GEODESIC PROPERTIES OF A GENERALIZED WASSERSTEIN EMBEDDING FOR TIME SERIES ANALYSIS

Shiying Li*  
Department of Biomedical Engineering  
University of Virginia  
Charlottesville, VA 22908, USA  
sl8jx@virginia.edu

Abu Hasnat Mohammad Rubaiyat  
Department of Electrical and Computer Engineering  
University of Virginia  
Charlottesville, VA 22904, USA  
ar3fx@virginia.edu

Gustavo K. Rohde  
Department of Biomedical Engineering & Department of Electrical and Computer Engineering  
Charlottesville, VA 22908, USA  
gustavo@virginia.edu

ABSTRACT

Transport-based metrics and related embeddings (transforms) have recently been used to model signal classes where nonlinear structures or variations are present. In this paper, we study the geodesic properties of time series data with a generalized Wasserstein metric and the geometry related to their signed cumulative distribution transforms in the embedding space. Moreover, we show how understanding such geometric characteristics can provide added interpretability to certain time series classifiers, and be an inspiration for more robust classifiers. The appendix can be found at https://arxiv.org/abs/2206.01984.

1 INTRODUCTION

Transport-based distances, such as Wasserstein distances [Villani (2003)], have been shown to be an effective tool in signal analysis and machine learning applications including image retrieval [Rubner et al. (2000)] and registration [Haker et al. (2004)], modeling biological morphology [Ozolek et al. (2014); Basu et al. (2014)], comparing probability distributions [Arjovsky et al. (2017)], and providing good low-dimensional embeddings of image manifolds [Hamm et al. (2022)], to name a few. The success of transport-based distances are in part due to their ability to incorporate information of spatial or time deformations naturally structured in signals [Kolouri et al. (2017)]. For example, they are observed to correctly recover biologically interpretable and statistically significant differences [Basu et al. (2014)] and quantify semantic differences between distributions that correlate well with human perception [Rubner et al. (2000)].

Within the set of problems in data science, modeling of time series data is considered a challenging problem. Deformation-based methods such as dynamic time warping [Abanda et al. (2019); Lines & Bagnall (2015)] have been shown successful in enabling the comparison of time series data more meaningfully. In addition to being able to align features from two different time series, Wasserstein-type distances are also true distances that can allow for a low-dimensional representation of dynamical systems in which time series can be classified and statistically analyzed [Muskulus & Verduyn-Lunel (2011)]. In recent years, many transport transform-based techniques have been developed to leverage Wasserstein distances and linearized ($L^2$) embeddings to facilitate the application of many standard data analysis [Kolouri et al. (2017)]. In particular, the cumulative distribution transform (CDT), based on the 1D Wasserstein embedding, was introduced in [Park et al. (2018)] as a means of classifying normalized non-negative signals, and has been extended to general signed signals via the signed cumulative distribution transform (SCDT) in [Aldroubi et al. (2022)].

Wasserstein embeddings based on the cumulative distribution transform (CDT) [Park et al. (2018); Rubaiyat et al. (2020); Aldroubi et al. (2022)] have recently emerged as a robust, computationally

*This author is currently with the University of North Carolina, Chapel Hill (email: shiyl@unc.edu).
efficient, and accurate end-to-end classification method for time series (1D signal) classification. They are particularly effective for classifying data emanating from physical processes where signal classes can be modeled as observations of a particular set of template signals under some unknown, possibly random, temporal deformation or transportation [Park et al. (2018); Shifat-E-Rabbi et al. (2021); Rubaiyat et al. (2022b)]. Efforts have been made to explain the success of these models by understanding the geometry of the transform embedding space [Park et al. (2018); Aldroubi et al. (2021); Moosmüller & Cloninger (2020)], where embedding properties and conditions when the data class becomes convex and linearly separable in the transform space are studied. In a nutshell, the template-deformation-based generative models capture the nonlinear structure of signals and the nonlinear transport transforms render signal classes that are nonlinear and non-convex into convex sets in transform embedding space (see Figure 1).

