Peak Persistence Diagrams for Signal Estimation under Additive and Warping Noise

Woo Min Kim[*,†], Sutanoy Dasgupta[‡], Pavan Turaga[⋆], and Anuj Srivastava[†]

[†]Department of Statistics, Florida State University, Tallahassee, FL 32306, USA

[‡]Department of Statistics, Texas A&M University, College Station, TX 77840, USA

[⋆]School of Arts Media and Engineering, Tempe, AZ 85287, USA

Abstract

Addressing the fundamental challenge of signal estimation from noisy data is a crucial aspect of signal processing and data analysis. Existing literature offers various estimators based on distinct observation models and criteria for estimation. This paper introduces an innovative framework that leverages topological and geometric features of the data for signal estimation. The proposed approach introduces a topological tool – *peak-persistence diagram* (PPD) – to analyze prominent peaks within potential solutions. Initially, the PPD estimates the unknown shape, incorporating details such as the number of internal peaks and valleys. Subsequently, a shape-constrained optimization strategy is employed to estimate the signal. This approach strikes a balance between two prior approaches: signal averaging without alignment and signal averaging with complete elastic alignment. Importantly, the proposed method provides an estimator within a statistical model where the signal is affected by both additive and warping noise. A computationally efficient procedure for implementing this solution is presented, and its effectiveness is demonstrated through simulations and real-world examples, including applications to COVID rate curves and household electricity consumption curves. The results showcase superior performance of the proposed approach compared to several current state-of-the-art techniques.

Keywords: signal estimation, peak persistence, elastic alignment, dynamic time warping, shape-constrained signal, shape estimation.

1 Introduction

Analyzing noisy data to estimate underlying signals is a central problem in signal processing and data processing. One observes an underlying true signal in form of curves that are corrupted by noise, and the

goal is to use curve data to estimate that signal. The choice of an estimator depends on the observation model, the noise distribution, the data structure, and the optimization criterion. The classical approach to signal estimation is identifying a signal space (typically a Hilbert space with an orthonormal basis), imposing an objective function, and optimizing it over the function space. We take a different approach. We will focus on the *shape* of the signal and introduce a novel topological feature, the *peak-persistence diagram* (PPD), to s earch for the estimate. This PPD enables us to estimate the shape of an unknown signal first, followed by a shape-constrained signal estimation.

We start the discussion from a traditional perspective and later make a case for a novel shape-based estimation. A simple, traditional model for signal estimation is the **additive noise model**: $f_i(t) = a_i g(t) + \epsilon_i(t)$, $i = 1, 2, \ldots, n$, for $t \in I$, where $g \in \mathcal{F}$ is a deterministic, unknown signal, $\epsilon_i \in \mathcal{F}$ is the zero mean, independent noise and $a_i \in \mathbb{R}_+$ is an independent random variable with mean one. Also, I is a compact observation interval. Here $\mathcal{F}$ denotes a signal space of interest; most commonly, $\mathcal{F}$ is the Hilbert space of square-integrable functions on an interval. In a simple case, with dense time samples, an estimate of g is given by the *cross-sectional mean* $\bar{f} = \frac{1}{n}\sum_{i=1}^{n} f_i$, because $\frac{1}{n}\sum_{i=1}^{n} \epsilon_i \to 0$ and $\frac{1}{n}\sum_{i=1}^{n} a_i \to 1$. The top left part of Fig. 1 shows a pictorial illustration of this estimation. Plot (a) shows the data $\{f_i\}$ and the true function g (in red), and (b) shows their cross-section mean superimposed in blue. In case of sparse data or other challenges, one can utilize a smooth basis for g or add a roughness penalty to modify the traditional least-squares criterion ([17, 8]).

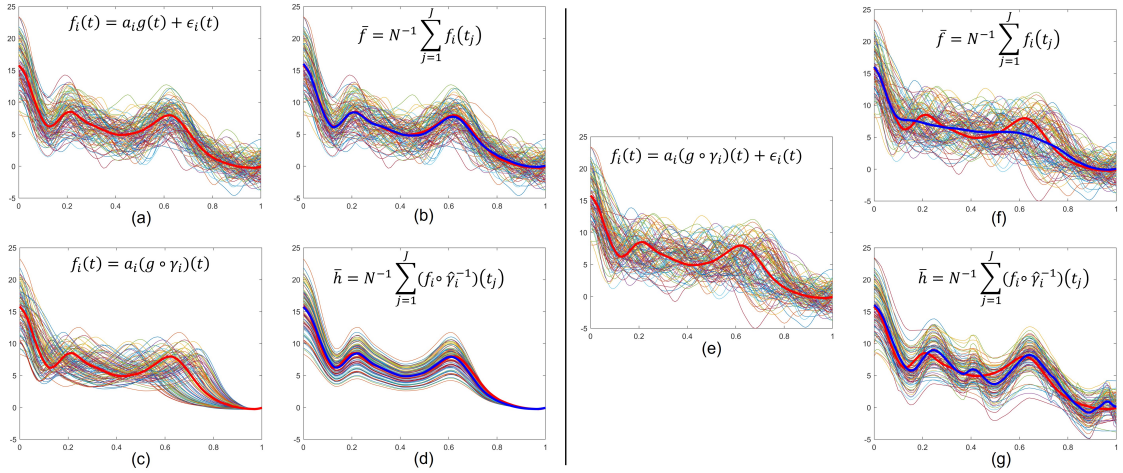


Figure 1: Illustration of different mean estimators on interval $I = [0, 1]$: Bold red lines denote the actual signal g, blue lines in (b),(f) denote the cross-sectional mean $\bar{f}$, and blue lines in (d),(g) denote the aligned mean $\bar{h}$. Functions in (a) have additive noise only, and those in (c) have phase noise only. Functions in (e) contain both additive and phase noises. The estimations in (b) and (d) perform well, but not the ones in (f) and (g).

In recent years, there has been a growing recognition of a different kind of noise or random variability

in observed signals; it is the so-called *phase* or *warping* noise. The observation model for this noise is given by the **phase noise model**: $f_i(t) = g(\gamma_i(t))$, $i = 1, 2, \ldots, n$, $tinI$, where γ_is are random warpings of the time domain. (A precise mathematical definition of time-warping is presented later.) To have identifiable g, one assumes that the phase noise has identity mean, *i.e.*, $\frac{1}{n}\sum_{i=1}^{n}\gamma_i(t) \to t$. One can verify that, in this case, the cross-sectional mean $\bar{f}$ does not directly estimate g. This is because $\frac{1}{n}\sum_{i=1}^{n} g(\gamma_i(t))$ does not converge to $g(t)$, in general, despite the fact that $(\frac{1}{n}\sum_{i=1}^{n}\gamma_i(t)) \mapsto t$. The quantity $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} g(\gamma_i(t))$ can be approximated by $\int g(t+x)\pi(x)\,dx$, where π is a normal density, and thus, the result is a Gaussian blurring of g. A better solution comes from solving for the γ_is explicitly and then performing an alignment of given data according to $\{\tilde{f}_i = f_i \circ \hat{\gamma}_i^{-1}\}$. This alignment is also called *phase-amplitude separation* in the literature ([23]). Several past signal processing approaches have used dynamic time warping or DTW to perform this separation. One can then estimate g using cross-sectional averaging of the *aligned signals* $\frac{1}{n}\sum_{i=1}^{n}\tilde{f}_i$ ([11]). We will call this average quantity the *fully-elastic* mean. The bottom left part of Fig. 1 shows a pictorial illustration of this estimation. Plot (c) shows the data $\{f_i\}$ and the true function g (in red), and (d) shows the aligned functions $\{\tilde{f}_i\}$ and their mean superimposed in blue. While several DTW approaches exist in the literature for signal alignment, a particularly efficient method is based on the nonparametric Fisher-Rao metric ([23]). It uses a square-root representation of signals and a dynamic programming (DP) algorithm to estimate $\{\hat{\gamma}_i\}$ and is remarkably successful in aligning peaks and valleys of f_is. The example shown in (d) uses this method.

The current paper focuses on a more general problem where *both* additive and phase noise are present in the observation. The data is now given by the **additive and phase noise model**:

$$f_i(t) = a_i(g \circ \gamma_i)(t) + \epsilon_i(t),\ i = 1, 2, \ldots, n, \text{ and } t \in I\,. \tag{1}$$

The components of the model remain the same as earlier. Next, we discuss the current state-of-the-art in handling this model.

1.1 Past Methods Including DTW

In signal processing literature, dynamic time warping, or DTW, is a classical technique for handling phase variability across datasets, especially where one expects periodic or cyclostationary observations. Sakoe and Chiba [19] popularized the use of DTW in speech recognition. [15] introduced a preprocessing tool for enhancing signal denoising, improving time-warping algorithm efficacy, and [9] estimated time-warping functions to transform irregularly cyclical data to regular form. [21] applied DTW for genomic signal alignment, while [24] used a Bayesian dynamic time warping to address Doppler-distorted signals in multipath

channels. Additionally, [14] characterized time-warping effects in almost-cyclostationary signals, focusing on periodic data. [13] employed DTW as a preprocessing method to align signals from individual trials to the averaged event-related potentials (ERPs) in electroencephalography (EEG) analysis. [16] introduced Progressive DTW for arterial blood pressure analysis as a solution to the limitations of traditional DTW, which struggles with noninvasive, continuous blood pressure estimation. [22] introduced a version of DTW that is robust to noise and outliers; it uses trend estimation and time warping in a multi-level framework for down-sampling and up-sampling.

