B}_{\mathbb{X}}(x_j, r)$ is either empty or contractible, then the Čech complex $\check{\text{Cech}}_{\mathcal{X}}^{\mathbb{X}}(r)$ is homotopic equivalent to the union of balls $\cup_{x \in \mathcal{X}} \mathbb{B}_{\mathbb{X}}(x, r)$.*

A.3 PERSISTENT HOMOLOGY AND DIAGRAMS

Persistent homology is a multiscale approach to represent topological features. A *filtration* \mathcal{F} is a collection of subspaces approximating the data points at different resolutions, formally defined as follows.

Definition A.8. A filtration $\mathcal{F} = \{\mathcal{F}_a\}_{a \in \mathbb{R}}$ is a collection of subsets of \mathbb{X} such that $a \leq b$ implies that $\mathcal{F}_a \subset \mathcal{F}_b$.

For a filtration \mathcal{F} and for each $k \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, the associated persistent homology $PH_k \mathcal{F}$ is an ordered collection of k -th dimensional homologies, one for each element of \mathcal{F} .

Definition A.9. Let \mathcal{F} be a filtration and let $k \in \mathbb{N}_0$. The associated k -th persistent homology $PH_k \mathcal{F}$ is a collection of groups $\{H_k \mathcal{F}_a\}_{a \in \mathbb{R}}$ equipped with homomorphisms $\{\iota_k^{a,b}\}_{a \leq b}$, where $H_k \mathcal{F}_a$ is the k -th dimensional homology group of \mathcal{F}_a and $\iota_k^{a,b} : H_k \mathcal{F}_a \rightarrow H_k \mathcal{F}_b$ is the homomorphism induced by the inclusion $\mathcal{F}_a \subset \mathcal{F}_b$.

For the k -th persistent homology $PH_k \mathcal{F}$, the set of filtration levels at which a specific homology appears is always an interval $[b, d) \subset [-\infty, \infty]$, i.e. a specific homology is formed at some filtration value b and dies when the inside hole is filled at another value $d > b$.

Definition A.10. Let \mathcal{F} be a filtration and let $k \in \mathbb{N}_0$. The corresponding k -th persistence diagram $Dgm_k(\mathcal{F})$ is a finite multiset of $(\mathbb{R} \cup \{\infty\})^2$, consisting of all pairs (b, d) where $[b, d)$ is the interval of filtration values for which a specific homology appears in $PH_k \mathcal{F}$. b is called a birth time and d is called a death time.

Another widely used and powerful approach for transforming complex topological summaries into structured, vectorized representations suitable for statistical and machine learning applications is the use of *silhouettes* (Chazal et al., 2014; 2015).

Silhouette. Suppose $Dgm_k(\mathcal{F})$ contains N off-diagonal birth-death pairs $\{(b_j, d_j)\}_{j=1}^N$. We define the *silhouette* as the power-weighted function,

$$\phi^{(q)}(t) = \frac{\sum_{j=1}^N |d_j - b_j|^q \Lambda_j(t)}{\sum_{j=1}^N |d_j - b_j|^q}. \quad (12)$$

for $0 < q < \infty$. The value q controls the trade-off between uniformly treating all pairs and considering only the most persistent pair in $Dgm_k(\mathcal{F})$. When $q \rightarrow \infty$, silhouettes converge to the first order landscapes.

A.4 STABILITY THEOREMS

Stability theorems have been established for various cases.

First, we consider the case when the filtration \mathcal{F} is generated from the sub-level sets or the super-level sets of a function. Let $f, g : \mathbb{X} \rightarrow \mathbb{R}$ be two functions, and let $Dgm_*(f)$ and $Dgm_*(g)$ be the corresponding persistence diagrams of the sublevel set filtrations $\{f \leq L\}_{L \in \mathbb{R}}$ and $\{g \leq L\}_{L \in \mathbb{R}}$.

We will impose a standard regularity condition for functions f and g , which is *tameness*.