As the CDT [Park et al. (2018)] is defined for probability measures (or their associated density functions) which are non-negative and normalized, in this manuscript we elucidate the geometry of general time series (signed, non-normalized) with a generalized Wasserstein metric in the SCDT embedding space with a $L^2$-type metric. In addition to the convexity and isometric embedding properties of the SCDT [Aldroubi et al. (2022)], which are shared by the CDT, we look at the geodesic properties of the SCDT and illustrate the differences between geometry of data in CDT and SCDT space. In particular, the SCDT embedding $\tilde{S} \subseteq (L^2(s_0) \times \mathbb{R})^2$ is not a geodesic space while geodesics exist between signals within a generative model (see Figure 1). As a preliminary application, we provide an interpretation of the nearest (transform) subspace classifiers proposed in [Rubaiyat et al. (2022a,b)] through visualizing paths between the test signals and their projections to various subspaces. In particular, we illustrate that using the training samples, these classifiers “correctly” generate a (local) subspace that models the generative clusters (see equation 9) to which the given test signals belong. We hypothesise this knowledge can lead to the design of more efficient and accurate pattern recognition tools and added interpretability of various classifiers.

Figure 1: Space of time series with a generalized Wasserstein metric (top) and its embedding $\tilde{S}$ (bottom) are not geodesic spaces while geodesics exist between any pair of signals in the same generative cluster $S_{p,G}$. Here $\tilde{S}$ is represented as the faces of the polyhedron (see Remark 2.7 and 2.8) and its ambient space $(L^2(s_0) \times \mathbb{R})^2$ is represented as the open cube.

Throughout the manuscript, we work with $L^1$ signals $s$ with finite second moments, where $\Omega_s \subseteq \mathbb{R}$ is the bounded domain over which $s$ is defined. We denote $S$ as the set of $L^1$ signals with finite second moments and $S_1$ as the set of non-negative $L^1$-normalized signals in $S$ and $\|s\|$ as the $L^1$-norm of $s$.

\[\text{For bounded } \Omega_s, \text{ it suffices to require that } s \in L^2(\Omega_s), \text{ in which case by Cauchy-Schwartz } \int_{\Omega_s} t^2|s(t)|dt < \infty; s \in L^1(\Omega_s) \text{ follows from the fact that } L^2(\Omega_s) \subseteq L^1(\Omega_s).\]
2 Transport Transforms and Geometric Properties of Time Series

We give a brief overview of the CDT and SCDT and their associated embedding properties in Section 2.1 and present the main theorems about the geodesic properties of time series with respect to a generalized Wasserstein metric and the geometry of the SCDT embedding in Section 2.2.

2.1 The Signed Cumulative Distribution Transform and a Generalized Wasserstein Metric

The Cumulative Distribution Transform (CDT) was introduced in Park et al. (2018) for non-negative maps \( s \) normalized version, i.e., we denote \( \tilde{s} := s \) for \( s \geq 0 \) and \( \hat{s} := s \) for \( s \leq 0 \). In particular, \( s^* \) is defined as the optimal transport map between reference \( s_0 \) and \( s \), which is the unique non-decreasing map \( \tilde{s} \) between reference \( s_0 \) and \( s \), given by \( \tilde{s} = F_t^* \circ F_{s_0} \),

\[
s^* = F_t^* \circ F_{s_0},
\]