From a statistical perspective, a seemingly natural idea to estimate the signal g is a penalized least-square solution:

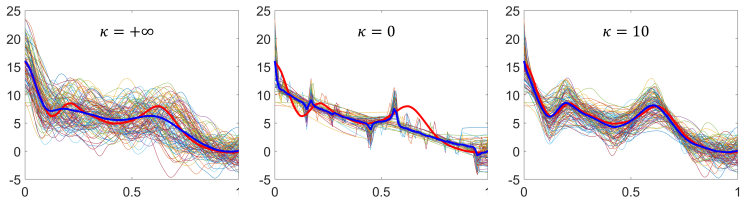
$$\underset{g\in\mathcal{F}}{\operatorname{argmin}} \left(\sum_{i=1}^{n} \min_{a_i\in\mathbb{R},\gamma_i\in\Gamma} \left(\|f_i - a_i(g\circ\gamma_i)\|^2 + \kappa\mathcal{R}(\gamma_i) \right) \right) , \tag{2}$$

where $\|f\| = \sqrt{\int_I f(t)^2\, dt}$ denotes the $\mathbb{L}^2$ norm and $\mathcal{R}$ denotes a roughness penalty on γ_is. This solution, when discretized and solved using the Dynamic Programming algorithm [2], closely relates to the classical DTW algorithm. However, this optimization has several problems. Firstly, for $\kappa = 0$, the optimization over γ_is is degenerate; this phenomenon is called the *pinching effect* ([17, 23]). While adding a penalty term (setting $\kappa > 0$) may avoid degeneracy, it does not entirely solve the problem. It only limits the search space for a solution and can have unintended consequences. Different values of κ may lead to widely different solutions.

Several papers have used Eqn. 2 but with a focus on aligning signal data and not necessarily estimating underlying signal. Others have studied a statistical mixed-effects model by adding a random effect term to Eqn. 1 ([17, 4]). A similar problem has been investigated in the registration of images ([1, 20]), albeit with focus on preserving visual features. Almost all these papers on DTW use the $\mathbb{L}^2$ objective function and must balance the two terms in Eqn. 2. Also, some papers have pursued a Bayesian approach for the alignment of given data $\{f_i\}$ by imposing a prior on γ_is ([10]).

In a classical DTW (as formulated in Eqn. 2), there are two main issues: (1) The problem is degenerate for $\kappa = 0$, giving rise to the *pinching effect*, and (2) How should one set the value of κ to avoid over smoothing and pinching? Although there are several ideas in the literature on choosing the smoothing parameter, pinching makes this selection complicated. One typically needs higher values of κ to avoid pinching but that can result in over smoothing. We illustrate this with an example using the first order roughness penalty $R(\gamma_i) = \int_I (1 - \sqrt{\dot{\gamma}_i(t)})^2 dt$. (The detailed algorithm is presented in [23].) Fig. 2 (a) displays the observed signals $\{f_i\}$ and their cross-sectional mean $\bar{f}$ (in blue). The red curve represents the ground truth, g, and the blue curves in plots (b)-(e) denote the cross-section mean of aligned curves, hereon referred to as $\hat{g}_{L^2}$. At $\kappa \in (0, 1)$, one can see the pinching effect in the estimate, but as κ increases, the estimate gets smoother.

This example shows that a carefully chosen κ is critical to avoid pinching and to reach a good estimate of g. The difficulty lies in finding the right κ from the data automatically.



(a) (b) (c)

Figure 2: Data alignment and signal estimation according to Eqn. 2. (a) The original data and their sample mean $\bar{f}$ (blue). (b,(c) The true g (red), the partially-aligned set, and the mean estimates (blue curves) for different κ.

1.2 Our Approach

Our approach is motivated by the fact that additive and phase noise models are special cases of Eqn. 1, and good solutions exist for each case individually. For additive noise only, direct or non-elastic averaging is optimal. For phase noise only, the fully elastic average is optimal. Under both types of noise, neither of these solutions is satisfactory or optimal. The cross-sectional mean over-smooths the data, while the fully elastic mean aligns even the noise artifacts, resulting in spurious peaks and valleys in the estimate. The right part of Fig. 1 shows an example: For the data in (e), plot (f) shows the non-elastic mean, and plot (g) shows the elastic mean. We argue that a good solution lies between these two extremes, and the balance should be adaptive to the data. To achieve this, we propose using partial alignments that control the elasticity of the signals during alignment via a parameter $\lambda \in \mathbb{R}_+$ (similar to κ mentioned above but under a different metric). The resulting partially-aligned signals are denoted by $\tilde{f}_{\lambda,i}$, and their mean is $\hat{g}_\lambda = \sum_{i=1}^{n} \tilde{f}_{\lambda,i}$. An important consideration is making λ adaptive to the given data.

Additionally, we must address the issue of degeneracy or the pinching effect. Our approach differs from Eqn. 2 in that the formulation is based on the Fisher-Rao distance rather than the $\mathbb{L}^2$ distance, thus avoiding the pinching problem altogether. However, the issue of choosing the smoothing parameter remains, and we tackle it using a novel topological feature defined next. Our approach has two main steps.

1. **Peak-Persistence Diagram** (PPD): The objective of this stage is twofold: to find an optimal λ and to estimate the *shape* of the unknown signal g. We introduce a novel topological tool called a Peak Persistence Diagram (PPD). When we change the parameter λ, the topological features, such as the number of peaks and valleys, in the average function $\hat{g}_\lambda$ change. Therefore, one can estimate λ by studying the persistence of the internal peaks of $\hat{g}_\lambda$ versus λ and selecting the most persistent peaks.

A display of this persistence is called a *Peak Persistence Diagram* or PPD. This process yields three quantities: an optimal λ^*, the number m of persistent internal peaks, and $\{\tilde{f}_{\lambda^*,i}, i = 1, 2, \ldots, n\}$, the set of partially-aligned signals under the optimal weight λ^*.

2. **Shape-Constrained Function Estimation**: In the second step, we find an optimal estimate of g by constraining it to have precisely m internal peaks and utilize a geometric approach to perform a shape-constrained estimation of g. The estimation process involves minimizing the squared error over the appropriate constraint space $\mathcal{F}_m$, using the data $\tilde{f}_{\hat{\lambda},i}$. Here, $\mathcal{F}_m$ refers to the set of all elements of $\mathcal{F}$ with m internal peaks.

This paper focuses on only the *internal* peaks and valleys of g. The potential peaks at the boundaries are also relevant, but they can be detected separately using simple tests and are ignored here. For the remainder of this paper, we will assume that signal data is defined on a compact time interval I and $\mathcal{F}$ is the set of all absolutely-continuous, real-valued functions on I. We will consider the model stated in Eqn. 1, with following independent components: (1) Scaling noise: $a_i \in \mathbb{R}_+ \sim N(1, \sigma_a^2)$; (2) Additive noise: $\epsilon_i \in \mathcal{F}$, a random function with mean $E[\epsilon_i(t)] = 0$ for all $t \in I$; (3) Phase noise: $\gamma_i \in \Gamma$ (where Γ is defined below), with the mean $E[\gamma_i(t)] = \gamma_{id}(t) = t$. Given a set of independent observations $\{f_i\}$, our goal is to estimate the number of internal peaks in g and the function g itself.

2 Proposed Method Step 1: Shape Estimation Using PPDs

In this step, we estimate the shape of g, *i.e.*, estimate the number of internal peaks and valleys in g, from the given data $\{f_i\}$ and form an initial estimate of g.

Background Material: We summarize a recent framework for partial elastic alignments of signals under the Fisher-Rao metric and refer the reader to the book [23] for details. For alignment purposes, a signal $f \in \mathcal{F}$ is represented by its square-root velocity function (SRVF): $q(t) = \text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$. Let Γ denote the group of all boundary-preserving diffeomorphisms of I. Any element $\gamma \in \Gamma$ is a smooth function with a smooth inverse and preserves the boundaries of I. The set Γ forms a group under composition, *i.e.*, for any $\gamma_1, \gamma_2 \in \Gamma$, we have $\gamma_1 \circ \gamma_2 \in \Gamma$. This group has an identity element $\gamma_{id}(t) = t$. Let a signal $f \in \mathcal{F}$ be time-warped by a $\gamma \in \Gamma$, resulting in $f \circ \gamma$. The SRVF of the $f \circ \gamma$ is given by $(q \circ \gamma)\sqrt{\dot{\gamma}}$, and denoted by $q \star \gamma$ for brevity.