Definition A.11 (tameness). ((Chazal et al., 2016b, Section 3.8)) Let $f : \mathbb{X} \rightarrow \mathbb{R}$. Then f is tame if the image $\text{im}(\iota_L^{L'})$ of the homomorphism $\iota_L^{L'} : H_k(f^{-1}(-\infty, L]) \rightarrow H_k(f^{-1}(-\infty, L'])$ induced from the inclusion $\iota_L^{L'} : f^{-1}(-\infty, L] \rightarrow f^{-1}(-\infty, L']$ is of finite rank for all $k \in \mathbb{N} \cup \{0\}$ and $L < L'$.

When two functions f and g satisfy the tameness condition, their bottleneck distance is bounded by their ℓ_∞ distance, an important and useful fact known as the stability theorem.

Theorem A.2 (Stability theorem for sublevel or superlevel sets of a function). Cohen-Steiner et al. (2007); Chazal et al. (2016b) For two tame functions $f, g : \mathbb{X} \rightarrow \mathbb{R}$,

$$d_B(Dgm_k(f), Dgm_k(g)) \leq \|f - g\|_\infty.$$

Second, we consider the case for Vietoris-Rips filtration. Let X and Y be two metric spaces, and let $Dgm_k(R_X)$ and $Dgm_k(R_Y)$ be the corresponding persistence diagrams of the Vietoris-Rips complex filtrations $\{R_X(r)\}_{r \in \mathbb{R}}$ and $\{R_Y(r)\}_{r \in \mathbb{R}}$.

We say that the metric space is totally bounded if it can be arbitrarily approximated by a finite set of points. For example, a bounded subset of Euclidean space is totally bounded.

Definition A.12. A metric space X is totally bounded if for any $\epsilon > 0$, there exists a finite set of points $x_1, \dots, x_n \in X$ that ϵ -approximates X , i.e. for all $x \in X$, there exists x_i such that $d(x, x_i) < \epsilon$.

Now, when two metric spaces X and Y are totally bounded, then their bottleneck distance between the corresponding Vietoris-Rips persistence homologies is bounded by their Gromov-Hausdorff distance, as the following stability theorem.

Theorem A.3 (Stability theorem for Vietoris-Rips complex). (Chazal et al., 2012, Theorem 5.2) Let X and Y be two totally bounded metric spaces. Then,

$$d_B(Dgm_k(R_X), Dgm_k(R_Y)) \leq d_{GH}(X, Y).$$

A.5 DISTANCE TO MEASURE.

The Distance to measure (DTM) (Chazal et al., 2011; 2016a) is a distance-like function that exhibits robustness to outliers. For a probability measure μ and parameters $m_0 \in [0, 1]$ and $r \geq 1$ (default is $r = 2$), the DTM function $d_{\mu, m_0} : \mathbb{R}^d \rightarrow \mathbb{R}$ is defined as

$$d_{\mu, m_0}(x) = \left(\frac{1}{m_0} \int_0^{m_0} \delta_{\mu, m}^r(x) dm \right)^{1/r},$$

where $\delta_{\mu, m}(x) = \inf\{t > 0 \mid \mu(B_x(t)) > m\}$ and $B_x(t)$ is a closed t -ball centered at x . Given data points X_1, \dots, X_n , an empirical version of the DTM is

$$\hat{d}_{m_0}(x) = \left(\frac{\sum_{X_i \in N_k(x)} w'_i \|X_i - x\|^r}{m_0 \sum_{i=1}^n w_i} \right)^{1/r} \quad (13)$$

where $N_k(x)$ is a subset of $\{X_1, \dots, X_n\}$ containing the k nearest neighbors of x . k is such that satisfies $\sum_{X_i \in N_{k-1}(x)} w_i < m_0 \sum_{i=1}^n w_i \leq \sum_{X_i \in N_k(x)} w_i$, and $w'_i = \sum_{X_j \in N_k(x)} w_j - m_0 \sum_{j=1}^n w_j$ if at least one of X_i 's is in $N_k(x)$ and $w'_i = w_i$ otherwise.

The parameter m_0 controls the scale at which features are extracted: smaller values of m_0 correspond to more local structures, while larger values capture more global ones. The DTM function is differentiable (Kim et al., 2020), and applying a sublevel or superlevel set filtration on the image of the DTM function yields a DTM filtration that is robust to outliers.