where \( F_{s_0} \) is the cumulative distribution function of \( s_0 \) and \( F_t^* := \inf \{ t \in \mathbb{R} : F(t) \geq x \} \) is the generalized inverse of \( F_t \). In particular, \( (s^*)^t s_0 = s \) and \( s^* \) minimizes the Wasserstein-2 cost \( d_{W^2}(s_0, s) := \sqrt{\int_{s = T_{s_0}} s_0(x) dx} \) between \( s_0 \) and \( s \). It is well-known (Santambrogio 2015) that \( (S_1, d_{W^2}(\cdot, \cdot)) \) is a geodesic space with the constant-speed geodesic between \( s \) and \( \tilde{s} \) given by for \( \alpha \mapsto p_{\alpha} = \left( (1 - \alpha)id + \alpha T^\ast \right)s \), where \( T^\ast \) is the optimal transport map between \( s \) and \( \tilde{s} \) in \( S_1 \) and \( \alpha \in [0, 1] \). Moreover, the CDT defines an isometric embedding from the space of non-negative normalized signals \( (S_1^+, L^2(s_0)) \) to the transform space \( (S_1^+, L^2(s_0)) \) (cf. Villani 2003), i.e.,

\[
d_{W^2}(s_1, s_2) = \| s_1^* - s_2^* \|_{L^2(s_0)}
\]

where \( S_1^+ \) is the set of CDTs of signals in \( S_1 \).

Now for a (non-zero) signed signal, the Jordan decomposition (Royden & Fitzpatrick 1988) is applied to \( s(t) = s^+(t) - s^-(t) \), where \( s^+(t) \) and \( s^-(t) \) are the absolute values of the positive and negative parts of the signal \( s(t) \). Given a fixed \( L^1 \)-normalized positive reference signal \( s_0 \) defined on \( \Omega_{s_0} \), the signed cumulative distribution transform (SCDT) (Aldroubi et al. 2022) of \( s(t) \) is then defined as the following tuple:

\[
\hat{s} := \left( \left( \frac{s^+}{\| s^+ \|} \right)^*, \| s^+ \|, \left( \frac{s^-}{\| s^- \|} \right)^*, \| s^- \| \right),
\]

where \( \left( \frac{s^+}{\| s^+ \|} \right)^* \) is the CDT (defined in eqn. (1)) of the normalized signal \( \frac{s^+}{\| s^+ \|} \). To simplify notations, from now on, for non-negative signals (e.g., \( s^+ \)), its CDT is defined as the CDT of its normalized version, i.e., we denote \( (\frac{s^+}{\| s^+ \|})^* \) simply as \( (s^+)^* \). Equivalently, equation 3 becomes \( \hat{s} = \left( (s^+)^*, \| s^+ \|, (s^-)^*, \| s^- \| \right) \). Like the CDT, the SCDT is isometric with the inverse transform (see A.5 for more details) given by

\[
s = \| s^+ \| \left( (s^+)^* s_0 \right) - \| s^- \| \left( (s^-)^* s_0 \right).
\]

The fact that \( s_0 \) and \( s \) have finite second moments guarantees the existence of a unique optimal transport map between them.

If \( F_t \) is strictly increasing on \( \Omega_s \), then \( F_t^* = F_{s_0}^{-1} \).

Here \( s^+(t) \) and \( s^-(t) \) can be seen as the density functions of corresponding measures in the Jordan decomposition of the measure \( \mu_s \) associated with \( s \) where \( d\mu_s(x) := s(x) dx \).

When \( s = 0 \), the CDT \( s^* \) is defined to be 0 and the SCDT \( \hat{s} := (0, 0, 0, 0) \).

Here \( (s^+)^* s_0 \) denote the push-forward signal of \( s_0 \) by the maps \( (s^+)^* \) respectively. In particular, when the maps \( (s^+)^* \) are differentiable, the inverse formula becomes

\[
s(t) = \| s^+ \| \left( (s^+)^{-1} \right)'(t) s_0 \left( (s^+)^{-1} (t) \right) - \| s^- \| \left( (s^-)^{-1} \right)'(t) s_0 \left( (s^-)^{-1} (t) \right).
\]
Similarly, a generalized Wasserstein-2 distance between two time series can be defined as follows:

\[ D_S(s_1, s_2) := \left( d_{W^2} \left( \frac{s_1^+}{\|s_1^+\|}, \frac{s_2^+}{\|s_2^+\|} \right) + d_{W^2} \left( \frac{s_1^-}{\|s_1^-\|}, \frac{s_2^-}{\|s_2^-\|} \right) + \left( \|s_1^+\| - \|s_2^+\| \right)^2 + \left( \|s_1^-\| - \|s_2^-\| \right)^2 \right)^{\frac{1}{2}} \]

