Shapes analysis for time series.

Anonymous Author(s)
Affiliation
Address
email

Abstract

Analyzing inter-individual variability of physiological functions is particularly appealing in medical and biological contexts to describe or quantify health conditions. Such analysis can be done by comparing individuals to a reference one with time series as biomedical data. This paper introduces an unsupervised representation learning (URL) algorithm for time series tailored to inter-individual studies. The idea is to represent time series as deformations of a reference time series. The deformations are diffeomorphisms parameterized and learned by our method called TS-LDDMM. Once the deformations and the reference time series are learned, the vector representations of individual time series are given by the parametrization of their corresponding deformation. At the crossroads between URL for time series and shape analysis, the proposed algorithm handles irregularly sampled multivariate time series of variable lengths and provides shape-based representations of temporal data. In this work, we establish a representation theorem for the graph of a time series and derive its consequences on the LDDMM framework. We showcase the advantages of our representation compared to existing methods using synthetic data and real-world examples motivated by biomedical applications.

1 Introduction

Our goal is to analyze the inter-individual variability of a time series dataset, which is of prime interest in medicine and biology [21, 46, 3, 17]. More specifically, we aim to find an unsupervised features representation method which encode the specificity of an individual compared to another. In physiology, studying the different *shapes* in a time series related to biological phenomena and their variations according to individual or pathology is common. However, a *shape* has no clear definition; it is more an intuitive way to speak about the silhouette of a pattern in a time series. In this paper, we refer to as the shape of a time series, the graph of this signal.

Although a community structure with representatives can be learned in an unsupervised way [44, 31] using contrastive loss [16, 43, 31] or similarity measures [1, 17, 37, 49], studying the inter-individual variability of shapes within a cluster [35, 41] is still an open problem in URL.

First, we propose not to see time series through their curve $\{s_t :\ t \in \mathsf{I}\}$, but through their graph $\mathsf{G}(s) = \{(t, s(t)) :\ t \in \mathsf{I}\}$. Then, building on the shape analysis litterature [4, 45], we follow the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework [4, 45] to analyze these graphs. The idea is to represent each element $(\mathsf{G}(s^j))_{j\in[N]}$ of the dataset as the transformation of a reference graph $\mathsf{G}(\mathrm{s}_0)$ by a diffeomorphism. Then, the diffeomorphism is learned by integrating an ordinary differential equation parameterized by a Reproducing Kernel Hilbert Space (RKHS). The parameters $(\alpha_j)_{j\in[N]}$ encoding the diffemorphisms $(\phi_j)_{j\in[N]}$ yield the representation features of the graphs $(\mathsf{G}(s^j))_{j\in[N]}$. Finally, these features encoding the shapes can feed any statistical or machine learning model as in URL.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

However, a graph time series transformation by a general diffeomorphism is not always a graph time series, see e.g. Figure 1, thus a graph time series is more than a simple curve [19]. Our contributions arise from this observation. We solve this issue by specifying the class of diffeomorphisms to consider and showing how to learn them. This change is fruitful in representing time transformation as illustrated in Figure 2.

Our contributions can be summarized as follows:

- We propose an unsupervised method (TS-LDDMM) to analyze inter-individual variability of shapes in a time series dataset. In particular, the method can handle *irregularly sampled* time series with *variable sizes*.
- We motivate our extension of LDDMM to time series by introducing a theoretical framework with a representation theorem for time series graph (Theorem 1) and kernels related to their structure (Lemma 1).
- We demonstrate the identifiability of the model by estimating the true generating parameter of synthetic data, and we highlight the sensitivity of our method with respect to its hyperparameters, also providing guidelines for tuning. We highlight the *interpretability* of TS-LDDMM for studying the inter-individual variability in a clinical dataset. We illustrate the quantitative interest of the representation on classification tasks on real shape-based datasets.

Our goal is to analyze the inter-individual variability within a time series dataset, an approach of prime interest in physiological contexts [21, 46, 3, 17]. More specifically, we aim to find an unsupervised features representation method that encodes individual time series specificities compared to a reference one.

Studying shape differences between time series related to biological mechanisms is a common practice in physiology to characterize healthy and pathological functioning CITE. For instance, the shapes of heartbeats in electrocardiograms are discriminant for some cardiovascular pathologies CITE. Several approaches have been proposed for such comparison. Some employ shape-based similarity measures between time series [1, 17, 37, 49], others embed time series as vectors of predefined features CITE, and, with the rise of deep neural networks, unsupervised learning representation of time series [44, 31] has shown to be a valuable approach CITE notably with contrastive learning [16, 43, 31]. However, shape-based representation of time series within cohorts [35, 41] remains an open problem in URL.

Our work focuses explicitly on learning shape-based representation of time series. First, we propose not to see the shape of a time series through its curve $\{s_t : \ t \in \mathsf{I}\}$, but rather through its graph $\mathsf{G}(s) = \{(t, s(t)) : \ t \in \mathsf{I}\}$. Then, building on the shape analysis literature [4, 45], we follow the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework [4, 45] to analyze these graphs. The idea is to represent each element $\mathsf{G}(s^j)$ of a dataset $(s^j)_{j\in[N]}$ as the transformation of a reference graph $\mathsf{G}(\mathsf{s}_0)$ by a diffeomorphism ϕ_j (ie $\mathsf{G}(s^j) \sim \phi_j.\mathsf{G}(\mathsf{s}_0)$). The diffeomorphism ϕ_j is learned by integrating an ordinary differential equation parameterized by a Reproducing Kernel Hilbert Space (RKHS). The parameters $(\alpha_j)_{j\in[N]}$ encoding the diffemorphisms $(\phi_j)_{j\in[N]}$ yield the representation features of the graphs $(\mathsf{G}(s^j))_{j\in[N]}$. Finally, these shape-encoding features can feed any statistical or machine-learning model.

However, a graph time series transformation by a general diffeomorphism is not always a graph time series, see e.g. Figure 1, thus a graph time series is more than a simple curve [19]. Our contributions arise from this observation: we specify the class of diffeomorphisms to consider and show how to learn them. This change is fruitful in representing transformations of time series graphs as illustrated in Figure 2.

Our contributions can be summarized as follows:

- We propose an unsupervised method (TS-LDDMM) to analyze the inter-individual variability of shapes in a time series dataset. In particular, the method can handle multivariate time series *irregularly sampled* and with *variable sizes*.
- We motivate our extension of LDDMM to time series by introducing a theoretical framework with a representation theorem for time series graph (Theorem 1) and kernels related to their structure (Lemma 1).

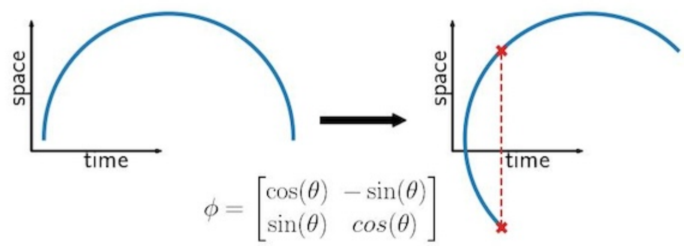

Figure 1: A time series' graph $\mathsf{G} = \{(t, s(t)) : \ t \in \mathsf{I}\}$ can lose its structure after applying a general diffeomorphism $\phi.\mathsf{G}$: a time value can be related to two values on the space axis.

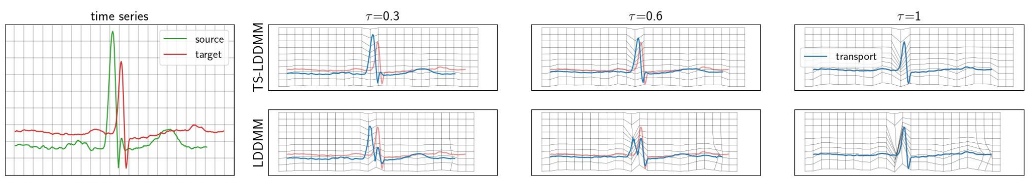

Figure 2: LDDMM and TS-LDDMM are applied to ECG data. We observe that LDDMM, using a general Gaussian kernel, does not learn the time translation of the first spike but changes the space values, i.e., one spike disappears before emerging at a translated position. At the same time, TS-LDDMM handles the time change in the shape. This difference of *deformations* implies differences in features *representations*.

- We demonstrate the identifiability of the model by estimating the true generating parameter of synthetic data, and we highlight the sensitivity of our method concerning its hyperparameters, also providing guidelines for tuning.
- We highlight the *interpretability* of TS-LDDMM for studying the inter-individual variability in a clinical dataset.
- We illustrate the quantitative interest of such representation on classification tasks on real shape-based datasets.

2 Related Works

Shape analysis focus on the statistical analysis of various mathematical objects invariant under rotations, dilations, or time parameterization. The main idea is to represent these different objects in a complete Riemannian manifold $(\mathcal{M}, \mathbf{g})$ with a metric $\mathbf{g}$ adapted to the geometry of the problem [32]. Then, any set of points in $\mathcal{M}$ can be represented as points in the tangent space of their Frechet mean $\mathbf{m}_0$ [36, 28] by considering their logarithms. The goal is then to find a well suited Riemannian struvture according to the studied object.

A time series graph can be seen as a curve and LDDMM structure is relevant to tackle curves as presented in [19]. However, time series graph has more structure than curves as depicted in Figure 1 due to the temporal evolution. [38] tracks anatomical shape changes in serial images using LDDMM, but distinguish from us by including the temporal evolution at a higher level: the goal is to perform longitudinal data modeling.

Leaving the LDDMM representation, [42, 22] address the representation of curves having unitary velocity by using the Square-Root Velocity (SRV) representation. However, the SRV representation applies after a reparametrization in time, such that the original time evolution of the time series is not represented in the final features. Again, the time series graph structure is not respected. Very recently in a functionnal data analysis framework, a paper [47] (Shape-FPCA) improved by giving a representation for the original time evolution. Nevertheless, this methods applies only on time series of *same size* and is made for *continuous objects*, making the estimation more sensitive to noise. [Ajouter littérature de URL demander à Sisi ?]

3 Notations

We denote by integer ranges by $[k : l] = \{k, \ldots, l\} \subset \mathcal{P}(\mathbb{Z})$ and $[l] = [1 : l]$ with $k, l \in \mathbb{N}$, by $\mathrm{C}^m(\mathsf{I}, \mathsf{E})$ the set of m-times continously differentiable function defined on an open set U to a normed vector space E, by $||u||_\infty = \inf_{x \in \mathsf{U}} |u(x)|$ for any bounded function $u : \mathsf{U} \to \mathsf{E}$, and by $\mathbb{N}_{>0}$ is the set of positive integers.