B A BRIEF REVIEW OF TAKENS' TIME-DELAY EMBEDDING THEOREM

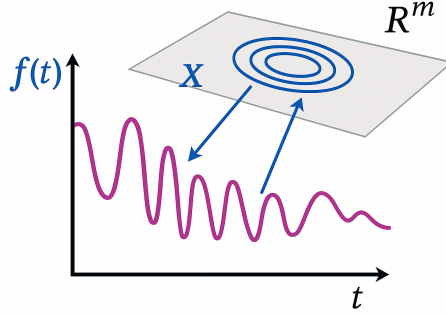


Figure 2: Illustration of the reconstructed attractor in the manifold in \mathbb{R}^m obtained through the trajectory matrix $X_{m,\tau}$, from the original time series observations $f(t)$.

A time series can be viewed as a sequence of projections from an underlying *dynamical system*, defined as a rule governing the temporal evolution of states in a system’s state space. Reconstructing the governing transition rules from observed time series is essential for understanding the underlying phenomena Packard et al. (1980). However, in the absence of prior knowledge, this reconstruction is generally difficult and often ill-posed Tong (1990); Eckmann & Ruelle (1992). Consequently, analyses frequently focus on *attractors*—sets of numerical values toward which a dynamical system evolves over time, largely independent of initial conditions Tong (1990); Packard et al. (1980); Eckmann & Ruelle (1992). Notably, two time series governed by the same transition rule may exhibit distinct observed waveforms, yet their corresponding attractors remain topologically similar. Studying attractors therefore offers a principled approach for modeling and comparing dynamical systems.

In principle, constructing an attractor requires infinitely many observations. In practice, however, we use a *quasi-attractor*, which can be estimated from a finite sample. One of the most widely adopted techniques for this purpose is time-delay embedding, including Takens’ embedding framework Takens (1981). Let \mathcal{M}_0 denote the manifold corresponding to the original dynamical system generating the observed time series. Takens’ theorem guarantees the existence of a smooth embedding $\Psi : \mathcal{M}_0 \rightarrow \mathcal{M}'$, where $\mathcal{M}' \subset \mathbb{R}^m$, such that \mathcal{M}_0 and \mathcal{M}' are topologically equivalent. Sauer et al. (1991) further showed that this embedding is valid if $m > d_0$, where d_0 is the box-counting dimension of the attractor on \mathcal{M}_0 . Robinson (2014) extended this result to infinite-dimensional systems with finite-dimensional attractors.

Although this implies that m must be sufficiently large relative to d_0 , the dimension d_0 is typically unobservable. As a practical alternative, the number of significant singular values of the delay-embedded matrix X can be used as a proxy, as suggested by Palu & Dvoak (1992); Torcu (2016). Accordingly, m is often selected as the smallest integer greater than or equal to this estimated rank. The choice of delay τ is similarly heuristic, often informed by autocorrelation or mutual information Small (2003). As a result, both m and τ are typically determined using data-driven, empirical approaches. For further discussion, see Bradley & Kantz (2015); Garland et al. (2016). Kim et al. (2020) used a simple heuristic approach to select optimal values for m and τ .

Hence, once Ψ is obtained, one can study \mathcal{M}' as a topologically equivalent representation to analyze the underlying dynamics of the given time series. This result has gained popularity as it enables the extraction of meaningful structural information from time series data within a Euclidean space.