(6)

\[ = \left( \|s_1^+\|^2_{L^2(s_0)} + \|s_1^-\|^2_{L^2(s_0)} + \|s_2^+\|^2_{L^2(s_0)} + \|s_2^-\|^2_{L^2(s_0)} + \left( \|s_1^+\| - \|s_2^+\| \right)^2 + \left( \|s_1^-\| - \|s_2^-\| \right)^2 \right)^{\frac{1}{2}} \]

(7)

where the second-to-last equality follows from the embedding property equation of the CDT. Hence the SCDT also defines an isometry (embedding) from the space \( (\tilde{S}, D_S(\cdot, \cdot)) \) of time series to its transform space \( \tilde{S} \times (\cdot, \cdot)((\tilde{S}, s_0)) \).

### 2.2 Geodesic Properties

We first present a sufficient condition on a pair of signals under which a geodesic exists between them.

**Theorem 2.1.** Fix a reference signal \( s_0 \in S_1 \). Given a signal \( s \in S \) and a strictly increasing differentiable function \( g : \mathbb{R} \to \mathbb{R} \), there is a constant speed geodesic between \( s \) and \( \bar{s} := g's \circ g \). In particular, the geodesic \( g \mapsto p_{\alpha} \) is given by \( \forall \alpha \in [0, 1], \)

\[ p_{\alpha} = [(1 - \alpha)\|s^+\| + \alpha\|\bar{s}^+\|](1 - \alpha)(s^+)^* + \alpha(\bar{s}^+)^* s_0 \]

\[ -[(1 - \alpha)\|s^-\| + \alpha\|\bar{s}^-\|](1 - \alpha)(s^-)^* + \alpha(\bar{s}^-)^* s_0. \]

(8)

Note that \( p_0 = s \) and \( p_1 = \bar{s} \).

**Proof.** It suffices to show that given \( \alpha_1, \alpha_2 \in [0, 1], D_S(p_{\alpha_1}, p_{\alpha_2}) = |\alpha_1 - \alpha_2|D_S(s, \bar{s}) \). By the composition property of the CDT (see A.4), \( (\bar{s}^\pm)^* = g^{-1} \circ (s^\pm)^* \). Hence we have that

\[ (1 - \alpha)(s^+)^* + \alpha(\bar{s}^+)^* = \left( (1 - \alpha)id + \alpha g^{-1} \right) \circ (s^\pm)^*. \]

Since \((1 - \alpha)id + \alpha g^{-1}\) is strictly increasing and \((s^\pm)^* s_0 \perp (\bar{s}^\pm)^* s_0 \) (see A.5), it follows that

\[ \left( (1 - \alpha)(s^+)^* + \alpha(\bar{s}^+)^* \right) s_0 \perp \left( (1 - \alpha)(s^-)^* + \alpha(\bar{s}^-)^* \right) s_0 \]

(cf. Lemma 5.4 in Aldroubi et al. (2022)). By the inverse formula in Proposition A.3, it is not hard to see that the expression in equation is the Jordan decomposition of \( p_{\alpha_1}, i.e.,\)

\[ p_{\alpha_i}^\pm = [(1 - \alpha)\|s^\pm\| + \alpha\|\bar{s}^\pm\|]

\[ \left( (1 - \alpha)(s^+)^* + \alpha(\bar{s}^+)^* \right) s_0. \]