4 Background on LDDMM

In this part, there is no novelty, we simply expose how to learn the diffeomorphisms $(\phi_j)_{j \in [N]}$ using LDDMM, initially introduced in [4]. In a nutshell, for any $j \in [N]$, ϕ_j corresponds to a differential flow related to a learnable velocity field belonging to a well-chosen Reproducing Kernel Hilbert Space (RKHS).

In the next section, the time series are going to be represented by diffeomorphism parameters $(\alpha_j)_{j \in [N]}$. That's why LDDMM is chosen since it offers a parametrization for diffemorphisms which is sparse and interpretable, two features particularly relevant in the biomedical context.

The basic problem that we consider in this section is the following. Given a set of targets $\mathbf{y} = (y_i)_{i \in [T_2]}$ in $\mathbb{R}^{d'}$[1], a set of starting points $\mathbf{x} = (x_i)_{i \in [T_1]}$ in $\mathbb{R}^{d'}$, we aim to find a diffeomorphism ϕ such that the finite set of points $\mathbf{y}$ is similar in a certain sense to the set of finite sets of transformed points $\phi \cdot \mathbf{x} = (\phi(x_i))_{i \in [T_1]}$. The function ϕ is occasionally referred to as a *deformation*. In general, these sets $\mathbf{x}, \mathbf{y}$ are meshes of continous objects, e.g surfaces, curves, images and so on.

Representing diffeomorpshims as deformations. Such *deformations* ϕ are constructed via differential flow equations, for any $x_0 \in \mathbb{R}^{d'}$ and $\tau \in [0, 1]$:

$$\frac{\mathrm{d}X(\tau)}{\mathrm{d}\tau} = v_\tau(X(\tau)), \quad X(0) = x_0 \;, \phi^v_\tau(x_0) = X(\tau), \quad \phi^v = \phi^v_1 \;, \tag{1}$$

where the velocity field is $v : \tau \in [0, 1] \mapsto v_\tau \in \mathsf{V}$ and V is a Hilbert space of continuously differentiable function on $\mathbb{R}^{d'}$. If $||\,\mathrm{d}u||_\infty + ||u||_\infty \le ||u||_{\mathsf{V}}$ for any $u \in \mathsf{V}$ and $v \in \mathrm{L}^2([0, 1], \mathsf{V}) = \{v \in \mathrm{C}^0([0, 1], \mathsf{V}) : \int_0^1 ||v_\tau||^2_{\mathsf{V}} \,\mathrm{d}\tau < \infty\}$, by [18, Theorem 5] ϕ^v exists and belongs to $\mathcal{D}(\mathbb{R}^{d'})$, where we denote by $\mathcal{D}(\mathsf{O})$ the set of diffeomorpshim defined on an open set O to O. Therefore, for any choice of v, ϕ^v defines a valid deformation. This offers a general recipe to construct diffeomorphism given a functional space V.

With this in mind, the velocity field v fitting the data can be estimated by minimizing $v \in \mathrm{L}^2([0, 1], \mathsf{V}) \mapsto \mathscr{L}(\phi^v.\mathbf{x}, \mathbf{y})$, where $\mathscr{L}$ is an appropriate loss function. However, two computational challenges arise. First, this optimization problem is ill-posed, and a penalty term is needed to obtain a unique solution. In addition, we have to find a parametric family $\mathsf{V}_\Theta \subset \mathrm{L}^2([0, 1], \mathsf{V})$, parameterized by Θ, which allows us to solve this minimization problem efficiently.

It has been proposed in [32] to interpret V as a tangent space relative to the group of diffeomorphisms $\mathsf{H} = \{\phi^v : \; v \in \mathrm{L}^2([0, 1], \mathsf{V})\}$. Following this geometric point of view, geodesics can be constructed on H by using the following squared norm

$$\mathscr{R}^2 : g \in \mathsf{H} \mapsto \inf_{v \in \mathrm{L}^2([0,1],\mathsf{V}) :\; g = \phi^v} \int_0^1 ||v_\tau||_{\mathsf{V}} \,\mathrm{d}\tau \tag{2}$$

By deriving differential constraints related to the minimum of (2) and using Cauchy lipshcitz conditions, geodesics can be defined only by giving the starting point and the initial velocity $v_0 \in \mathsf{V}$ [32], as straight lines in Euclidean space. Denoting by $w(v_0)$ the geodesic starting from the identity with inital velocity v_0, the exponential map is defined as $\varphi^{\{v_0\}} \triangleq \phi^v$ and the previous matching problem becomes a *geodesic shooting problem*:

$$\inf_{v_0 \in \mathsf{V}} \mathscr{L}(\varphi^{\{v_0\}}.\mathbf{x}, \mathbf{y}). \tag{3}$$

Using $\varphi^{\{v_0\}}$ instead of ϕ^v for any $v \in \mathrm{L}^2([0, 1], \mathsf{V})$ regularizes the problem and induces a sparse representation for the learning diffeomorphisms. Moreover, by setting V as an RKHS, the geodesic shooting problem has a unique solution and becomes tractable, as described in the next section.

[1] Note that we denote by $d' \in \mathbb{N}$ the ambient space

Discrete parametrization of diffeomorpshim. In this part, V is chosen as an RKHS [5] generated by a smooth kernel K (e.g., Gaussian). We follow [13] and define a discrete parameterization of the velocity fields to perform geodesics shooting (3). The initial velocity field v_0 is chosen as a finite linear combination of the RKHS basis vector fields, $\mathbf{n}_0$ control points $\mathsf{X}_0 = (x_{k,0})_{k\in[\mathbf{n}_0]} \in (\mathbb{R}^{d'})^{\mathbf{n}_0}$ and momentum vectors $\alpha_0 = (\alpha_{k,0})_{k\in[\mathbf{n}_0]} \in (\mathbb{R}^{d'})^{\mathbf{n}_0}$ are defined such that for any $x \in \mathbb{R}^{d'}$,

$$v_0\left(\alpha_0, \mathsf{X}_0\right)(x) = \sum_{k=1}^{\mathbf{n}_0} K(x, x_{k,0})\alpha_{k,0} \ . \tag{4}$$

In our applications, the control points $(x_{k,0})_{k\in[\mathbf{n}_0]}$ can be understood as the discretized graph $(t_k, \mathbf{s}_0(t_k))_{k\in[\mathbf{n}_0]}$ of a starting time series $\mathbf{s}_0$. With this parametrization of v_0, **(author?)** [32] show that the velocity field v of the solution of (3) keeps the same structure along time, such that for any $x \in \mathbb{R}^{d'}$ and $\tau \in [0, 1]$,

$$v_\tau(x) = \sum_{k=1}^{\mathbf{n}_0} K(x, x_k(\tau))\alpha_k(\tau) \ ,$$

$$\begin{cases} \dfrac{\mathrm{d}x_k(\tau)}{\mathrm{d}\tau} = v_\tau(x_k(\tau)) \ , \quad \dfrac{\mathrm{d}\alpha_k(\tau)}{\mathrm{d}\tau} = -\displaystyle\sum_{k=1}^{\mathbf{n}_0} \mathrm{d}_{x_k(\tau)}K(x_k(\tau), x_l(\tau))\alpha_l(\tau)^\top\alpha_k(\tau) \\ \alpha_k(0) = \alpha_{k,0}, \quad x_k(0) = x_{k,0} \ , k \in [\mathbf{n}_0] \end{cases} \tag{5}$$

These equations are derived from the hamiltonian $H : (\alpha_k, x_k)_{k\in[\mathbf{n}_0]} \mapsto \sum_{k,l=1}^{\mathbf{n}_0} \alpha_k^\top K(x_k, x_l)\alpha_l$, such that the velocity norm is preserved $||v_\tau||_{\mathsf{V}} = ||v_0||_{\mathsf{V}}$ for any $\tau \in [0, 1]$. By (5), the velocity field related to a geodesic v^* is fully parametrized by its initial control points and momentum $(x_{k,0}, \alpha_{k,0})_{k\in[\mathbf{n}_0]}$. Thus, given a set of targets $\mathbf{y} = (y_i)_{i\in[T_2]}$ in $\mathbb{R}^{d'}$, a set of starting points $\mathbf{x} = (x_{i,0})_{i\in[T_1]}$ in $\mathbb{R}^{d'}$, a RKHS's kernel $K : \mathbb{R}^{d'} \times \mathbb{R}^{d'} \to \mathbb{R}^{d'\times d'}$, a distance on sets $\mathscr{L}$, a numerical integration scheme of ODE and a penalty factor $\lambda > 0$, the basic geodesic shooting step minimizes the following function using a gradient descent method:

$$\mathcal{F}_{\mathbf{x},\mathbf{y}} : (\alpha_k)_{k\in[T_1]} \mapsto \mathscr{L}\left(\varphi^{\{v_0\}}.\mathbf{x}, \mathbf{y}\right) + \lambda||v_0||_{\mathsf{V}}^2 \ , \tag{6}$$

where v_0 is defined by (4) and $\varphi^{\{v_0\}}.\mathbf{x}$ is the result of the numerical integration of (5) using control points $\mathbf{x}$ and initial momentums $(\alpha_k)_{k\in[T_1]}$.

Relation to Continuous Normalizing Flows. One particular popular choice to address the problem of learning a diffeomorphism or a velocity field is Normalizing Flows [39, 27] (NF) or their continuous counterpart [9, 20, 40] (CNF). However, we do not rely on this class of learning algorithms for several reasons. Indeed, existing and simple normalizing flows are not suitable for the type of data that we are interested in this paper [15, 12]. In addition, they are primarily designed to have tractable Jacobian functions, while we do not require such property in our applications. Finally, the use of a differential flow solution of an ODE (1) trick is also at the basis of CNF, which then consists of learning a velocity field to address in fitting the data through a loss aiming to address the problem at hand. Nevertheless, the main difference between CNF and LDDMM lies in the parametrization of the velocity field. LDDMM uses kernels to derive closed form formula and enhance interpretability while NF and CNF take advantage of deep neural networks to scale with large dataset in high dimensions.