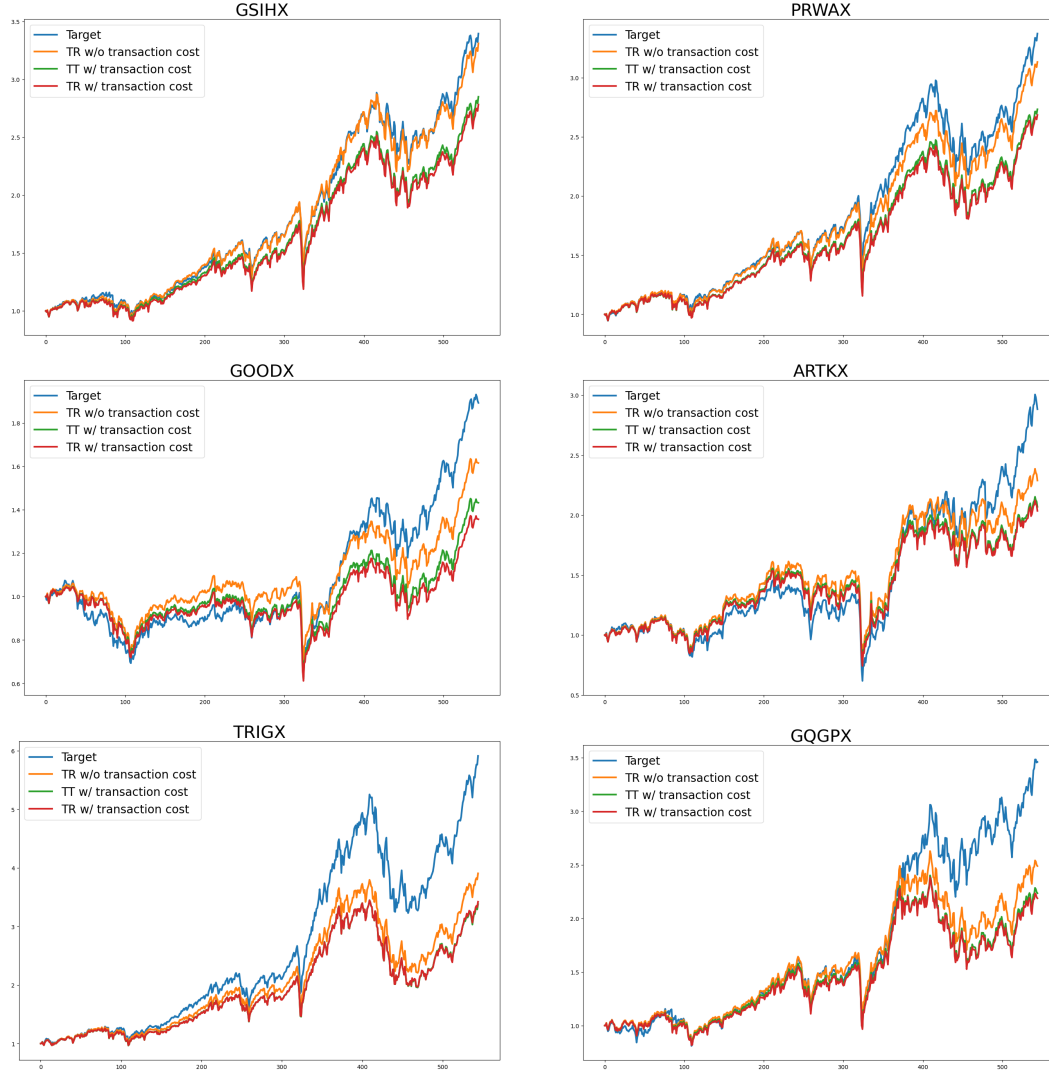


Figure 3: Net asset value of target index and model predictions over entire evaluation period.

C EXPERIMENT DETAILS AND VISUALIZATION OF THE INDEX TRACKING PERFORMANCE

In this section, we provide additional details about our simulations, with a focus on performance visualizations.

Optimizer and Computing resources. We use `Pytorch` (Paszke, 2019), which provides automatic differentiation for gradient-based optimization. All experiments were conducted on a Mac Studio equipped with an Apple M4 Max chip, featuring 32 CPU cores and 64 GPU cores, and 512GB memory.

Choice of TDA hyperparameters. Selecting appropriate hyperparameter values for the topological regularizer is crucial to ensure that the underlying structure of the time series data is effectively captured. For each step in Algorithm 1, the hyperparameters that require specification include

1. Time-delay embedding: (i) delay parameter τ , and (ii) embedding dimension m .
2. Computation of persistence diagram: (iii) choice of filtration.
3. Computation of persistence landscape: (iv) maximum homology dimension K_{max} , (v) number of landscapes J_{max} , (vi) interval $[0, L_{max}]$, and (vii) resolution κ

Another hyperparameter that must be determined is (viii) γ , which governs the strength of the topological regularization term. While determining the optimal hyperparameters is a non-trivial task, standard heuristics such as grid search or random search can be employed to identify suitable values that yield enhanced performance. In this experiment, a hybrid approach combining grid search and random search is implemented to determine effective hyperparameter configurations. The selection process for each of the hyperparameters is outlined below.