With this setup, we can solve for elastic alignment of any two signals $f_1, f_2 \in \mathcal{F}$ as follows. Let q_1, q_2 denote the SRVFs of f_1, f_2, respectively. Then, the optimal time warping to align f_2 to f_1 is given by: $\gamma^* = \text{argmin}_{\gamma \in \Gamma} \|q_1 - (q_2 \star \gamma)\|^2$. (Recall that $\|\cdot\|$ denotes the $\mathbb{L}^2$ norm.) However, if we want to penalize

the level of elasticity in their alignment, we can do so using: $\text{argmin}_{\gamma\in\Gamma}\left(\|q_1 - (q_2\star\gamma)\|^2 + \lambda\mathcal{R}(\gamma)\right)$. The term $\mathcal{R}(\gamma)$ denotes a penalty on the roughness of γ and forces it to be close to the identity element γ_{id}. In this paper, we have chosen the first-order penalty $\mathcal{R} = \|1 - \sqrt{\dot{\gamma}}\|^2$, but one can use other penalties instead. The constant $\lambda > 0$ controls the amount of elasticity in the alignment: $\lambda = 0$ is fully elastic and $\lambda = \infty$ is fully non elastic. To align multiple signals, say $f_1, f_2, \ldots, f_n$, let $q_1, q_2, \ldots, q_n$ denote their respective SRVFs. Then, a joint alignment of these signals is performed using the following optimization:

$$\min_{q\in\mathbb{L}^2}\left[\sum_{i=1}^{n}\left(\min_{\gamma_i\in\Gamma}\left(\|q - (q_i\star\gamma_i)\|^2 + \lambda\mathcal{R}(\gamma_i)\right)\right)\right]. \tag{3}$$

We can rearrange this equation using the facts that $\|q_1 - q_2\star\gamma\| = \|q_1\star\gamma^{-1} - q_2\|$, for all $q_1, q_2 \in \mathbb{L}^2$ and $\gamma\in\Gamma$, and that Γ is a group. This results in:

$$\min_{q\in\mathbb{L}^2}\left[\sum_{i=1}^{n}\left(\min_{\gamma_i\in\Gamma}\left(\|q_i - (q\star\gamma_i)\|^2 + \lambda\mathcal{R}(\gamma_i)\right)\right)\right],$$

which can be directly compared to Eqn. 2. We ignore the optimization over a_is here as their average converges to one and disappears in any mean-based estimation. The main difference between this solution and Eqn. 2 (or other DTW approaches) lies in the use of SRVFs and the invariance properties of the elastic metric. Fundamentally, the difference comes from the fact that $\|q\| = \|q\star\gamma\|$, for all $q\in\mathbb{L}^2$ and $\gamma\in\Gamma$, while $\|f\| \neq \|f\circ\gamma\|$ in general.

The SRVF-based optimization (Eqn. 3) does not provide an estimate of g directly but results in several quantities of interest: (1) If $\hat{\gamma}_{\lambda,i}$ denotes the optimal time warping inside the summation, for each i, then $\tilde{f}_{\lambda,i} = f_i\circ\hat{\gamma}_{\lambda,i}$ are the resulting partially-aligned functions; (2) Let $\hat{g}_\lambda$ be the cross-sectional mean of these $\tilde{f}_{\lambda,i}$. Algorithm 1 summarizes the main steps in this partial elastic alignment. Fig. 3 shows an illustration of the output of Algorithm 1. It offers several results, each showing partially-aligned signals and their cross-sectional mean for a different value of λ. The original data is shown in the bottom right panel. On one extreme, $\lambda = 0$ results in a perfect alignment of peaks and valleys. Conversely, $\lambda = \infty$ provides no alignment at all. Importantly, note that there is no pinching effect in the alignment for any value of λ. Note that $\hat{g}_{infty} = \bar{f}$, the cross-sectional mean mentioned earlier.

2.1 Peak Persistence Diagrams

Given this context, the subsequent task involves discovering an automated method to select the best λ from the available data. This is achieved via a topological approach, by utilizing the concept of peak persistence diagrams. As λ changes from 0 to ∞, some peaks and valleys in $\hat{g}_\lambda$ start to diminish and even disappear

Algorithm 1 Signal-Alignment Algorithm

Require: Data $\{f_1, f_2, \ldots, f_n\}$

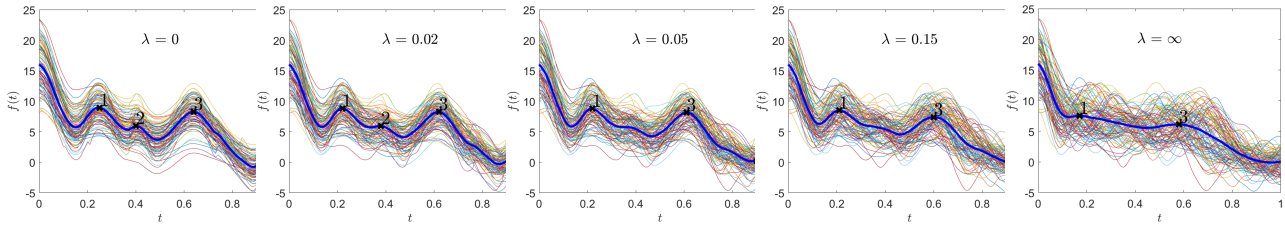
for $i = 1, \ldots, n$ **do**
  $q_i \leftarrow \text{sign}(\dot{f}_i)\sqrt{|\dot{f}_i|}$
end for
$\bar{q} \leftarrow \text{argmin}_{q \in \{q_1, \ldots, q_n\}} \left\| q - \frac{1}{n}\sum_{i=1}^{n} q_i \right\|$
while $\epsilon > tol$ **do**
  for $i = 1, 2, \ldots, n$ **do**
    $\hat{\gamma}_{\lambda,i} \leftarrow \arg\inf_{\gamma \in \Gamma} \left(\|\bar{q} - (q_i \star \gamma)\|^2 + \lambda \mathcal{R} \right)$
    $f_i^* \leftarrow f_i \circ \hat{\gamma}_{\lambda,i}$, and $q_i^* \leftarrow \text{sign}(\dot{f}_i^*)\sqrt{|\dot{f}_i^*|}$
  end for
  $\bar{q}^* \leftarrow \frac{1}{n}\sum_{i=1}^{n} q_i^*$, $\epsilon \leftarrow \|\bar{q} - \bar{q}^*\|^2$
  if $\epsilon > tol$ **then**
    $\bar{q} \leftarrow \bar{q}^*$
  end if
end while
$\bar{\gamma}^{-1} \leftarrow (\frac{1}{n}\sum_{i=1}^{n} \hat{\gamma}_{\lambda,i})^{-1}$, $\bar{q}^* \leftarrow (\bar{q}^* \star \bar{\gamma}^{-1})$
for $i = 1, 2, \ldots, n$ **do**
  $\hat{\gamma}_{\lambda,i} \leftarrow \arg\inf_{\gamma_i \in \Gamma} \left(\|\bar{q}^* - (q_i \star \gamma_i)\|^2 + \lambda \mathcal{R} \right)$
  $\tilde{f}_{\lambda,i} \leftarrow f_i \circ \hat{\gamma}_{\lambda,i}$
end for
return $\hat{g}_\lambda = \frac{1}{n}\sum_{i=1}^{n} \tilde{f}_{\lambda,i}$

altogether. Sometimes new peaks may also be formed. We propose to analyze this behavior using a peak persistence diagram.

Definition 1 *The peak persistence diagram (PPD) of a set of signals* $\{f_i, i = 1, 2, \ldots, n\}$ *is a visual presentation of the* **significant, internal peaks** *of their partial elastic mean* $\hat{g}_\lambda$ *plotted versus* λ*. Essentially, a PPD serves to identify the existence, magnitudes, and positions of significant internal peaks of* $\hat{g}_\lambda$*.*

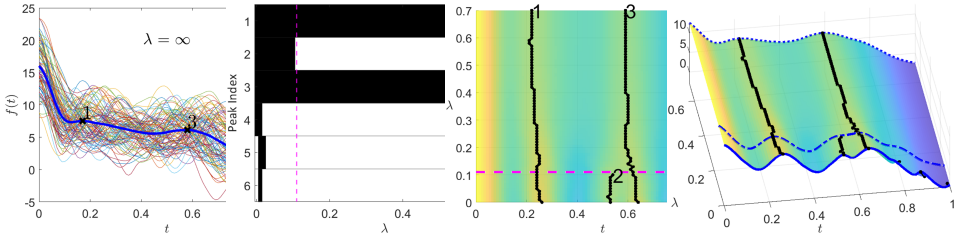
A PPD can be displayed in several ways. Plotting only the peak indicators across the range of λ, we obtain a PPD bar chart. If we are also interested in peak locations versus λ, we obtain a PPD surface. A vital issue to address is what qualifies as a *significant* peak, which we discuss in the next section. The idea of tracking significant peaks has been previously utilized in density estimation ([3]), although in the context of bandwidth selection for kernel-based methods. A PPD is akin to persistence homology in topological data analysis ([25]), where one traces topological features at different data resolutions. In contrast, a PPD tracks peaks of registered signals.