5 Methodology

We consider in this paper observations which consist in a population of N multivariate time series, for any $j \in [N]$, $s^j \in \mathrm{C}^1(\mathsf{I}_j, \mathbb{R}^d)$. However, we can only access a n_j-samples $\tilde{s}^j = (\tilde{s}_i^j = s^j(t_i^j))_{i\in[n_j]}$ collected at timestamps $(t_i^j)_{i\in[n_j]}$ for any $j \in [N]$. Note that **the number of samples n_j is not necessary the same accross individuals** and the timestamps can be **irregularly sampled**. We assume the time series population is globally homogeneous regarding their "shapes" even if inter-individual variability exists. Intuitively speaking, the "shape" of a time series $s : \mathsf{I} \to \mathbb{R}^d$ is encoded in its graphs $\mathsf{G}(s)$ defined as the set $\{(t, s(t)) : \ t \in \mathsf{I}\}$ and not only in its values $s(\mathsf{I}) = \{s(t) : \ t \in \mathsf{I}\}$ since the time axis is crucial. As a motivating use-case, s^j can be the time series of a heartbeat extracted from an individual's electrocardiogram (ECG), see Figure 2. The homogeneity in a resulting dataset comes from the fact that humans have similar shapes of heartbeat [48, 30].

The deformation problem. In this paper, we aim to study the inter-individual variability in the dataset by finding a relevant representation of each time series. Inspired from the framework of shape analysis [45], addressing similar problems in morphology, we suggest to represent each time series' graph $\mathsf{G}(s^j)$ as the transformation of a reference graph $\mathsf{G}(\mathbf{s}_0)$, related to a time series $\mathbf{s}_0 : \mathsf{I} \to \mathbb{R}^d$, by a diffeomorphism ϕ_j on $\mathbb{R}^{d+1}$, for any $j \in [N]$,

$$\phi_j.\mathsf{G}(\mathbf{s}_0) = \{\phi_j\left(t, \mathbf{s}_0(t)\right),\ t \in \mathsf{I}\}\ . \tag{7}$$

$\mathbf{s}_0$ will be understood as the typical representative shape common to the collection of time series $(s^j)_{j\in[N]}$. As $\mathbf{s}_0$ is supposed to be fixed, then the representation of the time series $(s^j)_{j\in[N]}$ boils down to the one of the transformation $(\phi_j)_{j\in[N]}$. We aim to learn $\mathsf{G}(\mathbf{s}_0)$ and $(\phi_j)_{j\in[N]}$.

Optimization related to (7)**.** Defining the *discretized graphs* of the time series $(s^j)_{j\in[N]}$ and a discretization of the reference graph $\mathsf{G}(\mathbf{s}_0)$ as, for any $j \in [N]$,

$$\mathbf{y}_j = \mathsf{G}(\tilde{s}^j) = (t_i^j, \tilde{s}_i^j)_{i\in[n_j]} \in (\mathbb{R}^{d+1})^{n_j}, \quad \tilde{\mathsf{G}}_0 = (t_i^0, \tilde{s}_i^0)_{i\in[\mathbf{n}_0]} \in (\mathbb{R}^{d+1})^{\mathbf{n}_0}\ ,$$

with $\mathbf{n}_0 = \mathrm{median}((n_j)_{j\in[N]})$, the representation problem given in (7) boils down solving:

$$\mathrm{argmin}_{\tilde{\mathsf{G}}_0, (\alpha_k^j)_{k\in[\mathbf{n}_0]}^{j\in[N]}} \sum_{j=1}^{N} \mathcal{F}_{\tilde{\mathsf{G}}_0, \mathbf{y}_j}\left((\alpha_k^j)_{k\in[\mathbf{n}_0]}\right)\ , \tag{8}$$

which is carried out by a gradient descent on the control points $\tilde{\mathsf{G}}_0$ and the momentums $\alpha_j = (\alpha_k^j)_{k\in[\mathbf{n}_0]}$ for any $j \in [N]$, initialized by a dataset's time series graph of size $\mathbf{n}_0$ and by $0_{(d+1)\mathbf{n}_0}$ respectively. The optimization hyperparameter details are given in Appendix D.1. The result of the minimization $\tilde{\mathsf{G}}_0$ is then considered as the $\mathbf{n}_0$-samples of a common time series $\mathbf{s}_0$ and the momentums α_j encoding ϕ_j yields a feature vector in $\mathbb{R}^{d\mathbf{n}_0}$ of s^j for any $j \in [N]$. Finally, the vectors $(\alpha_j)_{j\in[N]}$ can be analyzed with any statistical or machine learning tools such as Principal Components Analysis (PCA), Latent Discriminant Analysis (LDA), longitudinal data analysis and so on.

Nevertheless, (8) ask to define a kernel and a loss in order to perform geodesic shooting 6, which is the purpose of the next subsection.

5.1 Application of LDDMM to time series analysis: TS-LDDMM

In this section, we present our theoretical contribution: we tailor the LDDMM framework to handle time series data. The reason is that applying a general diffeomorphism ϕ from $\mathbb{R}^{d+1}$ to a time series' graph $\mathsf{G}(s)$ can result in a set $\phi.\mathsf{G}(s)$ that does not correspond to the graph of any time series, as illustrated in the Figure 1. Thus, Time series graph have more structure than a simple 1D curve [19] and deserve their special analysis which will prove fruitful as demonstrated in 6.

To address this challenge, we need to identify an RKHS kernel $K : \mathbb{R}^{d+1} \times \mathbb{R}^{d+1} \to \mathbb{R}^{(d+1)^2}$ that generates deformations preserving the structure of the time series graph. This goal motivates us to clarify, in Theorem 1, the specific representation of diffeomorphisms we require before presenting a class of kernels that produce deformations with this representation.

Similarly, selecting a loss function on sets $\mathscr{L}$ that considers the temporal evolution in a time series' graph is crucial for meaningful comparisons with time series data. Consequently, we introduce the oriented Varifold distance.

A representation separating space and time. We prove that two time series graphs can always be linked by a time transformation composed of a space transformation. Moreover, a time series graph transformed by this kind of transformation is always a time series graph. We define $\Psi_\gamma \in \mathcal{D}(\mathbb{R}^{d+1}) : (t,x) \in \mathbb{R}^{d+1} \to (\gamma(t), x)$ for any $\gamma \in \mathcal{D}(\mathbb{R})$ and $\Phi_f : (t,x) \in \mathbb{R}^{d+1} \to (t, f(t,x))$ for any $f \in \mathrm{C}^1(\mathbb{R}^{d+1}, \mathbb{R}^d)$. We have the following representation theorem. All proofs are given in Appendix A.

Denote by $\mathsf{G}(s) \triangleq \{(t, s(t)) :\ t \in \mathsf{I}\}$ the graph of a time series $s : \mathsf{I} \to \mathbb{R}^d$ and $\phi.\mathsf{G}(s) \triangleq \{\phi(t, s(t)) :\ t \in \mathsf{I}\}$ the action of $\phi \in \mathcal{D}(\mathbb{R}^{d+1})$ on $\mathsf{G}(s)$.

Theorem 1. *Let $s : \mathsf{J} \to \mathbb{R}^d$ and $\mathbf{s}_0 : \mathsf{I} \to \mathbb{R}^d$ be two continuously differentiable time seriess with I, J two intervals of $\mathbb{R}$. There exist $f \in \mathrm{C}^1(\mathbb{R}^{d+1}, \mathbb{R}^d)$ and $\gamma \in \mathcal{D}(\mathbb{R})$ such that $\gamma(\mathsf{I}) = \mathsf{J}$ and $\Phi_f \in \mathcal{D}(\mathbb{R}^{d+1})$,*

$$\mathsf{G}(s) = \Pi_{\gamma,f}.\mathsf{G}(\mathbf{s}_0),\ \Pi_{\gamma,f} = \Psi_\gamma \circ \Phi_f .$$

Moreover, for any $\bar{f} \in \mathrm{C}^1(\mathbb{R}^{d+1}, \mathbb{R}^d)$ and $\bar{\gamma} \in \mathcal{D}(\mathbb{R})$, there exists a continously differentiable time series $\bar{s}$ such that $\mathsf{G}(\bar{s}) = \Pi_{\bar{\gamma},\bar{f}}.\mathsf{G}(\mathbf{s}_0)$

Remark 2. *that for any $\gamma \in \mathcal{D}(\mathbb{R})$ and $s \in \mathrm{C}^0(\mathsf{I}, \mathbb{R}^d)$,*

$$\{(\gamma(t), s(t)),\ t \in \mathsf{I}\} = \{(t, s \circ \gamma^{-1}(t)) :\ t \in \gamma(\mathsf{I})\} .$$

As a result, Ψ_γ can be understood as a temporal reparametrization and Φ_f encodes the transformation about the space.

Choice for the kernel associated with the RKHS V As depicted on Figure 1-2, we can not use any kernel K to apply the previous methodology to learn deformations on time series' graphs. We describe and motivate our choice in this paragraph. Denote the one-dimensional Gaussian kernel by $K^{(a)}_\sigma(x, y) = \exp(-|x - y|^2/\sigma)$ for any $(x, y) \in (\mathbb{R}^a)^2$, $a \in \mathbb{N}$ and $\sigma > 0$. To solve the geodesic shooting problem (6) on $\mathbb{R}^{d+1}$, we consider for V the RKHS associated with the kernel defined for any $(t, x), (t', x') \in (\mathbb{R}^{d+1})^2$:

$$\begin{aligned} K_{\mathsf{G}}((t,x),(t',x')) &= \begin{pmatrix} c_0 K_{\text{time}} & 0 \\ 0 & c_1 K_{\text{space}} \end{pmatrix} , \\ K_{\text{space}} = K^{(1)}_{\sigma_{T,1}}(t,t') K^{(d)}_{\sigma_x}(x,x') \mathrm{I}_d &, K_{\text{time}} = K^{(1)}_{\sigma_{T,0}}(t,t') , \end{aligned} \tag{9}$$

parametrized by the widths $\sigma_{T,0}, \sigma_{T,1}, \sigma_x > 0$ and the constants $c_0, c_1 > 0$. This choice for K_{G} is motivated by the representation Theorem 1 and the following result.

Lemma 1. *If we denote by V the RKHS associated with the kernel K_{G}, then for any vector field v generated by* (5) *with v_0 satisfying* (4)*, there exist $\gamma \in \mathsf{D}(\mathbb{R})$ and $f \in \mathrm{C}^1(\mathbb{R}^{d+1}, \mathbb{R}^d)$ such that $\phi^v = \Psi_\gamma \circ \Phi_f$.*

Parler des Cauchy kernel en appndice et du choix de la loss

Remark 3. *With this choice of kernel, the features associated to the time transformation can be extracted from the momentums $(\alpha_{k,0})_{k\in[\mathbf{n}_0]} \in (\mathbb{R}^{d+1})^{\mathbf{n}_0}$ in* (4) *by taking the coordinates related to time. However, the features related to the space transformation are not only in the space coordinates since the related kernel K_{space} depends on time as well.*

In Appendix C, we give guidelines for selecting the hyperparameters $(\sigma_{T,0}, \sigma_{T,1}, \sigma_x, c_0, c_1)$.