1. τ : grid search over $\{2, 5, 10\}$.
2. m : fixed at 2 for simplicity.
3. filtration: DTM filtration in (13).
 - m_0 : fixed at 0.01.
 - domain of x : grid search over $\{[-0.2, 0.2]^2, [-0.15, 0.15]^2\}$.
4. K_{max} : fixed at 1.
5. J_{max} : fixed at 2.
6. $[0, L_{max}]$: grid search over $\{[0, 0.02], [0, 0.03]\}$.
7. κ : fixed at a value that discretizes the interval into 52 grid points.
8. γ : random search using $\gamma \sim \text{Uniform}(0, 1)$.

For each target fund, we choose the hyperparameters that yield the best performance under the evaluation criteria in Section 5.

Visualization. We provide a visualization comparing our model’s predictions to the target index in terms of net asset value. We consider three scenarios: (i) TR model in the absence of transaction cost, (ii) TT model with transaction cost, and (ii) TR model with transaction cost. The results are illustrated in Figure 3.

D PROOFS

D.1 PROOF OF LEMMA 3.1

Proof. For notational simplicity, for the true signal f , we define $\mathbb{X} := \{SW_{m,\tau}f(t) : t \in [0, T]\}$ and write the corresponding point cloud $X \equiv X_{m,\tau}^f$, and let X' denote the observed point cloud generated by the trajectory matrix equation 3 $X_{m,\tau}^{f+\zeta}$. Moreover, since our result holds for any fixed homological dimension k , we simplify the notation by writing $Dgm_k(X)$, $Dgm_k(X')$, and $Dgm_k(f)$ as D_X , $D_{X'}$, and D_f , respectively. Furthermore, let X_i and X'_i denote the i th rows of X and X' , respectively. Define $\tilde{N} := N - (m - 1)\tau$, and interpret X and X' as point clouds represented by $\{X_i\}_{i \leq \tilde{N}}$ and $\{X'_i\}_{i \leq \tilde{N}}$, respectively.

By the triangle inequality,

$$d_B(D_{X'}, D_f) \leq d_B(D_{X'}, D_X) + d_B(D_X, D_f). \quad (14)$$

For the first term, it follows that

$$d_B(D_X, D_{X'}) \leq d_H(X, X') \leq \sup_i \|X_i - X'_i\|_2 \leq \sqrt{m} \|\zeta\|_\infty. \quad (15)$$

Next, we consider the second term in equation 14, which arises due to time discretization. Since f is assumed to be L_f -Lipschitz continuous, i.e., $|f(t_1) - f(t_2)| \leq L_f|t_1 - t_2|$, $\forall t_1, t_2 \in [0, T]$, it follows that

$$\begin{aligned} \|SW_{m,\tau}f(t_1) - SW_{m,\tau}f(t_2)\|_2 &= \sqrt{\sum_{k=0}^{m-1} |f(t_1 + k(m-1)\tau) - f(t_2 + k(m-1)\tau)|^2} \\ &\leq \sqrt{m} L_f |t_1 - t_2|, \end{aligned}$$

which implies that $SW_{m,\tau}f$ is Lipschitz continuous with constant $\sqrt{m}L_f$.