Figure 4 displays PPD charts for the data shown in panel (a). Panel (b) plots the PPD bar chart indicating detected peaks for different values of λ. The x axis represents λ, while the y axis indicates peak labels. The black color denotes significant peaks, and grey regions denotes insignificant peaks. Panels (c) and (d) show a 3D surface plot, where the values of $\hat{g}_\lambda(t)$ are represented by colors ranging from blue to yellow, based



(a) (b) (c) (d) (e)

Figure 3: Partial elastic alignments for different values of λ according to Eqn. 3. Each plot shows the true g (red), the partially aligned set $\{\tilde{f}_{\lambda,i}\}$, and their cross-sectional mean $\hat{g}_\lambda$ (blue curve). Compare with Fig. 2 for a visual comparison of results.



(a) (b) (c) (d)

Figure 4: (a) The original signalal data, (b) the PPD barchart, and (c), (d) PPD surface generated by interpolating $\hat{g}_\lambda$'s for different λ and the movements of the peaks.

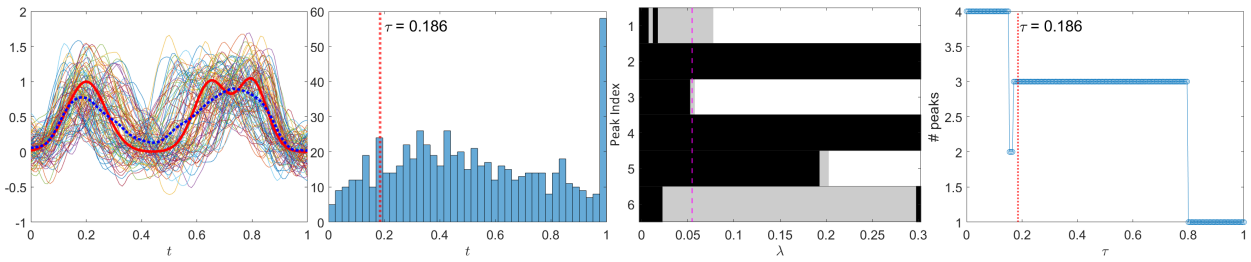
on their height. The y axis is λ, the x axis is the time t, and the labeled black lines indicate the t positions of the peaks as λ varies. At $\lambda = 0$, $\hat{g}_\lambda$ exhibits four internal peaks, but as λ increases the peaks labeled 2 and 4 disappear. Only those labeled 1 and 3 persist throughout indicating two dominating peaks.

2.2 Defining Peak Significance and Persistence

Peak Significance Parameter τ: Several reasons, including noise or numerical errors, can introduce small peaks in $\hat{g}_\lambda$. This can happen, for example, in parts of the domain where g is constant and an alignment of noise results in spurious peaks. Our goal is to keep only those peaks that are significant in some way. The question is: When is a peak considered significant? We define the *strength* of a peak located at $t_0 \in I$ to be the value $\frac{-\hat{g}''_\lambda(t_0)}{\max_{t\in I} -\hat{g}''_\lambda(t)}$. This quantity measures the curvature of $\hat{g}_\lambda$ at point t_0, normalized appropriately. If this quantity is less than a predetermined value, say τ, then the peak at t_0 is considered *insignificant*, and we discard it. Otherwise, it is significant, and we include it in PPD. To determine the value of cutoff τ from the data, we measure the strengths $\left\{\frac{-f_i''(t_{i,j})}{\max_{t\in I} -f_i''(t)}\right\}$, at all the peaks located at $\{t_{i,j}\}$ of all the original signals $\{f_i, i = 1,\ldots,n\}$. We then create an empirical distribution (a histogram) to assess the variability of strengths in the original data. We then set τ as the 15th percentile of the histogram. We have found the results robust for this choice – a range from 10 to 40 percentile works well in all case studies.

Peak Persistence Parameter λ^*: The next challenge is to decide when a peak is labeled as persistent. We define the *persistence measure* of a peak to be the range of λ-values in which that peak is significant, *i.e.*, $p_k = |\{\lambda \geq 0 \mid \text{peak } k \text{ in } \hat{g}_\lambda \text{ is significant}\}|$. Strong peaks in the signal g should appear in $\hat{g}_\lambda$ over a large range of λ. One can label a peak k as *persistent* if its persistence measure p_k exceeds a threshold λ^*, but how should one determine λ^*? Since persistence and non-persistence are the only two available labels for any peak, we use a clustering method to divide the set $\{p_k\}$ into two groups: one with large values is labeled persistence, and the other is non-persistent. (We use Ward's method for hierarchical clustering.) Then, we select the smallest value of λ that results in those persistent peaks in $\hat{g}_\lambda$ and call it λ^*. This choice is motivated by the observation that extrema of $\hat{g}_\lambda$ typically diminish as λ increases. The number of persistent peaks is denoted by m.

Fig. 5 shows these steps with a simulated dataset. Plot (a) shows the data, its cross-sectional mean (dotted blue line), and the ground truth g (solid red line). Plot (b) shows a histogram of strengths of peaks in the given data $\{f_i\}$, with the 15^{th} percentile located at $\tau = 0.186$ (indicated by a vertical dotted line). Plot (c) shows the PPD bar chart of peak indicators with significance (black) and insignificance (grey) color coded. From PPD, we compute persistence values, $p_1, ..., p_6$ for all the six peaks and we cluster them into two groups. This identifies peaks numbered 2, 4, and 5 as persistent, resulting in $m = 3$ and $\lambda^* = 0.05$. For this λ^*, we also compute the cross-sectional mean $\hat{g}_{\lambda^*} = \frac{1}{n}\sum_{i=1}^{n} \tilde{f}_{i,\lambda^*}$. Finally, plot (d) shows the variation in the estimated number of significant peaks versus τ, highlighting this estimate's robustness.



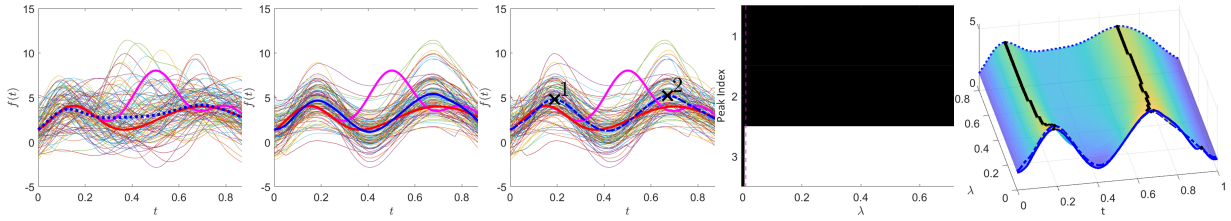
(a) (b) (c) (d)

Figure 5: (a) The original signals, their mean (dotted blue line), and the underlying g (red line). Plot (b) The histogram of *strengths* of all peaks in the original dataset $\{f_i\}$. The vertical line shows the 15^{th} percentile, and the corresponding τ at 0.186. (c) The resulting PPD plot with gray denoting insignificant peaks and black denoting significant peaks. (d) The plot of the number of peaks selected by PPD for different values of τ, underscoring the robustness of results to τ.

2.3 Experimental Results on PPD-Based Peak Estimation

Now we present examples of PPD-based shape estimation on a simulated and a real dataset. Note that additional experimental results are presented in the Supplementary Material.

Example 1: Here we generate data from a bimodal signal, labeled as g_1 (red curve), which we corrupt by adding a few tri-modal curves g_2 (magenta curve), and randomly time warpg them and add some noise. The complete data consists of 80 random perturbations of g_1 and 20 of g_2. The objective is to estimate m from this noisy data.

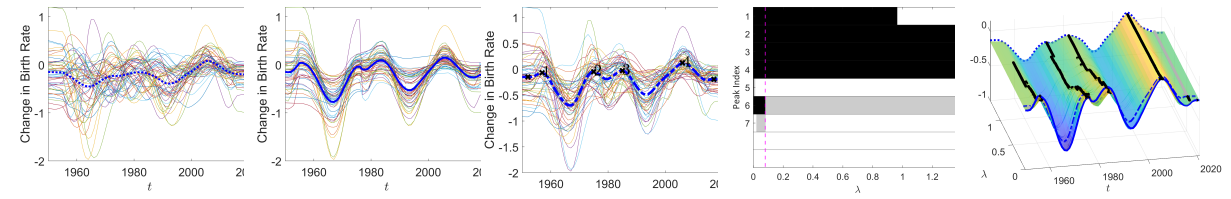


(a) (b) (c) (d) (e)

Figure 6: Example 1: Data consists of 80 and 20 signals generated from bimodal g_1 and tri-modal g_2, respectively. Plots (a), (b), and (c) display $\hat{g}_\infty$, $\hat{g}_0$, and $\hat{g}_{\lambda^*}$, respectively. PPDs in (d) and (e) suggest two peaks (1 and 2) are significant and persistent. Figure Key: Solid Red - g_1, Solid Magenta - g_2, Dotted Blue - $\hat{g}_\infty$, Solid Blue - $\hat{g}_0$, Dashed Blue - $\hat{g}_{\lambda^*}$.