Loss This section specifies the distance function $\mathscr{L}$ introduced in the loss function defined in (6).

In practice, we can only access discretized graphs of time series, $(t^j_i, \tilde{s}^j_i)_{i\in[n_j]}$ for any $j \in [N]$, that are potentially of different sizes n_j and sampled at different timestamps $(t^j_i)_{i\in[n_j]}$ for any $j \in [N]$. Usual metrics, such as the Euclidean distance, are not appealing as they make the underlying assumptions of equal size sets and the existence of a pairing between points. Distances between measures on sets (taking the empirical distribution), such as Maximum Mean Discaprency (MMD) [14, 6], alleviate those issues; however, MMD only accounts for positional information and lacks information about the time evolution between sampled points. A classical data fidelity metric from shape analysis corresponding to the distance between *oriented varifolds* associated with curves alleviates this last issue [26]. Intuitively, an oriented varifold is a measure that accounts for positional and tangential information about the underlying curves at sample points. More details and information about *oriented varifolds* can be found in Appendix B.

More precisely, given two sets $\mathsf{G}_0 = (g^0_i)_{i\in[T_0]}, \mathsf{G}_1 = (g^1_i)_{i\in[T_1]} \in (\mathbb{R}^{d+1})^{T_1}$ and a kernel[2] $k : (\mathbb{R}^{d+1} \times \mathbb{S}^d)^2 \to \mathbb{R}$ verifying [26, Proposition 2 & 4], for any $\xi \in \{0, 1\}$ and $i \in [T_\xi - 1]$, denoting the center and length of the i^{th} segment $[g^\xi_i, g^\xi_{i+1}]$ by $c^\xi_i = (g^\xi_i + g^\xi_{i+1})/2$, $l^\xi_i = \|g^\xi_{i+1} - g^\xi_i\|$, and

[2] $\mathbb{S}^d = \{x \in \mathbb{R}^{d+1} :\ |x| = 1\}$

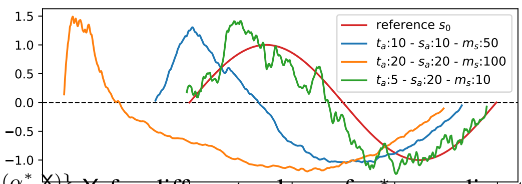

Figure 3: Plots of $\varphi^{\{v_0(\alpha^*, \mathsf{X})\}}.\mathsf{X}$ for different values of α^* according to its sampling parameter t_a, s_a, m_s, taking $\mathsf{X} = \mathsf{G}(s_0)$ with $s_0 : k \in [300] \to \sin(2\pi k/300)$.

Table 1: Values of $\mathscr{L}(\varphi^{\{v_0(\alpha^*, \mathsf{X})\}}.\mathsf{X}, \varphi^{\{\hat{v}_0\}}.\mathsf{X})$ as α^* is sampled according to Gen(10,10,50) and $\hat{v}_0$ is estimated using K_G with varying parameters $\sigma_{T,1}, \sigma_x$.

$\sigma_{T,0} \backslash \sigma_x$	1	10	50	100	200	300
0.1	2e+0	3e-4	1e-5	4e-6	7e-4	4e-3
1	4e-2	1e-4	1e-5	4e-6	7e-4	4e-3
100	4e-2	2e-4	1e-5	4e-6	7e-4	4e-3

$\overrightarrow{v_i}^\xi = (g_{i+1}^\xi - g_i^\xi)/l_i^\xi$, the varifold distance between G_0 and G_1 is defined as,

$$d^2_{\mathsf{W}^*}(\mathsf{G}_0, \mathsf{G}_1) = \sum_{i,j=1}^{T_0-1} l_i^0 k((c_i^0, \overrightarrow{v_i}^0), (c_j^0, \overrightarrow{v_j}^0)) l_j^0 - 2 \sum_{i=1}^{T_0-1} \sum_{j=1}^{T_1-1} l_i^0 k((c_i^0, \overrightarrow{v_i}^0), (c_j^1, \overrightarrow{v_j}^1)) l_j^1$$
$$+ \sum_{i,j=1}^{T_1-1} l_i^1 k((c_i^1, \overrightarrow{v_i}^1), (c_j^1, \overrightarrow{v_j}^1)) l_j^1$$

In practice, we set the kernel k as the product of two anisotropic Gaussian kernels, k_{pos} and k_{dir}, such that for any $(x, \overrightarrow{u}), (y, \overrightarrow{v}) \in (\mathbb{R}^{d+1} \times \mathbb{S}^d)^2$

$$k((x, \overrightarrow{u}), (y, \overrightarrow{v})) = k_{\text{pos}}(x, y) k_{\text{dir}}(\overrightarrow{u}, \overrightarrow{v}) \ .$$

The specific kernels $k_{\text{pos}}, k_{\text{dir}}$ that we use in our experiments are given Appendix B.1. Note that the loss kernel k has nothing to do with the velocity field kernel denoted by K_G or K specified in Section 5.1. Finally, we define the data fidelity loss function, $\mathscr{L}$, as $d^2_{\mathsf{W}^*}$, which is differentiable with regards to its first variable. For further readings on curves and surfaces representation as varifolds, readers can refer to [26, 8].

Parler de méthode adaptatif ici

6 Experiments

[L'intro est ce necessaire ?]

First, we show on synthetic data that the proposed representation is identifiable provided that the hyperparameters and the reference graph are wisely selected, i.e., the parameter v_0^* generating a deformation $\varphi^{\{v_0^*\}}$ of a time series graph G can be estimated from the data $\mathsf{G}, \varphi^{\{v_0^*\}}.\mathsf{G}$ by solving the geodesic shooting problem (6). Secondly, we illustrate the qualitative interest of TS-LDDMM in studying inter-individual variability on a clinical dataset. Thirdly, we demonstrate the quantitative performance of our representation by performing classification on shape-based datasets. The method is implemented on Python using the library JAX[3]. The code was compiled on a server with NVIDIA RTX A2000 12GB GPU, Intel(R) Xeon(R) Gold 5220R CPU @ 2.20GHz, and 250 GB of RAM. The code will be available on Github.

6.1 Synthetic experiments

First, we show the model identifiability when the kernel K_G is well specified: the estimated parameter is a good approximation of the generating parameter when the generation and the estimation procedure use the same hyperparameters for the RKHS kernel K_G. All the hyperparameter values for generation and estimation are given in Appendix D.2. We fix the initial control points

[3]https://github.com/google/jax

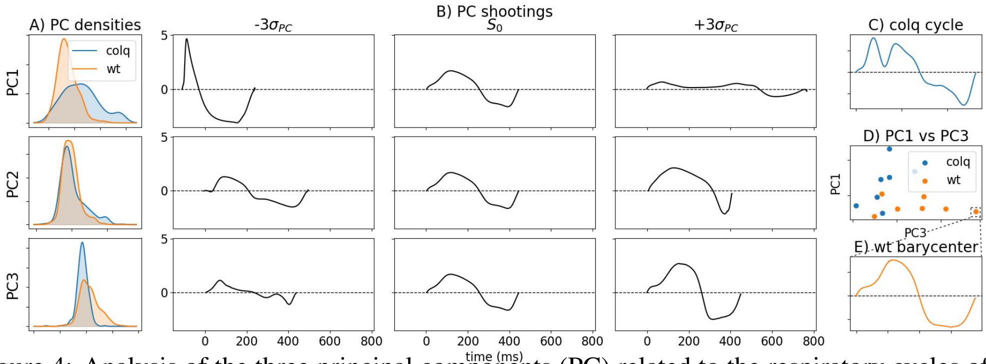

Figure 4: Analysis of the three principal components (PC) related to the respiratory cycles of the mouse before exposure. In Figure A), the densities of each genotype according to each PC are displayed. In Figure B), the deformations of the reference graph S_0 along each PC are given. In Figure D), the graph of reference S^j, also called barycenter, related to each mouse, is displayed according to their coordinates on PC1 and PC3. In Figure C) et E), illustrations of respiratory cycles related to mice coming from the **wt** and **colq** group are displayed.

as $\mathsf{X} = (x_k = (k, \sin(2\pi k/300)))_{k\in[300]}$. Given $m_s \in \mathbb{N}_{>0}$ and $t_a, s_a > 0$, we randomly generate initial momentums $\alpha^* = (\alpha_k^*)_{k\in[\mathbf{n}_0]}$ with the following sampling, called Gen(m_s, t_a, s_a): For any $k \in [\mathbf{n}_0]$, α'_k is sampled according to a Gaussian normal distribution $\mathcal{N}(0_{d+1}, I_{d+1})$. Then, $(\alpha'_k)_{k\in[\mathbf{n}_0]}$ is regularized by a rolling average of size m_s, we get $\bar{\alpha}' = (\bar{\alpha}'_k)_{k\in[\mathbf{n}_0]}$. Finally, we normalize $\bar{\alpha}'$ to derive α^* such that $|([\alpha_k^*]_t)_{k\in[\mathbf{n}_0]}| = t_{\text{amp}}$ and $|([\alpha_k^*]_s)_{k\in[\mathbf{n}_0]}| = s_{\text{amp}}$ for any $k \in [\mathbf{n}_0]$, denoting by $[\alpha_k^*]_t, [\alpha_k^*]_s$ the time and space coordinates of α_k^* respectively. Note that the regularizing step $(\alpha'_k)_{k\in[\mathbf{n}_0]} \to \bar{\alpha}'$ is necessary to obtain realistic deformations which take into account the regularity induced by the RKHS V. Then, using $v_0(\alpha^*, \mathsf{X})$ as defined in (4) with initial momentums α^* and control points X, we apply the induced deformation $\varphi^{\{v_0\}}$ by (5) to X and obtain $\varphi^{\{v_0\}}.\mathsf{X}$. Finally, we solve (6) to recover an estimation $\hat{\alpha}$ of α^* and report the average relative error (ARE) $|v_0(\hat{\alpha}, \mathsf{X}) - v_0(\alpha^*, \mathsf{X})|_{\mathsf{V}}/|v_0(\alpha^*, \mathsf{X})|_{\mathsf{V}}$ on 50 repetitions. This procedure is performed for any $m_s, t_a, s_a \in \{10, 50, 100\} \times \{5, 10, 15, 20\}^2$. Mean, standard deviation, and maximum of the ARE on all these hyperparameters choices are respectively $\mathbf{0.10}, \mathbf{0.03}, \mathbf{0.17}$. Therefore, the estimation procedure (6) offers a good approximation of the true parameter when the kernel K_{G} is well specified. We observe that the estimation is difficult when $t_a \ll s_a$ because the time series can be very noisy as illustrated in Figure 3: this impacts the Varifold loss which is sensitive to tangents.