Since X is obtained by sampling \mathbb{X} with grid size $\frac{T}{N}$, the Hausdorff distance between X and \mathbb{X} is bounded by

$$d_H(X, \mathbb{X}) \leq \frac{\sqrt{m}L_f T}{N}.$$

Since both X and \mathbb{X} are bounded subsets of a Euclidean space, we may apply the above bound to Theorem A.3, which yields the following upper bound on the bottleneck distance between the corresponding persistence diagrams:

$$d_B(D_X, D_f) \leq \frac{\sqrt{m}L_f T}{N}. \quad (16)$$

Hence, by applying equation 15 and equation 16 to equation 14, we obtain the desired bound as

$$d_B(D_{X'}, D_f) \leq \sqrt{m}\|\zeta\|_\infty + \frac{\sqrt{m}L_f T}{N}.$$

□

D.2 PROOF OF THEOREM 3.2

Proof. The result is immediate by Theorem A.1 in Bubenik (2015) and Lemma 3.1, as

$$\begin{aligned} \|\lambda_{k,X'} - \lambda_{k,f}\|_\infty &\leq d_B(D_{X'}, D_f) \\ &\leq \sqrt{m}\|\zeta\|_\infty + \frac{\sqrt{m}L_f T}{N}. \end{aligned}$$

□

D.3 PROOF OF LEMMA 3.3

Proof. Let $\mathbb{X}^{f+\zeta} := \{SW_{m,\tau}(f + \zeta)(t) : t \in [0, T]\}$. Note that $f + \zeta$ is a Lipschitz function with Lipschitz constant $L_f + L_\zeta$, i.e.,

$$|(f + \zeta)(t_1) - (f + \zeta)(t_2)| \leq (L_f + L_\zeta) |t_1 - t_2|.$$

Since $X_{m,\tau,N_1}^{f+\zeta}$ is obtained by sampling $\mathbb{X}^{f+\zeta}$ with grid size $\frac{T}{N_1}$, and $X_{m,\tau,N_2}^{f+\zeta}$ is obtained by sampling with grid size $\frac{T}{N_2}$, the Hausdorff distance between $X_{m,\tau,N_1}^{f+\zeta}$ and $X_{m,\tau,N_2}^{f+\zeta}$ is bounded by

$$d_H(X_{m,\tau,N_1}^{f+\zeta}, X_{m,\tau,N_2}^{f+\zeta}) \leq \frac{\sqrt{m}(L_f + L_\zeta)T}{\min\{N_1, N_2\}}.$$

Hence applying Theorem A.3 yields the following upper bound on the bottleneck distance between the corresponding persistence diagrams:

$$d_B(Dgm_k(X_{m,\tau,N_1}^{f+\zeta}), Dgm_k(X_{m,\tau,N_2}^{f+\zeta})) \leq \frac{\sqrt{m}(L_f + L_\zeta)T}{\min\{N_1, N_2\}}.$$

Similarly, since $X_{m,\tau,N}^{f+\zeta}$ is obtained by sampling $\mathbb{X}^{f+\zeta}$ with grid size $\frac{T}{N}$, the Hausdorff distance between $X_{m,\tau,N}^{f+\zeta}$ and $\mathbb{X}^{f+\zeta}$ is bounded by

$$d_H(X_{m,\tau,N}^{f+\zeta}, \mathbb{X}^{f+\zeta}) \leq \frac{\sqrt{m}(L_f + L_\zeta)T}{N}.$$

Hence applying Theorem A.3 yields the following upper bound on the bottleneck distance between the corresponding persistence diagrams:

$$d_B(Dgm_k(X_{m,\tau,N}^{f+\zeta}), Dgm_k(f + \zeta)) \leq \frac{\sqrt{m}(L_f + L_\zeta)T}{N}.$$

□

D.4 PROOF OF THEOREM 3.4

Proof. The result is immediate by Theorem A.1 in Bubenik (2015) and Lemma 3.3, as

$$\begin{aligned} \left\| \lambda_{j, X_{m, \tau, N_1}^{f+\zeta}} - \lambda_{j, X_{m, \tau, N_2}^{f+\zeta}} \right\|_{\infty} &\leq d_B(Dgm_k(X_{m, \tau, N_1}^{f+\zeta}), Dgm_k(X_{m, \tau, N_2}^{f+\zeta})) \\ &\leq \frac{\sqrt{m}(L_f + L_{\zeta})T}{\min\{N_1, N_2\}}, \end{aligned}$$

and

$$\begin{aligned} \left\| \lambda_{j, X_{m, \tau, N}^{f+\zeta}} - \lambda_{j, f+\zeta} \right\|_{\infty} &\leq d_B(Dgm_k(X_{m, \tau, N}^{f+\zeta}), Dgm_k(f + \zeta)) \\ &\leq \frac{\sqrt{m}(L_f + L_{\zeta})T}{N}. \end{aligned}$$