Figure 6 presents PPD outputs of this data. Plot (a) shows the original data, and (b) shows the full elastic alignment with the mean, $\hat{g}_0$, drawn in blue. This mean curve has three peaks: two dominant and a minor third peak. Plot (c) shows the outcomes of partial-elastic alignment with $\lambda^* = 0.01$ chosen via PPDs displayed in panels (d) and (e). The bar chart in (d) indicates two significant and persistent peaks (1 and 2), consistent with g_1. This example suggests that the PPD method can successfully estimate the shape of the underlying signal even when the data is contaminated by a differently shaped signal.

Example 2: Birth Rate Changes in European Countries: Here, we study historical records of birth rate changes in 50 European countries during $1950 - 2021$ [18]. The response variable here is the yearly birth rate, defined as the count of live births per 1000 individuals each year. We smooth the raw data first using a window size of 15 years. As shown in Fig. 7, the PPD finds that $m = 4$ peaks are significant and persistent. The optimal parameter for this peak is $\lambda^* = 0.09$. Plot (b) shows that $\hat{g}_0$ exhibits five peaks, but the fifth peak disappears quickly as the value of λ increases. On the other hand, (a) and (d) show that $\hat{g}_\infty$ has four shallow peaks. However, (c) shows $\hat{g}_{\lambda^*}$ with four significant peaks.



(a) (b) (c) (d) (e)

Figure 7: Example 2: The estimated number of peaks by PPDs is four (peaks 1,2, 3, and 4). Figure Key: Solid Red - g_1, Solid Magenta - g_2, Dotted Blue - $\hat{g}_\infty$, Solid Blue - $\hat{g}_0$, Dashed Blue - $\hat{g}_{\lambda^*}$.

3 Proposed Method Step 2: Peak-constrained Curve Estimation

So far, we have partially aligned observations, and their sample mean $\hat{g}_{\lambda^*}$. While this may be a candidate for estimating g, it does not have any optimality criterion attached to it. Therefore, we introduce a second step of the framework that involves searching over a pre-defined shape class, *i.e.*, having m peaks under a penalized maximum-likelihood criterion. The objective here is to limit the exploration only to signals with m peaks and determine the optimal element of that class through geometric search. This technique adapts the shape-constrained density estimation method suggested in [5] and has been customized for our signal estimation. It involves altering an initial estimate $\hat{g}_{init}$ while maximizing a penalized log-likelihood function. (We use the peaks and valleys, including locations and heights, in $\hat{g}_{\lambda^*}$ and smooth interpolations between them to form an initial estimate, $\hat{g}_{init}$.) The optimization is conducted over all signals with m peaks.

3.1 Constrained Signal Estimation using Penalized-MLE

Let M be the number of *extrema points* of $\hat{g}_{init}$, which comprises interior peaks, valleys, and the two endpoints. M lies between $2m+1$ to $2m+3$ (inclusive), where m is the estimated number of interior peaks. For instance, if both endpoints are minima and there are $m-1$ interior valleys, then $M = 2m+1$.

Set of Shape-Constrained Signals: Define $\mathcal{F}_m \subset \mathcal{F}$ as the set of all signals with m internal peaks on I. Any two elements of $\mathcal{F}_m$ differ in the locations and heights of their extrema points. Correspondingly, we are going to define two sets of variables. **Locations**: We will use the time-warping functions – elements of Γ – to vary the locations of the extrema points. That is, the composition $\hat{g}_{init} \circ \gamma$, for a $\gamma \in \Gamma$, changes the locations of the extrema points while maintaining their heights and order. **Heights**: We also define $\mathbf{s} = \{s_1, s_2, \ldots, s_M\}$ as the heights of the extremal points in $\hat{g}_{init}$. There are some natural constraints on the values of $\mathbf{s}$. A valley's height should be less than the heights of its neighboring peaks. Let $\mathcal{S}$ be the set of all vectors $\mathbf{s}$ that satisfy these constraints. We will use $f_{\mathbf{s}} \in \mathcal{F}_m$ to denote a signal that has $\mathbf{s}$ as the height vector of its extrema.

Starting with the initial estimate $\hat{g}_{init}$, we adjust the locations and the heights of the extrema points in order to explore the set $\mathcal{F}_m$. The final estimate is given by: $\hat{g} = (f_{\mathbf{s}^*} \circ \gamma^*)$, where

$$(\gamma^*, \mathbf{s}^*) = \underset{\gamma \in \Gamma, \mathbf{s} \in \mathcal{S}}{\operatorname{argmin}} \left(\sum_{i=1}^{n} \int_0^1 (f_{\mathbf{s}}(\gamma(t)) - \tilde{f}_{\lambda^*,i}(t))^2 dt + \rho \int_0^1 \ddot{f}_{\mathbf{s}}(\gamma(t))^2 dt \right) \tag{4}$$

Here $\rho > 0$ is an infinitesimal weight ($\sim 10^{-8}$) for favoring smooth functions. This solution differs from

Eqn. 2 in several ways. Here we optimize over only one γ while Eqn. 2 uses a γ_i for each observation. Also, we use the partially aligned data that favors m peaks rather than using the original data.

The optimization problem in Eqn. 4 is solved using the `fmincon` function in Matlab. However, since Γ is a nonlinear manifold of infinite dimension, direct optimization poses a challenge. To overcome this, we employ an SRVF map followed by an inverse exponential map to represent the warping functions in a vector space, and use an orthogonal basis to represent γ by its coefficients. Interested readers can find the optimization details in [6]. Algorithm 2 outlines the steps for shape-constrained estimation of g and Fig. 8 shows an example. The gray dotted points represent the partially-aligned signals $\{\tilde{f}_{\lambda^*,i}\}$, while the initial estimate $\hat{g}_{init}$ is shown in cyan, and the outcome of Algorithm 2 is depicted in green.

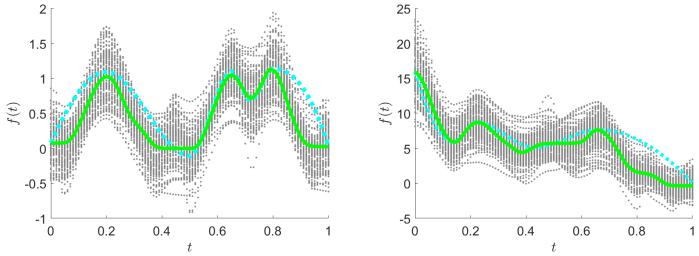


Figure 8: Two examples of shape-constrained signal estimation: In each case, the dotted points represent the partially-aligned functions $\{\tilde{f}_{\lambda^*,i}\}$. The cyan and the green lines are the initial and final function estimates of g.

Algorithm 2 Peak-constrained Curve Estimation

Require: Data: $\{\tilde{f}_{\lambda^*,i}\}_{i=1}^n$, initial shape estimate $\hat{g}_{init}(t)$
1: Initialize the heights of the extrema points of $\hat{g}_{init}$: $\{s_1, s_2, ..., s_M\}$,
2: Initialize $\gamma \leftarrow \gamma_{id}$.
3: Obtain γ^* and $\mathbf{s}^*$ by solving Eqn. 4 by `fmincon`.
4: **return** $\hat{g}(t) = f_{\mathbf{s}^*}(\gamma^*(t))$

3.2 Bootstrapping for Estimating Confidence Bands

To evaluate the performance of the signal estimate $\hat{g}$, we compute pointwise confidence bands for our estimator and display them with the estimation results. In the case of simulated data, we check if these estimated confidence bands contain the ground truth and validate the results. Algorithm 3 lays out the steps for computing these confidence bands.

Algorithm 3 Constructing Pointwise Confidence Band

Require: $\{f_i\}_{i=1}^n$, Initial shape estimate, $\hat{g}_{init}$, Estimated smoothing parameter, λ^*

1: **for** $j = 1, 2, \ldots, B$ **do**
2: Resample n observations with replacement to obtain a bootstrapped set of functions, $\{f_i^{(j)}\}_{i=1}^n$.
3: Obtain partially aligned data with smoothing parameter, λ^*, $\{\tilde{f}_{\lambda^*,i}^{(j)}\}_{i=1}^n$
4: Given $\hat{g}_{init}$ and $\{\tilde{f}_{\lambda^*,i}^{(j)}\}_{i=1}^n$, obtain the best estimate of g by Algorithm 2: $\hat{g}^{(j)}$
5: **end for**
6: With $\{\hat{g}^{(j)}\}_{j=1}^B$, **return** $(\alpha/2)$ and $(1-\alpha/2)$ quantiles to set the lower and upper bound of the confidence bound point-wisely.

4 Experimental Results: Shape and Signal Estimation

This section demonstrates and evaluates the performance of our proposed method using a range of simulated and real datasets. We primarily compare our estimate $\hat{g}$ with two prior solutions, which include (1) the mean of unaligned functions, $\bar{f} = \hat{g}_\infty$, and (2) the fully-elastic mean, $\hat{g}_0$. In addition, we study the results obtained using Eqn. 2 (denoting DTW solutions) in the case of simulated data for several κ values. We use RMSE as the metric for comparing the estimation errors and provide 95% pointwise bootstrap confidence bands to visualize the variability of our estimator.