Secondly, we demonstrate a weak identifiability when the kernel K_{G} is misspecified: we can reconstruct the graph time series' after deformations even if the hyperparameters of K_{G} are different during the generation and the estimation. The hyperparameters of K_{G} during generation are $(c_0, c_1, \sigma_{T,0}, \sigma_{T,1}, \sigma_x) = (1, 0.1, 100, 1, 1)$ and we fix $\sigma_{T,1}, c_0, c_1 = (1, 1, 0.1)$ for K_{G} during estimation. We aim to understand the impact of $\sigma_{T,1}, \sigma_x$ on the reconstruction since they are encoding the smoothness of the transformation according to time and space.

For any choice of the hyperparameters $\sigma_{T,1}, \sigma_x \in \{1, 10, 50, 100, 200, 300\} \times \{0.1, 1, 100\}$ related to K_{G} in the estimation, we average $\mathscr{L}(\varphi^{\{v_0(\alpha^*,\mathsf{X})\}}.\mathsf{X}, \varphi^{\{\hat{v}_0\}}.\mathsf{X})$ on 50 repetitions when α^* is sampled according to Gen$(10, 10, 50)$ and $\hat{v}_0 = v_0(\hat{\alpha}, \mathsf{X})$ denoting by $\hat{\alpha}$ the result of the minimization (6). We observe in Table 1 that the reconstruction is almost perfect except in the case when $\sigma_{t,0} = 1$ during estimation, while $\sigma_{t,0} = 100$ during generation. Compared to $\sigma_{T,0}$, σ_x has nearly no impact on the reconstruction. In Appendix B.1-C, we propose guidelines to drive future hyperparameters tuning and further discussions related to $\sigma_{T,1}, c_0, c_1$.

6.2 Qualitative analysis of respiratory behavior in mice

This experiment highlights the *interpretability* of TS-LDDMM for studying the inter-individual variability in a clinical dataset. We consider a time series dataset recording the evolution of the respiratory airflow of mice exposed to an irritant molecule altering respiratory functions [34]. The dataset is divided into two groups, one composed of 7 control mice (**wt**) and the other of 7 mice (**colq**) deficient in an enzyme involved in the control of respiration. For each mouse, the respiratory

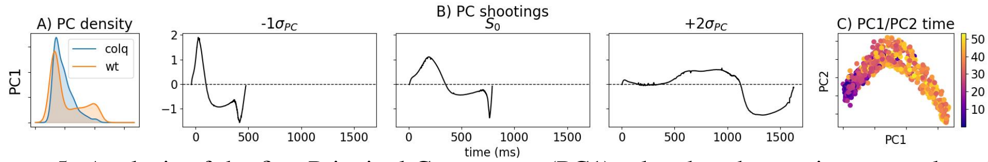

Figure 5: Analysis of the first Principal Component (PC1) related to the respiratory cycles of the mouse before and after exposure. In Figure A), the densities of each genotype according are displayed. In Figure B), the deformations of the reference graph S_0 PC1 is given. In Figure C), respiratory cycles displayed with respect to time and according to their coordinates on PC1 and PC2

airflow was recorded for 15 to 20 minutes before exposure to the irritant molecule and then for 35 to 40 minutes. A complete description of the dataset is given in the Appendix E. By comparing the shape of individual respiratory cycles (inspiration + expiration, see Figure 4-C)), we show that TS-LDDMM features can encode genotype distinctive breathing behaviors and their evolution after exposure to the irritant molecule.

We first compare breathing behaviors before exposure. Solving (8), we derive the reference respiratory cycle's graph S_0 and the TS-LDDMM features representations $(\alpha_j)_{j\in[N_1]}$ related to $N_1 = 700$ respiratory cycles extracted according to the procedure [17]. Then, we perform a kernel PCA on the initial velocity field $(v_0(\alpha_j, S_0))_{j\in[N_1]} \in \mathsf{V}^{N_1}$ defined in (4). In Figure 4, we focus on the analysis of the three Principal Components (PC).

As observable from Figure 4-B), principal components refer to different types of deformations. By interpreting Figure 4-B): Only PC1 accounts for time warping, PC2 expresses the trade-off between inspiration and expiration duration, and PC3 corresponds to a change in signal amplitude. Compared to **wt** mice, the distribution of **colq** mice TS-LDMMM feature representation along the PC1 axis has a heavy tail and the associated deformation (+3 σ_{PC}) shows an inspiration with two peaks. As illustrated in Figure 4-A), such respiratory cycles are preponderant with **colq** mice and may be caused by motor impairment due to their enzyme deficiency, [17]. In addition, the **colq** mice were smaller than the **wt** mice due to a delay in growth caused by their lack of an enzyme. This difference can be seen on PC3 since the volumes of air (area under the curve) inspired and exhaled are smaller for the smaller mice. In correlation, the distribution of **wt** mice TS-LDDMM feature representations along the PC3 axis have a heavy tail corresponding to large air volume as depicted by the deformation (+3 σ_{PC}) in Figure 4-B). Finally, Figure 4-D) shows that PC1 and PC3 capture the main differences between the two groups as their respective reference graphs S^j are located in different parts of the space.

We perform a second experiment to analyze the evolution of breathing behaviors when mice are exposed to the irritant molecule. We follow the same procedure as before. However, we take $N_2 = 1400$ with 25% (resp. 75%) before (resp. after) exposure. In Figure 5, we focus on the first principal component PC since it encodes the effect of the irritant molecule as depicted in Figure 5-C) (the exposure occurs at 20 minutes). Figure 5-B) shows that the deformation (+3 σ_{PC}) leads to longer respiratory cycles that include pauses, as observed in [17]. As well, Figure 5-A) shows that TS-LDDMM features distributions are less spread out for **colq** mice compared to **wt** mice. Indeed, the irritant molecule inhibits the action of the deficient enzyme, **wt** mice strongly react to the irritant molecule, whereas **colq** mice are better adapted due to their deficiency.

6.3 Quantitative performances of the TS-LDDMM representation in classification

Combined with a Support Vector Classifier (SVC) [24], TS-LDDMM representation can be used for classification tasks using the kernel associated with the initial velocity space V. We compare TS-LDDMM-SVC classification performances with another SVC using representation learned with T-loss [16], an unsupervised deep learning feature representation method for time series. We also include fully supervised methods in deep learning -ResNet, CNN [25]- and machine learning: Catch22 [29], Rocket [11], Dynamic Time Wrapping k-Nearest Neighbors (DTW-kNN) [33]. Methods are compared using f1-score on several shape-based UCR/UEA datasets [10, 2] introduced in Appendix F. All implementation details are given in Appendix D.4. Table 2 presents the reuslts. TS-LDDMM-SVC consistently outperforms the other unsupervised methods. It is ranked 1,3,4,3 for all methods combined, demonstrating its competitiveness as an unsupervised method on time series dataset homogeneous regarding shape.

Table 2: Classification results in f1-score (U: unsupervised, S: supervised, DL: deep learning, ML: machine learning). **x** best unsupervised method, <u>x</u> best supervised method.

			ArrowHead	ECG200	GunPoint	NATOPS
U		TS-LDDMM-SVC	**0.84**	**0.82**	**0.94**	**0.93**
		T-loss-SVC	0.57	0.76	0.82	0.88
		DTW-kNN	0.70	0.75	0.91	0.88
S	DL	CNN	0.70	0.79	0.85	<u>0.96</u>
		ResNet	0.77	0.87	0.97	0.95
	ML	Catch22	0.73	0.81	0.96	0.89
		Rocket	<u>0.81</u>	<u>0.91</u>	<u>1.00</u>	0.88

7 Conclusion

In this paper, we propose a feature representation method, TS-LDDMM, designed for shape comparison in homogeneous time series datasets. We show on a real dataset its ability to study, with high interpretability, the inter-individual shape variability. As an unsupervised approach, it is user-friendly and enables knowledge transfer for different supervised tasks such as classification. Although TS-LDDMM is already competitive for classification, its performances can be leveraged on more heterogeneous datasets using a hierarchical clustering extension, which is relagated for future work.

References

[1] Asal Asgari. Clustering of clinical multivariate time-series utilizing recent advances in machine-learning. 2023.

[2] Anthony Bagnall, Hoang Anh Dau, Jason Lines, Michael Flynn, James Large, Aaron Bostrom, Paul Southam, and Eamonn Keogh. The uea multivariate time series classification archive, 2018. *arXiv preprint arXiv:1811.00075*, 2018.

[3] Ziv Bar-Joseph, Anthony Gitter, and Itamar Simon. Studying and modelling dynamic biological processes using time-series gene expression data. *Nature Reviews Genetics*, 13(8):552–564, 2012.

[4] M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61:139–157, 2005.

[5] Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.

[6] Karsten M Borgwardt, Arthur Gretton, Malte J Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006.

[7] Claudio Carmeli, Ernesto De Vito, Alessandro Toigo, and Veronica Umanitá. Vector valued reproducing kernel hilbert spaces and universality. *Analysis and Applications*, 8(01):19–61, 2010.

[8] Nicolas Charon and Alain Trouvé. The varifold representation of nonoriented shapes for diffeomorphic registration. *SIAM journal on Imaging Sciences*, 6(4):2547–2580, 2013.

[9] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[10] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. The ucr time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, 2019.

[11] Angus Dempster, François Petitjean, and Geoffrey I Webb. Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5):1454–1495, 2020.

[12] Ruizhi Deng, Bo Chang, Marcus A Brubaker, Greg Mori, and Andreas Lehrmann. Modeling continuous stochastic processes with dynamic normalizing flows. *Advances in Neural Information Processing Systems*, 33:7805–7815, 2020.

[13] Stanley Durrleman, Stéphanie Allassonnière, and Sarang Joshi. Sparse adaptive parameterization of variability in image ensembles. *International Journal of Computer Vision*, 101:161–183, 2013.