Hence

\[ D_S^2(p_{\alpha_1}, p_{\alpha_2}) = D_S^2(p_{\alpha_1}^+, p_{\alpha_2}^+) + D_S^2(p_{\alpha_1}^-, p_{\alpha_2}^-) \]

\[ = \|p_{\alpha_1}^+\|^2_{L^2(s_0)} + \|p_{\alpha_2}^+\|^2_{L^2(s_0)} + \left( \|p_{\alpha_1}^-\| - \|p_{\alpha_2}^-\| \right)^2 + \left( \|p_{\alpha_1}^-\| - \|p_{\alpha_2}^-\| \right)^2 \]

\[ = |\alpha_1 - \alpha_2|^2 \left( \|s^+\|^2_{L^2(s_0)} + \|s^-\|^2_{L^2(s_0)} + \left( \|s^+\| - \|\bar{s}^+\| \right)^2 + \left( \|s^-\| - \|\bar{s}^-\| \right)^2 \right) \]

\[ = |\alpha_1 - \alpha_2|^2 D_S^2(s, \bar{s}), \]

where the second-to-last equality follows from the fact that \( (p_{\alpha_i}^\pm)^* = (1 - \alpha_i)(s^\pm)^* + \alpha_i(\bar{s}^\pm)^*, i = 1, 2 \) (see A.5 for more details).

**Remark 2.2.** The same conclusion holds if \( \bar{s} := \lambda g's \circ g \) for any constant \( \lambda > 0 \) and \( g \) strictly increasing and differentiable.

**Corollary 2.3.** Let \( G = \{g : \mathbb{R} \to \mathbb{R} : g \) is strictly increasing and differentiable\} and \( p \in S \). Consider the following template-deformation based generative model\( Park et al. (2018) \)

\[ S_{p,G} := \{p_g = g'p \circ g : g \in G\}. \]

(9)

Then for any two signals in \( S_{p,G} \), a unique constant-speed geodesic exists with respect to the generalized Wasserstein-2 metric defined in equation.
Proof. Let \( p_1 = g_1 \circ g_1 \in S_{p,G}, i = 1, 2. \) It follows that \( p_2 = (g_1^{-1} \circ g_2) p_1 \circ (g_1^{-1} \circ g_2). \) Since \( g_1^{-1} \circ g_2 \) is also strictly increasing, by Theorem 2.1 there exists a constant speed geodesic between \( p_1 \) and \( p_2. \)

Remark 2.4. More generally, one can show the existence of geodesics between pairs of signals in any subset \( U \) of \( S \) such that \( U \) is convex, the proof of which is presented in ??.

Theorem 2.5. The metric space \( (S, D_S(\cdot, \cdot)) \) is not a geodesic space.

Proof. Recall that a metric space is geodesic means that for any two given elements, there exists a path between such that the length of the path equals the distance between them. We prove this theorem by giving an example where the length of any path between some \( s_1, s_2 \in S \) is larger than their distance \( D_S(s_1, s_2) \). Let \( s_1 = \mathbb{1}_{[-1,0)} \mathbb{1}_{[0,1]} \) and \( s_2 = \mathbb{1}_{[0,1]} \mathbb{1}_{[-1,0]} \). Assume by contradiction that there is a path \( \gamma : [0, 1] \rightarrow S \) such that \( \gamma(0) = s_1 \) and \( \gamma(1) = s_2 \) such that \( D_S(s_1, s_2) = L \gamma(1) \). By the embedding property equation \( \text{7} \) of the CDT, it follows that \( \tilde{\gamma}_s(\alpha \in [0, 1]) \) defines a geodesic in \( \tilde{S} \subseteq (L^2(s_0) \times \mathbb{R})^2 \). Since the space \( (L^2(s_0) \times \mathbb{R})^2 \) is uniquely geodesic (see A.6), then \( \tilde{\gamma}_s = (1 - \alpha) \delta_1 + \alpha s_2. \) Let \( \tilde{\gamma}_s = (f_{1/2}^+, 1, f_{1/2}^+, 1) \in \tilde{S} \) and it is not hard to see that \( s_2 = (f_{1/2}^+, 1, f_{1/2}^+, 1) \).