4.1 Simulation Studies

First, we try two simulation scenarios on the domain $I = [0, 1]$. (Additional experimental results are presented in the Supplementary Material). In each experiment, we choose a different signal g and generate $n = 100$ samples according to Eqn. 1. To compute and compare different solutions, we use 50 independent replications of each scenario and compute statistics of RMSEs of the estimators.

1. ***Simulated Dataset 1***: We set the true signal g as:$g(t) = -4 + 6\phi(t, -0.05, 0.14) + 1.5\phi(t, 0.25, 0.075) + 5.8\phi(t, 0.59, 0.195)$, where $\phi(t, \mu, \sigma)$ denotes the Gaussian probability density function evaluated at $t \in I$ with mean μ and standard deviation σ. Figure 9 (top) shows results from this experiment. The original data is shown in panel (a), with the true function g drawn in red. Plots (a) and (b) show the alignment of functions under the extreme values of λ: ∞ and 0, respectively. The cross-sectional means are shown in blue in each case. The PPD bar chart in (c) estimates two significant and persistent internal peaks. The optimal parameter $\lambda^* = 0.11$ is shown in the magenta. Plot (d) is the PPD surface displaying gradual changes in $\hat{g}_\lambda$ when λ increases. The black lines on this surface are traces of significant peaks. Next, we use these quantities to estimate the signal g as described in Section 3. The green dashed curve in (e) is the final estimate $\hat{g}$. We also display a pointwise bootstrap confidence band using a gray region. We can see that both the function estimate (in green) and the band (in gray)

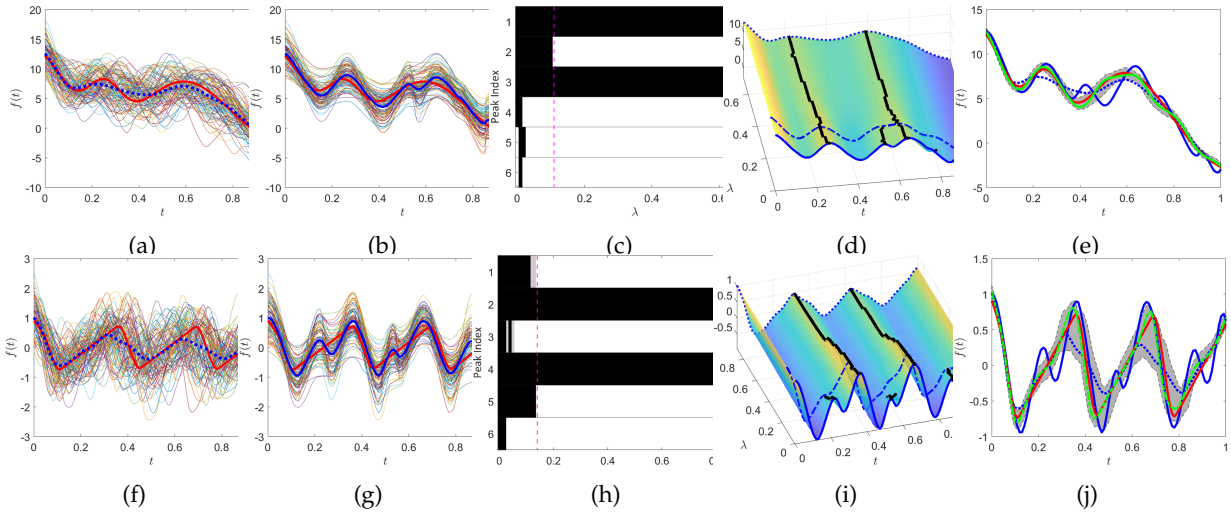


Figure 9: (Sumlated Data), top row is Datset 1 and bottom row is Datset 2. In each row, (a) and (b) are aligned functions at $\lambda = \infty$ and $\lambda = 0$. These figures also show $\hat{g}_0$ (bold blue), $\hat{g}_{0.05}$ (bold dotted-line blue) and $\hat{g}_\infty$ (bold dotted blue). The ground truth g and the final estimate $\hat{g}$ are drawn in red and green, respectively. Plot (c) is the PPD with $\lambda^* = 0.11$. Plot (d) shows the evolution of $\hat{g}_\lambda$ versus λ. The black lines on the surface indicate the interior peaks. Panel (e) compares different estimates with the ground truth g. A 95% bootstrap confidence band is drawn using gray regions. Figure Key: Solid Red - g, Dotted Blue - $\hat{g}_\infty$, Solid Blue - $\hat{g}_0$, Dashed Blue - $\hat{g}_{\lambda^*}$, Dashed Green: $\hat{g}$.

differ from $\hat{g}_0$ and $\hat{g}_\infty$. Furthermore, $\hat{g}$ is the closest function to the ground truth g in terms of the $\mathbb{L}^2$ distance. A detailed quantitative evaluation of the results is presented later.

2. ***Simulated Dataset 2***: In this example, the signal g is $g(t) = \text{sawtooth}(6(t - 0.07)\pi)$, where the matlab function 'sawtooth' generates a sawtooth wave pattern. To handle the non-differentiability at the sawtooth wave's interior extrema, we employ a locally weighted scatterplot smoothing (LOWESS) with a span of 12 data points to smooth the data. This helps to eliminate any sharp spikes or dips in the signal and provides a smoother signal. The data is presented in Fig. 9(f). The PPD bar chart in (h) successfully detects two significant internal peaks, labeled 2 and 4, and estimates $\lambda^* = 0.14$. The evolution of peaks is seen in bottom (i with PPD surface plot. Plot (j) shows the final estimate $\hat{g}$ (in green) with a bootstrap confidence band in gray. As shown, the elastic mean, $\hat{g}_0$ (solid blue), has spurious peaks while $\hat{g}_\infty$ (dotted blue) underestimates the heights of peaks and valleys. Our estimate $\hat{g}$ (in green) provides an excellent estimate of g.

Performance Quantification: So far we have visualized estimates from different methods. To quantify and compare estimation performances, we generate 50 replications of each experiment (discussed above) and calculate the $\mathbb{L}^2$-based RMSE $\sqrt{\int_I (g(t) - \hat{g}(t))^2 \, dt}$ for each replication. Each row in Fig. 10 shows results for a simulation experiment. The left two panels display boxplots of the mean and standard deviations of the RMSE for different estimators. The first plot shows boxplots for $\hat{g}_\infty$, $\hat{g}_0$, and $\hat{g}$. As expected, $\hat{g}$ has the

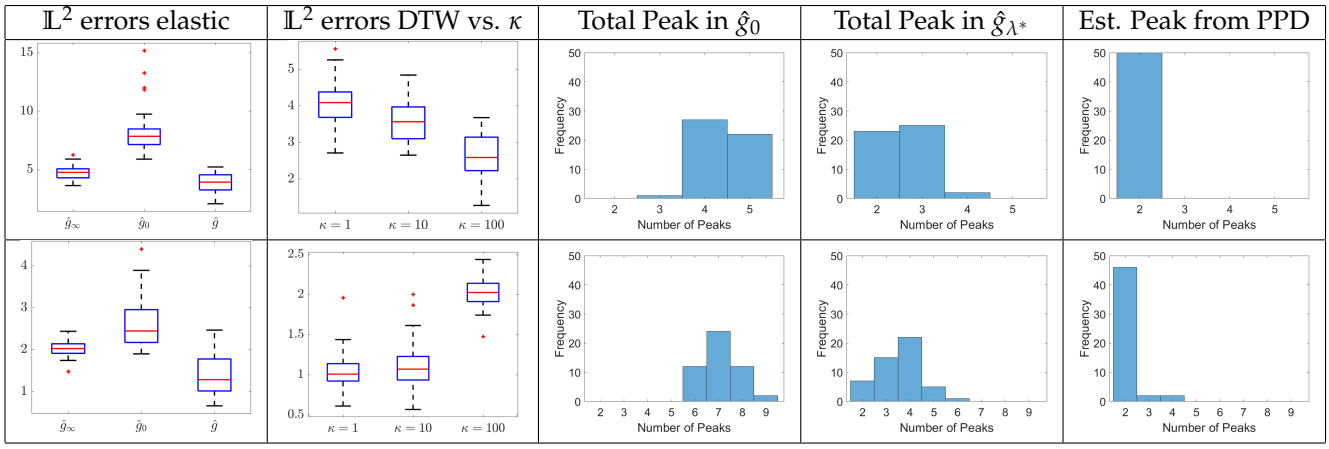


Figure 10: Figure summarises results from two simulation experiments using 50 replicates. Each row shows a different simulation, from left to right: (1) the boxplots of the RMSE of three elastic estimators, (2) the boxplots of RMSE for the penalized-$\mathbb{L}^2$ estimators (for three different values of κ), (3) histograms of total peak counts in $\hat{g}_0$, (4) histograms of total estimated peak counts (both insignificant and significant) in $\hat{g}_{\lambda^*}$, and (5) histograms of dominant (persistence and significant) peaks detected using PPD.

smallest error as λ^* is estimated from the data. This also supports the prior observation that $\hat{g}_\infty$ underestimates signal structure (peaks and valleys and their heights), and $\hat{g}_0$ tends to overestimate structures. The second column shows the error statistics for the DTW technique under three values of κ, at 1, 10, and 100. The boxplots of both cases highlight the difficulty in choosing an optimal κ.