[14] Gintare Karolina Dziugaite, Daniel M Roy, and Zoubin Ghahramani. Training generative neural networks via maximum mean discrepancy optimization. *arXiv preprint arXiv:1505.03906*, 2015.

[15] Shibo Feng, Chunyan Miao, Ke Xu, Jiaxiang Wu, Pengcheng Wu, Yang Zhang, and Peilin Zhao. Multi-scale attention flow for probabilistic time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[16] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. Unsupervised scalable representation learning for multivariate time series. *Advances in neural information processing systems*, 32, 2019.

[17] Thibaut Germain, Charles Truong, Laurent Oudre, and Eric Krejci. Unsupervised classification of plethysmography signals with advanced visual representations. *Frontiers in Physiology*, 14:781, 2023.

[18] Joan Glaunes. Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique. *These de sciences, Université Paris*, 13, 2005.

[19] Joan Glaunes, Anqi Qiu, Michael I Miller, and Laurent Younes. Large deformation diffeomorphic metric curve mapping. *International journal of computer vision*, 80:317–336, 2008.

[20] Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, and David Duvenaud. Scalable reversible generative models with free-form continuous dynamics. In *International Conference on Learning Representations*, page 7, 2019.

[21] Ella Guscelli, John I Spicer, and Piero Calosi. The importance of inter-individual variation in predicting species' responses to global change drivers. *Ecology and Evolution*, 9(8):4327–4339, 2019.

[22] Tae-Young Heo, Joon Myoung Lee, Myung Hun Woo, Hyeongseok Lee, and Min Ho Cho. Logistic regression models for elastic shape of curves based on tangent representations. *Journal of the Korean Statistical Society*, pages 1–19, 2024.

[23] Heinz Gerd Hoymann. Lung function measurements in rodents in safety pharmacology studies. *Frontiers in pharmacology*, 3:156, 2012.

[24] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification, 2003.

[25] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4):917–963, 2019.

[26] Irene Kaltenmark, Benjamin Charlier, and Nicolas Charon. A general framework for curve and surface comparison and registration with oriented varifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3346–3355, 2017.

[27] Ivan Kobyzev, Simon JD Prince, and Marcus A Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):3964–3979, 2020.

[28] Huiling Le. Locating fréchet means with application to shape spaces. *Advances in Applied Probability*, 33(2):324–338, 2001.

[29] Carl H Lubba, Sarab S Sethi, Philip Knaute, Simon R Schultz, Ben D Fulcher, and Nick S Jones. catch22: Canonical time-series characteristics: Selected through highly comparative time-series analysis. *Data Mining and Knowledge Discovery*, 33(6):1821–1852, 2019.

[30] Putri Madona, Rahmat Ilias Basti, and Muhammad Mahrus Zain. Pqrst wave detection on ecg signals. *Gaceta Sanitaria*, 35:S364–S369, 2021.

[31] Qianwen Meng, Hangwei Qian, Yong Liu, Yonghui Xu, Zhiqi Shen, and Lizhen Cui. Unsupervised representation learning for time series: A review. *arXiv preprint arXiv:2308.01578*, 2023.

[32] Michael I Miller, Alain Trouvé, and Laurent Younes. Geodesic shooting for computational anatomy. *Journal of mathematical imaging and vision*, 24:209–228, 2006.

[33] Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[34] Aurélie Nervo, André-Guilhem Calas, Florian Nachon, and Eric Krejci. Respiratory failure triggered by cholinesterase inhibitors may involve activation of a reflex sensory pathway by acetylcholine spillover. *Toxicology*, 424:152232, 2019.

[35] Vit Niennattrakul and Chotirat Ann Ratanamahatana. Inaccuracies of shape averaging method using dynamic time warping for time series data. In *Computational Science–ICCS 2007: 7th International Conference, Beijing, China, May 27-30, 2007, Proceedings, Part I 7*, pages 513–520. Springer, 2007.

[36] Susovan Pal, Roger P Woods, Suchit Panjiyar, Elizabeth Sowell, Katherine L Narr, and Shantanu H Joshi. A riemannian framework for linear and quadratic discriminant analysis on the tangent space of shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 47–55, 2017.

[37] John Paparrizos and Luis Gravano. k-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pages 1855–1870, 2015.

[38] Anqi Qiu, Marilyn Albert, Laurent Younes, and Michael I Miller. Time sequence diffeomorphic metric mapping and parallel transport track time-dependent shape changes. *NeuroImage*, 45(1):S51–S60, 2009.

[39] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.

[40] Hadi Salman, Payman Yadollahpour, Tom Fletcher, and Kayhan Batmanghelich. Deep diffeomorphic normalizing flows. *arXiv preprint arXiv:1810.03256*, 2018.

[41] Gota Shirato, Natalia Andrienko, and Gennady Andrienko. Identifying, exploring, and interpreting time series shapes in multivariate time intervals. *Visual Informatics*, 7(1):77–91, 2023.

[42] Anuj Srivastava, Eric Klassen, Shantanu H Joshi, and Ian H Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE transactions on pattern analysis and machine intelligence*, 33(7):1415–1428, 2010.

[43] Sana Tonekaboni, Danny Eytan, and Anna Goldenberg. Unsupervised representation learning for time series with temporal neighborhood coding. *arXiv preprint arXiv:2106.00750*, 2021.

[44] Patara Trirat, Yooju Shin, Junhyeok Kang, Youngeun Nam, Jihye Na, Minyoung Bae, Joeun Kim, Byunghyun Kim, and Jae-Gil Lee. Universal time-series representation learning: A survey. *arXiv preprint arXiv:2401.03717*, 2024.

[45] Marc Vaillant, Michael I Miller, Laurent Younes, and Alain Trouvé. Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 23:S161–S169, 2004.

[46] Kai Wang, Youjin Zhao, Qingyu Xiong, Min Fan, Guotan Sun, Longkun Ma, Tong Liu, et al. Research on healthy anomaly detection model based on deep learning from multiple time-series physiological signals. *Scientific Programming*, 2016, 2016.

[47] Yuexuan Wu, Chao Huang, and Anuj Srivastava. Shape-based functional data analysis. *TEST*, 33(1):1–47, 2024.

[48] Can Ye, BVK Vijaya Kumar, and Miguel Tavares Coimbra. Heartbeat classification using morphological and dynamic features of ecg signals. *IEEE Transactions on Biomedical Engineering*, 59(10):2930–2941, 2012.

[49] Lexiang Ye and Eamonn Keogh. Time series shapelets: a new primitive for data mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 947–956, 2009.

[50] Juntang Zhuang, Tommy Tang, Yifan Ding, Sekhar C Tatikonda, Nicha Dvornek, Xenophon Papademetris, and James Duncan. Adabelief optimizer: Adapting stepsizes by the belief in observed gradients. *Advances in neural information processing systems*, 33:18795–18806, 2020.

A Proofs

Denote by $\mathsf{G}(s) \triangleq \{(t, s(t)) : t \in \mathsf{I}\}$ the graph of a time series $s : \mathsf{I} \to \mathbb{R}^d$ and $\phi.\mathsf{G}(s) \triangleq \{\phi(t, s(t)) : t \in \mathsf{I}\}$ the action of $\phi \in \mathcal{D}(\mathbb{R}^{d+1})$ on $\mathsf{G}(s)$.

Theorem 4. *Let $s : \mathsf{J} \to \mathbb{R}^d$ and $\mathbf{s}_0 : \mathsf{I} \to \mathbb{R}^d$ be two continuously differentiable time seriess with I, J two intervals of $\mathbb{R}$. There exist $f \in \mathrm{C}^1(\mathbb{R}^{d+1}, \mathbb{R}^d)$ and $\gamma \in \mathcal{D}(\mathbb{R})$ such that $\gamma(\mathsf{I}) = \mathsf{J}$ and $\Phi_f \in \mathcal{D}(\mathbb{R}^{d+1})$,*

$$\mathsf{G}(s) = \Pi_{\gamma,f}.\mathsf{G}(\mathbf{s}_0),\ \Pi_{\gamma,f} = \Psi_\gamma \circ \Phi_f .$$

Moreover, for any $\bar{f} \in \mathrm{C}^1(\mathbb{R}^{d+1}, \mathbb{R}^d)$ and $\bar{\gamma} \in \mathcal{D}(\mathbb{R})$, there exists a continously differentiable time series $\bar{s}$ such that $\mathsf{G}(\bar{s}) = \Pi_{\bar{\gamma},\bar{f}}.\mathsf{G}(\mathbf{s}_0)$

Proof. Let $s : \mathsf{J} \to \mathbb{R}^d$ and $\mathbf{s}_0 : \mathsf{I} \to \mathbb{R}^d$ be two continuously differentiable time seriess with $\mathsf{I} = (a, b), \mathsf{J} = (\alpha, \beta)$ two intervals of $\mathbb{R}$. By setting $\gamma : t \in \mathbb{R} \mapsto (\beta - \alpha)(t - a)/(b - a) + \alpha \in \mathbb{R}$, we have $\gamma(\mathsf{I}) = \mathsf{J}$ and $\gamma \in \mathcal{D}(\mathbb{R})$. By defining $f : (t, x) \in \mathbb{R}^{d+1} \mapsto x - \mathbf{s}_0(t) + s \circ \gamma(t)$, the map $\Phi_f \in \mathcal{D}(\mathbb{R}^{d+1})$, indeed, its inverse is $\Phi_f^{-1} : (t, x) \in \mathbb{R}^{d+1} \mapsto (t, x + \mathbf{s}_0(t) - s(t))$ and is continuously differentiable. Moreover, we have $\Pi_{\gamma,f}.\mathsf{G}(\mathbf{s}_0) = \{(\gamma(t), s \circ \gamma(t)) : t \in \mathsf{I}\} = \mathsf{G}(s)$.