Hence \( \tilde{\gamma}_1/2 := (f_{1/2}^+, a_1,f_{1/2}^+, b) = (f_{1/2}^+, 1, f_{1/2}^+, 1) \), which is a contradiction to the fact that \( \tilde{\gamma}_1/2 \in \tilde{S} \) (since by definition \( (f_{1/2}^+, a_1,f_{1/2}^+, b) \) should be mutually singular, see A.5).

Remark 2.6. Using the fact that the space \( (L^2(s_0) \times \mathbb{R})^2 \) is uniquely geodesic, one can show that if the geodesic in \( (S, D_S(\cdot, \cdot)) \) exists between two signals \( s \) and \( \bar{s} \), it should be of the form equation \( \text{8}. \)

Remark 2.7. Observing from Proposition A.1, it is not hard to see that \( \tilde{S} \) is not convex. This is in contrast to the fact the embedding CDT space \( S^*_1 \) is convex [Aldroubi et al., 2021].

Remark 2.8. The set \( G \) in Corollary 2.3 is a group with the group operation being composition and defines an equivalence relation in \( S \) via \( s \sim \bar{s} \) if and only if \( s = g/s \circ g \) for some \( g \in G \). It follows that \( S \) can be partitioned into disjoint clusters \( \bigcup_{i=1}^n \tilde{S}_{p_i,G} \) (see equation \( \text{9} \)). Hence by Remark 2.4 \( \tilde{S} \) can be partitioned as a union of convex subsets \( \bigcup_{i=1}^n \tilde{S}_{p_i,G}. \)

2.3 Examples of Geodesics

In this section, we give two examples (Figure 2) where a unique geodesic exists between the given signals and two examples (Figure 3) where a geodesic does not exist. By Remark 2.6, the geodesic \( p_\alpha \) for \( \alpha \in [0, 1] \) is of the form equation \( \text{8} \) when it exists. In the case when a geodesic does not exist, we plot a path in \( S \) according to equation \( \text{8} \). For a geodesic \( p_\alpha \) with \( p_0 = s \) and \( p_1 = \bar{s} \), it follows by definition that \( \sum_{i=1}^n D_S(p_{\alpha_{i-1}}, p_{\alpha_i}) = D_S(s, \bar{s}) \) for \( 0 = \alpha_0 < \alpha_1 < \cdots < \alpha_n = 1; \) while \( \sum_{i=1}^n D_S(p_{\alpha_{i-1}}, p_{\alpha_i}) > D_S(s, \bar{s}) \) if \( \alpha \mapsto p_{\alpha}, \alpha \in [0, 1] \) is not a geodesic path. In particular, in the following plots the signals \( p_{\alpha_i} \) where \( \alpha_i = \frac{i}{4}, i = 0, \ldots, 4 \) are presented and distances \( D_i := D_S(p_{\alpha_{i-1}}, p_{\alpha_i}), i = 1, 2, 3, 4 \) and \( D := D_S(s, \bar{s}) \) are shown.

3 A Preliminary Application: Interpreting the SCDT Subspace Classifier

In this section, we attempt to utilize the preliminary geodesic properties discussed above to partially interpret the decisions made by the SCDT subspace classifiers proposed in Rubaiyat et al. (2022b) and Rubaiyat et al. (2022a), which were shown to achieve very high accuracy in classifying segmented time series events.