□

D.5 PROOF OF PROPOSITION 4.1

Proof. We provide only a sketch of the proof, omitting algebraic details that are not particularly illuminating. It is immediate to see that $\mathcal{L}(w)$ is differentiable almost everywhere. Let $z := Rw$ and let Z be the point cloud embedding of z corresponding to (3). Given a simplicial complex K and $\tau, \sigma \in K$, consider a monotonic filtration function $f : K \rightarrow \mathbb{R}$ such that $f(\tau) \leq f(\sigma)$ whenever τ is a face of σ . By defining $K(a) := f^{-1}(-\infty, a]$, we can obtain a filtration $\mathcal{F} = \{K(a) | a \in \mathbb{R}\}$ of nested simplicial complexes. Accordingly, let us denote the filtration constructed from Z using a filtration function f as \mathcal{F}_Z . Then, using chain rule, the gradient can be expressed as

$$\frac{\partial \mathcal{R}_{top}(y, Rw)}{\partial w} = \frac{\partial z}{\partial w} \frac{\partial Z}{\partial z} \sum_{k=0}^{K_{\max}} \sum_{j=1}^{J_{\max}} \frac{\partial Dgm_k(\mathcal{F}_Z)}{\partial Z} \frac{\partial \lambda_{j,z}^k}{\partial Dgm_k(\mathcal{F}_Z)} \frac{\partial \|\lambda_{j,y}^k - \lambda_{j,z}^k\|_2^2}{\partial \lambda_{j,z}^k}.$$

Since all other gradient terms exist almost everywhere, it remains to verify the existence of the derivative $\frac{\partial Dgm_k(\mathcal{F}_Z)}{\partial Z} = \left\{ \left(\frac{\partial b_i}{\partial Z}, \frac{\partial d_i}{\partial Z} \right) \right\}_{(b_i, d_i) \in Dgm_k(\mathcal{F}_Z)}$. For every birth-death point $(b_i, d_i) \in Dgm_k(\mathcal{F}_Z)$, there exists a birth simplex τ_i and a death simplex σ_i that forms and destroys the corresponding k -dimensional homological feature. Such simplices satisfy $f(\tau_i) = b_i$ and $f(\sigma_i) = d_i$, which yields $\frac{\partial Dgm_k(\mathcal{F}_Z)}{\partial Z} = \left\{ \left(\frac{\partial f(\tau_i)}{\partial Z}, \frac{\partial f(\sigma_i)}{\partial Z} \right) \right\}_{(b_i, d_i) \in Dgm_k(\mathcal{F}_Z)}$. Consequently, $\mathcal{R}_{top}(y, Rw)$ is differentiable as long as the filtration function is differentiable. □

D.6 PROOF OF PROPOSITION 4.2

Lemma D.1. Given $f_b, f_s, w_{prev} \in [0, 1]^N$, let $\mathcal{T} : [0, 1]^N \rightarrow [0, 1]$ be

$$\mathcal{T}(w) = (f_b \odot b - f_s \odot (1 - b))^{\top} (w - w_{prev}),$$

where $b = \mathbb{1}(w > w_{prev})$. Then \mathcal{T} is a convex function.

Proof for Lemma D.1. Rewrite $\mathcal{T}(w)$ as

$$\mathcal{T}(w) = \max \left\{ (f_b \odot b)^{\top} (w - w_{prev}), (f_s \odot (1 - b))^{\top} (w_{prev} - w) \right\}.$$

Then $\mathcal{T}(w)$ is maximum of two affine functions, and hence convex. □

Proof of Proposition 4.2. \mathcal{C} can be written as

$$\mathcal{C} = \{w \in [0, 1]^N : C_t \mathcal{T}(w) \leq \delta\} = \mathcal{T}^{-1} \left(-\infty, \frac{\delta}{C_t} \right].$$

Then, by Lemma D.1, \mathcal{T} is a convex function. Since \mathcal{C} is a sublevel set of a convex function, it follows that \mathcal{C} is a convex set. □