Let $\bar{f} \in \mathrm{C}^0(\mathbb{R}^{d+1}, \mathbb{R}^d)$, $\bar{\gamma} \in \mathcal{D}(\mathbb{R})$ and $\mathbf{s}_0 \in \mathrm{C}^0(\mathsf{I}, \mathbb{R}^d)$ with I an interval of $\mathbb{R}$. We have :

$$\begin{aligned} \Pi_{\gamma,f}.\mathsf{G}(\mathbf{s}_0) &= \{(\gamma(t), f(t, \mathbf{s}_0(t))),\ t \in \mathsf{I}\,\} \\ &= \left\{\left(t, f\left(\gamma^{-1}(t), \mathbf{s}_0(\gamma^{-1}(t))\right)\right),\ t \in \gamma(\mathsf{I})\ \right\}. \end{aligned} \tag{10}$$

By defining $\bar{s} : t \in \gamma(\mathsf{I}) \to f\left(\gamma^{-1}(t), \mathbf{s}_0(\gamma^{-1}(t))\right)$, we have $\bar{s} \in \mathrm{C}^0(\gamma(\mathsf{I}), \mathbb{R}^d)$ by composition of continuous functions and $\mathsf{G}(\bar{s}) = \Pi_{\gamma,f}.\mathsf{G}(\mathbf{s}_0)$ by (10), which concludes the proof. □

Lemma 2. *If we denote by V the RKHS associated with the kernel K_G, then for any vector field v generated by* (5) *with v_0 satisfying* (4)*, there exist $\gamma \in \mathsf{D}(\mathbb{R})$ and $f \in \mathrm{C}^1(\mathbb{R}^{d+1}, \mathbb{R}^d)$ such that $\phi^v = \Psi_\gamma \circ \Phi_f$.*

Proof. Let v be a vector field generated by (5) with v_0 satisfying (4). We remark that the first coordinate of the velocity field v_τ denoted by v_τ^{time} only depends on the time variable t for any $\tau \in [0, 1]$. Thus, when computing the first coordinate of the deformation ϕ^v, denoted by γ, we integrate (1) with v_τ replaced by v_τ^{time}, thus γ is independant of the variable x. Moreover, $\gamma \in \mathcal{D}(\mathbb{R})$ since a Gaussian kernel induced an Hilbert space V satisfying $|f|_V \leq |f|_\infty + |\,\mathrm{d}f|_\infty$ for any $f \in \mathsf{V}$ by [18, Theorem 9]. For the same reason, we have $\phi^v \in \mathcal{D}(\mathbb{R}^{d+1})$, and thus its last coordinates denoted by f belongs to $\mathrm{C}^1(\mathbb{R}^{d+1}, \mathbb{R}^d)$, and by construction $\phi^v = \Psi_\gamma \circ \Phi_f$. □

B Oriented varifold

In this section, we introduce the *oriented varifold* associated with curves. For further readings on curves and surfaces representation as varifolds, readers can refer to [26, 8]. We associate to $\gamma \in \mathrm{C}^1((a,b), \mathbb{R}^{d+1})$ an *oriented varifold* μ_γ, i.e. a distribution on the space $\mathbb{R}^{d+1} \times \mathbb{S}^d$ defined as follows, for any smooth test function $\omega : \mathbb{R}^{d+1} \times \mathbb{S}^d \to \mathbb{R}$,

$$\mathbb{E}_{Y \sim \mu_\gamma}\left[\omega(Y)\right] = \mu_\gamma(\omega) = \int_a^b \omega\left(\gamma(t), \frac{\dot\gamma(t)}{|\dot\gamma(t)|}\right) |\dot\gamma(t)| \, \mathrm{d}t \, .$$

Denoting by W the space of smooth test function, we have that μ_γ belongs to its dual W^*. Thus, a distance on W^* is sufficient to set a distance on oriented varifolds associated to curve and thus on $\mathrm{C}^1((a,b), \mathbb{R}^{d+1})$ by the identification $\gamma \to \mu_\gamma$. Remark that in (TS-LDDMM), γ should be the parametrization of a time series' graph $\mathsf{G}(s)$, i.e. $\gamma : t \in \mathsf{I} \to (t, s(t)) \in \mathbb{R}^{d+1}$ denoting by $s : \mathsf{I} \to \mathbb{R}^d$ the time series. However, in practice, we work with discrete objects. That is why, we set W as an RKHS to use its representation theorem. More specifically [26, Proposition 2 & 4] encourages us to consider a kernel $k : (\mathbb{R}^{d+1} \times \mathbb{S}^d)^2 \to \mathbb{R}$ such that there exist two positive and continuously differentiable kernels k_{pos} and k_{dir}, such that for any $(x, \overrightarrow{u}), (y, \overrightarrow{v}) \in (\mathbb{R}^{d+1} \times \mathbb{S}^d)^2$

$$k((x, \overrightarrow{u}), (y, \overrightarrow{v})) = k_{\mathrm{pos}}(x, y) k_{\mathrm{dir}}(\overrightarrow{u}, \overrightarrow{v}) \, ,$$

with moreover $k_{\mathrm{dir}} > 0$ and k_{pos} which admits an RKHS $\mathsf{W}_{\mathrm{pos}}$ dense in the space of continous function on $\mathbb{R}^{d+1}$ vanishing at infinite [7].

Given such a kernel $k : (\mathbb{R}^{d+1} \times \mathbb{S}^d)^2 \to \mathbb{R}$ verifying [26, Proposition 2 & 4], we have that for any $(x, v) \in \mathbb{R}^{d+1} \times \mathbb{S}^d$, $\delta_{(x, \overrightarrow{v})}$ belongs to W^* as a distribution and that the dual metric $\langle \cdot, \cdot \rangle_{\mathsf{W}^*}$ satisfies for any $(x_1, v_1), (x_2, v_2) \in \left(\mathbb{R}^{d+1} \times \mathbb{S}^d\right)^2$,

$$\langle \delta_{(x_1, \overrightarrow{v}_1)}, \delta_{(x_2, \overrightarrow{v}_2)} \rangle_{\mathsf{W}^*} = k((x_1, \overrightarrow{v}_1), (x_2, \overrightarrow{v}_2)) \, .$$

Thus, given two sets of triplets $X = (l_i, x_i, \overrightarrow{v}_i)_{i \in [T_0 - 1]} \in (\mathbb{R} \times \mathbb{R}^{d+1} \times \mathbb{S}^d)^{T_0 - 1}, Y = (l'_i, y_i, \overrightarrow{w}_i)_{i \in [T_1]} \in (\mathbb{R} \times \mathbb{R}^{d+1} \times \mathbb{S}^d)^{T_1 - 1}$ and denoting by

$$\mu_X = \sum_{i=1}^{T_0} l_i \delta_{(x_i, \overrightarrow{v}_i)}, \mu_Y = \sum_{i=1}^{T_1} l'_i \delta_{(y_i, \overrightarrow{w}_i)} \, , \tag{11}$$

we have,

$$|\mu_X - \mu_Y|^2_{\mathsf{W}^*} = \sum_{i,j=1}^{T_0-1} l_i k((x_i, \overrightarrow{v_i}), (x_i, \overrightarrow{v_i^0})) l_j - 2 \sum_{i=1}^{T_0-1} \sum_{j=1}^{T_1-1} l_i k((x_i, \overrightarrow{v_i}), (y_i, \overrightarrow{w_i})) l'_j + \sum_{i,j=1}^{T_1-1} l'_i k((y_i, \overrightarrow{w_i}), (y_i, \overrightarrow{w_i})) l'_j \, .$$

Then, using the identification $X \to \mu_X, Y \to \mu_Y$, we can define a distance on sets of triplets as $d_{\mathsf{W}^*,3}(X, Y) = |\mu_X - \mu_Y|^2_{\mathsf{W}^*}$.

Now, we aim to discretize the oriented varifold μ_{G} related to a time series' graph $\mathsf{G}(s)$ by using a set of triplets. This is carried out by using a discretized version of $\mathsf{G}(s)$, i.e. $\tilde{\mathsf{G}} = (g_i = (t_i, s(t_i)))_{i \in [T]} \in (\mathbb{R}^{d+1})^T$, in the following way: For any $i \in [T-1]$, denoting the center and length of the i^{th} segment $[g_i, g_{i+1}]$ by $c_i = (g_i + g_{i+1})/2$, $l_i = \|g_{i+1} - g_i\|$, and the unit norm vector of direction $\overrightarrow{g_i g_{i+1}}$ by $\overrightarrow{v_i} = (g_{i+1} - g_i)/l_i$, we define the set of triplets $X(\tilde{\mathsf{G}}) = (l_i, c_i, \overrightarrow{v_i})_{i \in [T-1]}$ and its related oriented varifold $\mu_{X(\tilde{\mathsf{G}})} = \sum_{i=1}^{T-1} l_i \delta_{c_i, \overrightarrow{v_i}}$ as in (11). This is a valid discretization of the oriented varifold μ_{G} according to [26, Proposition 1]: $\mu_{X(\tilde{\mathsf{G}})}$ converges towards μ_{G} as the size of the descretization mesh $\sup_{i \in [T-1]} |t_{i+1} - t_i|$ converges to 0.

Finally, we define a distance on discretized time series' graphs $\tilde{\mathsf{G}}_1, \tilde{\mathsf{G}}_2$ as $d_{\mathsf{W}^*}(\tilde{\mathsf{G}}_1, \tilde{\mathsf{G}}_2) = d_{\mathsf{W}^*,3}(X(\tilde{\mathsf{G}}_1), X(\tilde{\mathsf{G}}_2))$.

B.1 Varifold kernels

Denote the one-dimensional Gaussian kernel by $K_\sigma^{(a)}(x, y) = \exp(-|x - y|^2/\sigma)$ for any $(x, y) \in (\mathbb{R}^a)^2$, $a \in \mathbb{N}$ and $\sigma > 0$. In the implementation, we use the following kernels, for any $((t_1, x_1), (t_2, x_2)) \in (\mathbb{R}^{d+1})^2, ((w_1, v_1), (w_2, v_2)) \in (\mathbb{S}^d)^2$,

$$k_{\mathrm{pos}}(x, y) = K^{(1)}_{\sigma_{\mathrm{pos},t}}(t_1, t_2) K^{(d)}_{\sigma_{\mathrm{pos},x}}(x_1, x_2), \quad k_{\mathrm{pos}}(x, y) = K^{(1)}_{\sigma_{\mathrm{dir},t}}(w_1, w_2) K^{(d)}_{\sigma_{\mathrm{dir},x}}(v_1, v_2) \, ,$$

where $\sigma_{\mathrm{pos},t}, \sigma_{\mathrm{pos},x}, \sigma_{\mathrm{dir},t}, \sigma_{\mathrm{dir},x} > 0$ are hyperparameters. In practice, we select $\sigma_{\mathrm{pos},x} \approx \sigma_{\mathrm{dir},x} \approx 1$ when the times series are centered and normalized. Otherwise we select $\sigma_{\mathrm{pos},x} \approx \sigma_{\mathrm{dir},x} \approx \bar{\sigma}_s$ with $\bar{\sigma}_s$ the average standard deviation of the time series. We choose $\sigma_{\mathrm{pos},t} \approx \sigma_{\mathrm{dir},t} = mf_e$ with f_e the sampling frequency of the time series and $m \in [5]$ an integer depending on the time change between the starting and the target time series graph. The more significant the time change, the higher m should be. The intuition comes from the fact that the width $\sigma_{\mathrm{pos},t}, \sigma_{\mathrm{dir},t}$ rules the time windows used to perform the comparison, and $\sigma_{\mathrm{pos},x}, \sigma_{\mathrm{dir},x}$ affects the space window. The size of the windows should be selected depending on the variations in the data.