\text{8}\) Here \( \mathbb{1}_{[a,b]} \) denotes the indicator function of interval \( [a, b]. \)

\text{9}\) By a direct computation (see A.8), we have that the geodesic is given by \( p_\alpha(t) = s(\frac{t}{\alpha}) \) for \( \alpha \in (0, 1] \) and \( p_0 := 0. \)
where a more general multi-template generative model was used, assuming that a signal class is generated from a set of templates under some unknown deformations. Formally, the multi-template generative model for signal class \( c \) is defined to be the set: 

\[
S^{(c)} = \bigcup_{m=1}^{M} S_{\varphi^{(c)}, G_m}^{(c)}, \quad \text{where } S_{\varphi^{(c)}, G_m}^{(c)} = \left\{ s_{j,m}^{(c)} \big| s_{j,m}^{(c)} = g_j \varphi^{(c)} \circ [g_j, g_j \in G_m] \right\} \text{ and } (G_m)_{\varphi^{(c)}}^{-1} = \left\{ \sum_{i=1}^{k} \alpha_i f_{i,m}^{(c)} \big| \alpha_i \geq 0 \right\}
\]

lies in a linear space of deformations. Each signal class is hypothesized to be modeled well by a union of subspaces \( \bigcup_{m=1}^{M} \tilde{V}_m^{(c)} \) in the embedding (transform) space, where 

\[
\tilde{V}_m^{(c)} = \text{span}(S_{\varphi^{(c)}, G_m}^{(c)}).
\]

The corresponding classifier searches for the nearest local subspace \( \tilde{V}_{m}^{(c)} \) to the test sample to predict the class label.

3.2 INTERPRETATION OF THE SCDT SUBSPACE CLASSIFIERS

As explained in the section above, the SCDT subspace classifiers search for the nearest subspace \( \tilde{V}^{(c)} \) to a given test signal in the embedding space. In particular, for the nearest subspace classifier, it can be shown that under certain assumptions, \( d^2(\hat{s}, \tilde{V}^{(c)}) < d^2(\hat{s}, \tilde{V}^{(p)}) \) for \( s \in S^{(c)} \) and \( p \neq c \), where \( d^2(\cdot, \cdot) \) denotes the \( L^2 \)-metric. Similar properties hold for the nearest local subspace classifier. The class labels are then predicted by searching for the (local) subspace with the smallest distance to the test signal.
In this paper, we provide a preliminary interpretation of the subspace classifiers using a synthetic dataset and a real dataset StarLightCurves\cite{Rebbapragada2009} hosted by the UCR time series classification archive\cite{Dau2018}. In particular, we present a visualization of the paths between a test signal and its projections onto the subspaces associated with different classes. The projection $\tilde{s}$ of a test signal $s$ onto the subspace $V^{(c)}$ associated with class $c$ is defined in the sense that $\tilde{s} := P_{V^{(c)}} s$, where $P_{V^{(c)}}$ denotes the $L^2$-projection operator to the subspace $V^{(c)}$.

If the classifier can successfully model a subspace (using training data) that contains a test signal $s$, then the projection $\tilde{s} = s$. In practice, we observe that the path defined by equation 8 between a test signal $s$ and its projection $\tilde{s}$ to a subspace associated with the same class as $s$ consists of signals of similar shapes (or looks like a geodesic lying in the same class) and $s \approx \tilde{s}$. On the other hand, the path between $s$ and its projection to a subspace associated with a different class seems to exhibit unpredictable behavior.

\textbf{Experiment 1:} For this experiment, we created a synthetic dataset of three classes consisting of scaled and translated signals of three template signals: Gabor wave, apodized sawtooth wave, and apodized square wave, respectively (shown in Figure 4a). The training samples were generated following the generative model given by: $S^{(c)} = \{ g \varphi^{(c)}(t) g \mid g \in G \}$, where $G = \{ g = \omega t + \tau, \omega > 0, \tau \in \mathbb{R} \}$ and $\varphi^{(c)}(t)$ denotes the template corresponding to class $c$ (see Figure 4a). Here the test sample (not present in training set) follows the generative model for class 1 (Gabor wave). The signal paths $p_\alpha$ between the test signal and its projections onto three subspaces formed via the subspace classifier\cite{Rubaiyat2022} are shown in Figure 5. We observe that the projection onto class 1 resembles the test signal and the corresponding path looks like a geodesic (i.e., the ratio $\sum_{i=1}^{4} D_i / D \approx 1$), indicating the subspace $V^{(1)}$ generated by the subspace classifier contains the SCDTs of the test signal $s$ and its cluster $S_s,G^{(9)}$. However the paths from the test signal to its projections onto the other two classes contain signals that belong to neither classes and seem to exhibit unpredictable patterns.