To emphasize the role of shapes, we also display the histograms of the estimated number of peaks for each experiment. The first histogram shows the total number of peaks in fully aligned mean, $\hat{g}_0$. The second histogram shows the total number of peaks (both significant and insignificant) in the partially aligned mean, $\hat{g}_{\lambda^*}$. The third shows the number of the significant and persistent peaks selected by PPD. In both simulation cases, the ground truth has two peaks. As these results show, the PPD approach is better able to estimate the correct number of peaks in g than $\hat{g}_{\lambda^*}$ which, in turn, is better than $\hat{g}_0$. This partly explains the improved performance of the proposed shape-constrained function estimator.

Finally, we analyze the computational cost of the past and proposed methods. Table 1 lists the time to perform different algorithmic steps. This cost is computed on a CPU (Intel(R)) i7-11800H @ 2.30GHz. We omit the cost of unaligned mean as it does not require any signal alignment. The cost for computing the fully elastic mean averaged over 50 replications is in the first column. The costs associated with different steps of our method – partial alignment, computing PPD, and function estimation – are listed in the three right columns. A majority of time is consumed in partial alignment of functions for different values of λ.

	$\lambda = 0$	$\lambda > 0$		
	Align	Alig (1st Step)	Form PPD (1st Step)	Estimate (2nd Step)
Sim 1	15.29	1868.4	0.13	17.87
Sim 2	17.12	2021.6	0.12	18.92

Table 1: Computational time (in seconds) for various steps in our procedure.

4.2 Real Data Studies

In this section, we estimate signals underlying real datasets using our framework. However, in these cases, we lack ground truth to evaluate performance. Therefore, the results must be interpreted and compared to other estimators in order to assess their performance. One can also use plausibility of solutions as a merit criterion for comparing different estimators.

COVID-19 Data Analysis for Europe: There is a great interest in analyzing COVID-19 pandemic data and associated surges, the so-called waves, in infection rates. While different communities experienced COVID peaks at different times, some common patterns underlying COVID-19 incidences and outcomes emerged. For instance, if we focus on COVID-19 outcomes for different states or countries in a region, there is arguably an overall common pattern of waves during the pandemic period. We aim to discover such patterns using raw COVID-19 data from OWID [12].

- **Daily Hospitalization Rate Curves**: We consider the daily hospitalization counts associated with COVID-19 in 25 European countries during the period from April 2020 to July 2021 ([12]).

  The top row in Fig. 11 (a) and (b) display the original infection rate data and their fully-elastic alignments, respectively. These plots also show their cross-sectional means, $\hat{g}_\infty$ and $\hat{g}_0$, respectively. These means differ in the heights of peaks and valleys, and $\hat{g}_0$ also shows some small spurious peaks. PPD surface plot in (d) shows the gradual changes in $\hat{g}_\lambda$ for $\lambda \in [0, 0.6]$, and the bar chart in (c) selects $\lambda^* = 0.1$ with three persistent peaks (1, 3 and 4). The 95% pointwise confidence band (in gray) in (e) is quite broad, which can be attributed to a small sample size $n = 25$. Why is $\hat{g}$ a better estimator than other choices? When we consider individual curves in (a), the curves have well-defined peaks (and valleys) and mostly have three peaks. Our final estimate (in green) in panel (e) finds three well-defined peaks and strikes a good balance between the two extremes. More importantly, it better matches individual shapes than the other estimators and, thus, seems more plausible as a pattern common to the 25 countries.

- **Daily Infection Rate Curves**: The daily infection rate curves of COVID-19 can be used to study the rate of spread of a variant or the effectiveness of a vaccine. Next, we study the infection rate curves

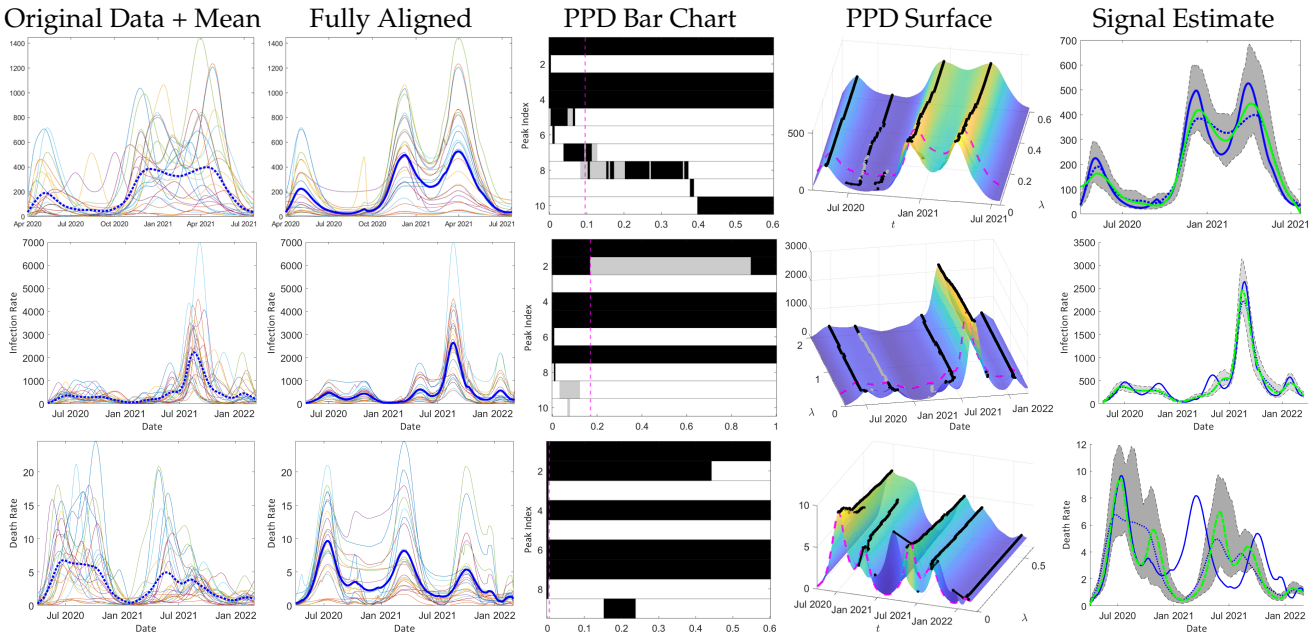


Figure 11: **Top row**: Daily hospitalization rate curves – PPD in suggests three peaks (1,3 and 4) are significant and persistent with $\lambda^* = 0.1$. **Top Middle row**: Daily infection rate curves – PPD in (labels four peaks (1, 4, 5, and 7) as significant. **Bottom row**: Daily death rate curves – PPD suggests that five peaks (1, 2, 4, 6, and 7) are significant. The final estimation (in green) differs from $\hat{g}_\infty$ and $\hat{g}_0$ in all three cases.
Figure Key: Dotted Blue - $\hat{g}_\infty$, Solid Blue - $\hat{g}_0$, Dashed Blue - $\hat{g}_{\lambda^*}$, Dashed Green: $\hat{g}$.

in the same 25 European countries from April 2020 to March 2022. The middle row of Fig. 11 shows the results: (a) and (b) show the original data and the fully-aligned signals, respectively. A simple averaging, $\hat{g}_\infty$, irons out most of the peaks in the first half of data. On the other hand, some of the peaks in $\hat{g}_0$ appear weak and need a closer look. Indeed, the PPD method rejects peak number 2 as insignificant and keeps the other four peaks (1, 4, 5, and 7). Plot (e) shows our final estimate (in green), and it shows a broad, shallow wave in the first half, and three significant and quick waves in the second half. The most damaging wave (peak 5) was during the outbreak of the delta variant in the Summer of 2021. It also discovers a prominent peak around January 2022, which is unclear from a quick inspection. The number of peaks in $\hat{g}$ is same as in $\hat{g}_{infty}$, but the heights of the peaks are stronger in $\hat{g}$.

- **Daily Death Rate Curves**: We also studied the daily death rates of COVID-19 in these 25 European countries from April 2020 to March 2022. The results are shown in the bottom row of Fig. 11. The standard estimates $\hat{g}_\infty$ (dotted blue) and $\hat{g}_0$ (solid blue) in (a) and (b) offer very different results: $\hat{g}_0$ has several sharp prominent peaks whereas $\hat{g}_\infty$ has broad and shallow peaks. A small peak in January 2021 in $\hat{g}_0$ seems artificial as no country has a peak at that time in the original data. The PPDs in (c)

and (d) estimate $\lambda^* = 0.025$, and the peaks labeled 3, 5, and 8 disappear immediately. Plot (e) shows that our estimate $\hat{g}$ (in green) differs significantly from the other estimates: $\hat{g}_\infty$ and $\hat{g}_0$. Although the partial alignment of functions is not too different from the original data, our estimate $\hat{g}$ picks up three distinct peaks in the early stage of COVID before January 2021.