C Tuning the hyperparameters of the TS-LDDMM kernel given in (9)

The parameter $\sigma_{T,0}$ should be chosen *large* compared the sampling frequency f_e and compared to average standard deviation $\bar{\sigma}_s$ of the time series, e.g $\sigma_{T,0} = 100$ as $\bar{\sigma}_s \approx f_e \approx 1$. It makes the time transformation smoother. If $\sigma_{T,0}$ is too small, for instance, $\sigma_{T,0} = f_e$, the effect of the time deformation is too localized, and there are not enough samples to make it visible.

The parameter $\sigma_{T,1}$ should be of the same order as f_e: two different points in time can have various space transformations. σ_x should be of the same order of $\bar{\sigma}_s$: two points with a big difference regarding space compared to $\bar{\sigma}_s$ can have very different space transformations.

We take $c_0 \approx 10c_1$, we want to encourage time transformation before space transformation. We take $(c_0, c_1) = (1, 0.1)$ in all experiments.

D Numerical details

A report of all the hyperparameters selected is given in Table 3.

D.1 Optimization details of (8)

Initialization At the initialization of (8), all the momentums parameter are set to 0 and the graph of reference is set to the graph of a time series in the dataset having a mediane samples size.

Gradient descent. The chosen gradient descent method is "adabelief" [50] implemented in the library OPTAX[4]. There are two main parameters in the gradient descent: the number of steps nb_steps, and the maximum value of step size η_M. The stepsize has a particular scheduling:

- Warmup period on $0.1\times$ nb_steps steps: the stepsize increases linearly from 0 to η_M. The goal is to learn progressively the parameters. If the stepsize is too large at the start, smaller steps at the end can't make up for the mistakes made at the beginning.
- Fine tuning periode on $0.9\times$ nb_steps : the stepsize decreases from η_M to 0 with a cosine decay implemented in the OPTAX scheduler, i.e. the decreasing factor as the form $0.5(1 + \cos(\pi t/T))$.

The sharper the deformations, the larger the number of steps and the maximum value of step size should be selected. We suggest nb_steps=300, $\eta_M = 0.1$ for small deformations and nb_steps=800, $\eta_M = 0.3$ for big ones (time dilation with a factor $\lambda \geq 2$).

D.2 Synthetic experiments

For any deformations generation in both experiments (well-specified and misspecified), we take $\sigma_{T,0}, \sigma_{T,1}, \sigma_x = (100, 1, 1)$ and $c_0, c_1 = (1, 0.1)$ for the kernel K_{G} and $\sigma_{\mathrm{pos},t}, \sigma_{\mathrm{pos},t}, \sigma_{\mathrm{dir},t}, \sigma_{\mathrm{dir},x} = (2, 1, 2, 0.6)$ for the varifold kernels $k_{\mathrm{pos}}, k_{\mathrm{dir}}$ related to the loss $\mathscr{L}$.

In both experiments, we have nb_steps=300 abd $\eta_M = 0.1$.

[4]https://optax.readthedocs.io/en/latest/

D.3 Mouse experiments

The number of steps is larger in the second experiment (before/after injection) because the deformations are sharper.

D.4 Classification experiments

We defined a default parametrization for all classifiers.

For classifiers: CNN, ResNet, Catch22, DTW-KNN, Rocket we used the aeon[5] implementations with their default settings.

For Tloss-SVC we used the implementation provided on github[6] with the following parameters for learning representations: batch_size: 10, channels: 40, depth: 10, nb_steps: 200, in_channels: 1, kernel_size: 3, lr: 0.001, nb_random_samples: 10, negative_penalty: 1, out_channels: 320, reduced_size: 160. We used the Support Vector Classifier (SVC) from scikit-learn with thee regularization term C: 1. Others parameters are set to default.

For TS-LDDMMM-SVC, all kernels' parameters et optimizer parameter are presented in Table 3. As well, we used the Support Vector Classifier from scikit-learn with thee regularization term C: 1. Others parameters are set to default.

Table 3: Parameters used in all the experiments. For synthetic data, K_{G} refers to the kernel used in the generation, which is the same for the estimation only in the well-specified case. $\bar{l}$ refers to the average time series length and N_d refers to the number of dimensions.

objects	Optimizer	$k_{\mathrm{pos}}, k_{\mathrm{dir}}$	K_{G}
Parameter	(nb_steps,η_M)	$(\sigma_{\mathrm{pos},t}, \sigma_{\mathrm{pos},t}, \sigma_{\mathrm{dir},t}, \sigma_{\mathrm{dir},x})$	$(c_0, c_1, \sigma_{T,0}, \sigma_{T,1}, \sigma_x)$
Synthetic data well-specified	(300,0.1)	$(2, 1, 2, 0.6)$	$(1, 0.1, 100, 1, 1)$
Synthetic data misspecified	(300,0.1)	$(2, 1, 2, 0.6)$	$(1, 0.1, 100, 1, 1)$
Mouse before injection	(400,0.3)	$(2, 1, 2, 0.6)$	$(1, 0.1, 100, 1, 1)$
Mouse before/after injection	(800,0.3)	$(5, 1, 5, 0.6)$	$(1, 0.1, 150, 1, 1)$
classification	(400,0.1)	$(\max(2, 0.03\bar{l}), N_d, \max(2, 0.03\bar{l}), 0.6)$	$(1, 0.1, 0.33\bar{l}, 1, N_d)$

E Mouse respiratory dataset

Ventilation is a simple physiological function that ensures a vital supply of oxygen and the elimination of CO2. Acetylcholine (Ach) is a neurotransmitter that plays an important role in muscular activity, notably for breathing. Indeed, muscle contraction information passes from the brain to the muscle through the nervous system. Achs are located in synapses of the nervous system (central and peripheral) and skeletal muscles. They ensure the information transmission from nerve to nerve. However, the transmission cannot end without the hydrolysis of Ach by the enzyme Acetylcholinesterase (AchE), allowing nerves to return to their resting state. Inhibition of (AchE) with, for instance, nerve gas, pesticide, or drug intoxication leads to respiratory arrests.

The dataset comes from the experiment [34], where they studied the consequences of partial deficits in AChE and AChE inhibition on mice respiration. AchE inhibition was induced with an irritant molecule called physostigmine (an AchE inhibitor). Mice nasal airflows were sampled at 2000Hz with a Double Chamber plethysmograph [23], as depicted in Figure 6-A). The flow is expressed in $ml.s^{-1}$; it has a positive value during inspiration and a negative value expiration Figure 6-B). Among the mice population, we selected 7 control mice (**wt**) and 7 ColQ mice (**colq**), which do not have AChE anchoring in muscles and some tissues. As described in [34], mice experiments were as follows:

1. The mouse is placed in a DCP for 15 or 20 min to serve as an internal control.
2. The mouse is removed from the DCP and injected with physostigmine.

[5] https://www.aeon-toolkit.org/en/stable/index.html
[6] https://github.com/mqwfrog/ULTS

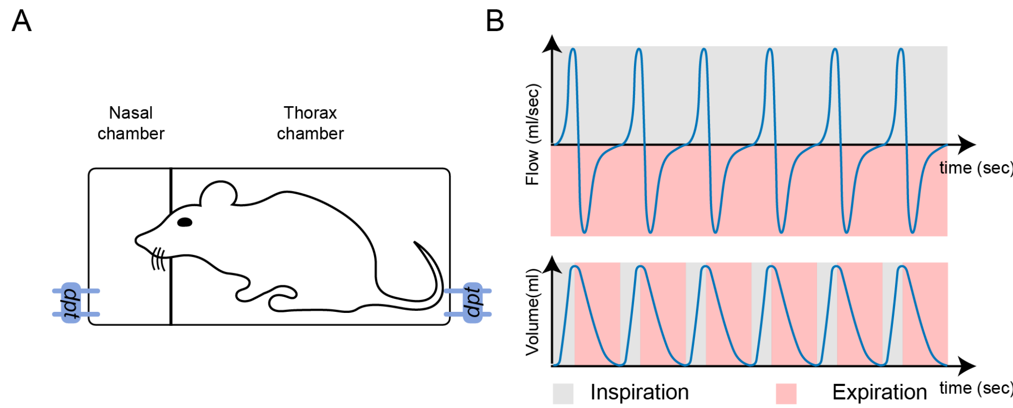

Figure 6: A: Illustration of a double-chamber plethysmograph. The term *dpt* stands for differential pressure transducer which measures the pressure in each compartment, the pressure then being converted to flow. B: Nasal airflow (top) and lung volume (bottom). During inspiration, airflow is positive (grey) and during expiration, airflow is negative (pink).

3. The mouse is placed back into the DCP, and its nasal flow is recorded for 35 or 40 min.

Respiratory cycles were extracted following procedure [17]. We removed respiratory cycles whose duration exceeds 1 second; the average respiratory cycle duration is 300 ms. We randomly sampled 10 respiratory cycles per minute and mouse. It leads to a dataset of 12,732 (time, genotype)-annotated respiratory cycles.

F Classification datasets

All datasets were taken from UCR/UEA archives [10, 2]. Among all available datasets[7], we selected 4 datasets related to time series shape comparison. All datasets were downloaded with the python package aeon[8] which already includes the train test split. Essential dataset information is summarized in Table 4.

Table 4: Time series datasets summary for shape based classification.

Dataset	Train size	test size	Lenghth	Number of classes	Number of dimensions	Type
ArrowHead	36	175	251	3	1	IMAGE
ECG200	100	100	96	2	1	ECG
GunPoint	50	150	150	2	1	MOTION
NATOPS	180	180	51	6	24	MOTION

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and precede the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

[7]https://timeseriesclassification.com
[8]https://www.aeon-toolkit.org/en/stable/

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS paper checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.

- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed

instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: **[TODO]**

Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: **[TODO]**

Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: **[TODO]**

Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained? .

Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.