\textbf{Experiment 2:} We conducted a similar experiment with a real dataset StarLightCurves\cite{Rebbapragada2009}, where two classes: ‘Cepheid’ (class 1) and ‘Eclipsing Binary’ (class 2) were used. Sample signals from these two classes are shown in Figure 4b. Signal paths $p_\alpha$ are presented in Figure 9. Since the same phenomenon is observed for any random test signal following the generative model $S_{\varphi^{(1)},G}$, which equals $S_s,G$ for an arbitrary signal $s$ in class 1 since $G$ is the set of affine deformations.
from a test signal $s_j$ in each class ($j = 1, 2$) to their projections $\tilde{s}_j^{(i)}$ corresponding to subspaces $V^{(i)}$ ($i = 1, 2$) produced by the local subspace classifier proposed in Rubaiyat et al. (2022a). We observe that the path from $s_j$ to its projection $\tilde{s}_j^{(j)}$ associated with the same class resembles a geodesic within class $j$ (see the first and the fourth row in Figure 6) while the path from $s_j$ to its projection $\tilde{s}_j^{(i)}$ associated with a different class $i \neq j$ contains very noisy signals which seem to belong to neither classes (see the second and third row of Figure 6).

Figure 6: Signal path $p_\alpha$ defined by eq. (8) between a test sample $s_j$ and its projections $\tilde{s}_j^{(i)}$ ($i, j = 1, 2$) in the local subspaces obtained by the subspace classifier Rubaiyat et al. (2022a) corresponding to two classes, respectively. Note that in the top panel the distance $D$ between $s_1$ and $\tilde{s}_1^{(1)}$ (i.e., $D_S(s_1, \tilde{s}_1^{(1)})$) is smaller, and in the bottom panel the distance $D$ between $s_2$ and $\tilde{s}_2^{(2)}$ is smaller. Hence the local classifier predicts the correct labels in both cases.

Remark 3.1. Given the stark differences between the path from a signal to its projection associated with a subspace of its own class and that to its projection associated with a subspace of a different class, one may consider using the length of the path (see e.g., $\sum_i D_i$ in Figure 5) as the classification metric, in contrast to the $D_S$ distance between the signal and its projection used in the aforementioned subspace classifiers, which we leave to future work.

4 CONCLUSION

In this preliminary study, we look into the geometry of time series with respect to a generalized Wasserstein metric and its embedding in a linear space. In particular, a geodesic may not exist between two arbitrary time series in $S$ but exists between signals within the same generative cluster (see equation 9). We utilize this knowledge to show whether the (local) subspaces formed by the training samples via the SCDT subspace classifiers do a good job in modeling the signal classes conforming to certain generative modeling assumptions. We illustrate it by visualizing the paths between random test signals and their projections onto the corresponding subspaces associated with different classes and observe that there is a path resembling a geodesic from a test signal to its projection in the subspace associated with the same class of the test signal. This shows some preliminary evidence that the classifiers produce a “good” subspace that models the corresponding generative clusters to which the test signals belong. In the future, we are interested in quantitative (numerical) characterization of geodesics for real/noisy data in the hope of building more robust models and classifiers for time series events.

5 ACKNOWLEDGMENTS

This work was supported in part by NIH grants GM130825, GM090033.
REFERENCES