Household Electricity Consumption Data: The objective here is to analyze the weekday electricity consumption data in domestic households [7]. This study focuses on two subsets of the data, corresponding to the months of January and July, consisting of 968 and 1, 219 households, respectively. In Fig. 12, the top row of (a) and (b) present the original and aligned curves of the January data. The PPD barchart reveals that five peaks (2, 6, 8, 9, and 10) are significant, with $\lambda^* = 0.235$. The upper panel (e) shows that there are vast differences among different estimators: $\hat{g}$ (in green), $\hat{g}_0$ (in solid blue), and $\hat{g}_\infty$ (in dotted blue). Our estimate$\hat{g}$ seems more plausible as a typical household electricity usage pattern due to its ups and downs nature. The full aligned mean $\hat{g}_0$ lumps the daily usage into a large peak during mid-day, and $\hat{g}_\infty$ oversimplifies the usage variations by ironing them out. The bottom row of Fig. 12 presents results for the July consumption data. The PPD barchart reveals two prominent peaks (2 and 5) with $\lambda^* = 0.18$, while $\hat{g}_0$ overestimates the peaks (it suggests 1, 2 and 3) and $\hat{g}_\infty$ underestimates the peaks (it suggests only one peak (2)). As shown in (e), $\hat{g}$ captures the important mid-afternoon peak, reflecting sustained air conditioning demand due to high temperatures, even when homes are unoccupied.

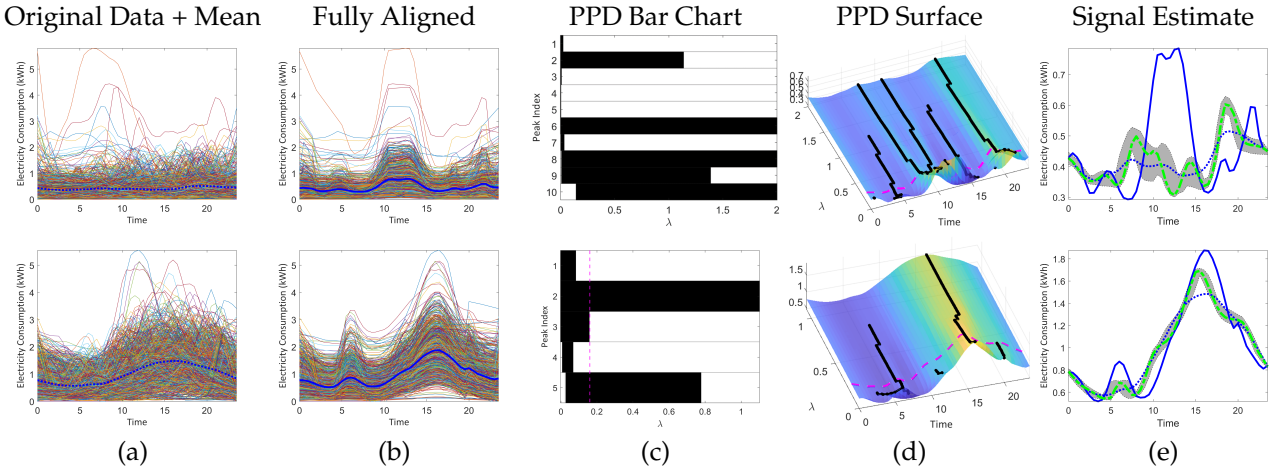
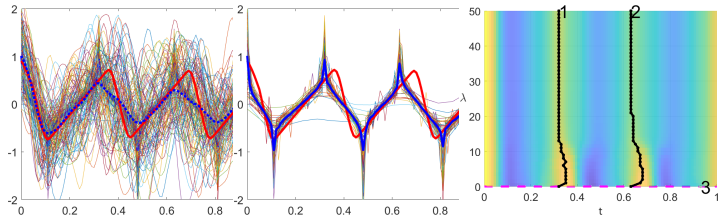


Figure 12: **Top row**: January electricity consumption curves – PPD in suggests five peaks (2,6,8,9 and 10) are significant and persistent with $\lambda^* = 0.235$. **Bottom row**: July electricity consumption curves – PPD suggests that two peaks (2 and 5) are significant. The final estimation (in green) differs from $\hat{g}_\infty$ and $\hat{g}_0$ in all three cases.
Figure Key: Dotted Blue - $\hat{g}_\infty$, Solid Blue - $\hat{g}_0$, Dashed Blue - $\hat{g}_{\lambda^*}$, Dashed Green: $\hat{g}$.

5 Discussion

The experimental results presented in this paper prove that our approach is successful in (1) estimating the number of peaks in noisy signals and (2) estimating the underlying unknown signal in a shape-constrained manner. A vital accomplishment is the automated tuning of parameter λ using PPDs. PPD, a topological feature of the data, is intuitive and effective in determining the number and locations of peaks in g, allowing for the selection of a reasonable value for the smoothing parameter, λ. Moreover, shape-constrained signal estimation further refines the initial estimate by maximizing the likelihood.

A pertinent question here is: Why not define the concept of PPDs in the original signal space? Why use the SRVF representation instead? We have provided the mathematical reasoning earlier. The $\mathbb{L}^2$ norm is invariant to the action of Γ in SRVF space but not in the original Hilbert space. Fig. 13 illustrates an example of the problem. It tries to form a PPD on one of the previously studied simulated data and concludes that $\kappa = 0$ is optimal. This result is unsatisfactory as the pinching effect is visible at $\kappa = 0$. We know that under the SRVF representation, the pinching effect is completely avoided, even for $\lambda = 0$.



(a) (b) (c)

Figure 13: (PPD in the original signal space): (a) Sawtooth dataset with the cross-sectional mean (in dotted blue) and ground truth g (in red). (b) Signal alignment using Eqn. 2 with mean (in blue) and optimal $\kappa^* = 0$. (c) The PPD plot in the original signal space, with $\kappa^* = 0$.

6 Conclusion

Estimating a mean signal g underlying a population of random observation is a classic signal processing problem. In many cases, the number, locations, or heights of extrema in g may be of direct interest themselves. In the presence of phase and additive noise, the classical unaligned mean $\hat{g}_\infty$ loses the geometric features, while fully elastic mean $\hat{g}_0$ can generate spurious features. This paper presents a geometric approach that explores the solution space by studying the geometry of $\hat{g}_\lambda$, for a range of smoothing parameter $\lambda \in [0, \infty)$. It introduces a novel topological tool *peak persistence diagram* (PPD) for investigating this solution space and focusing on dominant peaks. This persistence of peaks helps us to discard insignificant peaks, estimate the shape of g, and reach an optimal λ^*. The latter two quantities feed into a shape-constrained es-

timation of g. This estimation refines $\hat{g}_{\lambda^*}$ and produces an optimal estimate, $\hat{g}$, under penalized MLE. Most importantly, the heights, the locations, and the number of extrema in $\hat{g}$ are interpretable and supported by data. In contrast, $\hat{g}_\infty$ often blurs the peaks, and $\hat{g}_0$ overestimates heights and the number of extrema.

References

[1] S. Allassonniére, Y. Amit, and A. Trouvé. Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society (B)*, 69(1):3–29, 2007.

[2] D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 1995.

[3] P. Chaudhuri and J. S. Marron. Sizer for exploration of structures in curves. *Journal of the American Statistical Association*, 94(447):807–823, 1999.

[4] G. Claeskens, E. Devijver, and I. Gijbels. Nonlinear mixed effects modeling and warping for functional data using b-splines. *Electronic Journal of Statistics*, 15, 2021.

[5] S. Dasgupta, D. Pati, I. H. Jermyn, and A. Srivastava. Modality-constrained density estimation via deformable templates. *Technometrics*, 63(4):536–547, 2021.

[6] S. Dasgupta, D. Pati, and A. Srivastava. A two-step geometric framework for density modeling. *Statistica Sinica*, 30(4):2155–2177, 2020.

[7] S. Dasgupta, A. Srivastava, J. Cordova, and R. Arghandeh. Clustering household electrical load profiles using elastic shape analysis. In *2019 IEEE Milan PowerTech*, pages 1–6, 2019.

[8] J. Fan and I. Gijbels. *Local polynomial modelling and its applications*. Chapman & Hall/CRC, 1996.

[9] W. A. Gardner. Statistically inferred time warping: extending the cyclostationarity paradigm from regular to irregular statistical cyclicity in scientific data. *EURASIP Journal on Advances in Signal Processing*, 2018(1):59, 2018. Article 59.

[10] S. Kurtek. A geometric approach to pairwise Bayesian alignment of functional data using importance sampling. *Electronic Journal of Statistics*, 11(1):502 – 531, 2017.

[11] S. Kurtek, A. Srivastava, and W. Wu. Signal estimation under random time-warpings and nonlinear signal alignment. In *Advances in Neural Information Processing Systems*, volume 24, 2011.

[12] E. Mathieu, H. Ritchie, L. Rodés-Guirao, C. Appel, C. Giattino, J. Hasell, B. Macdonald, S. Dattani, D. Beltekian, E. Ortiz-Ospina, and M. Roser. Coronavirus-19 pandemic. *Our World in Data*, 2020.

[13] M. Molina, L. J. Tardón, A. M. Barbancho, I. De-Torres, and I. Barbancho. Enhanced average for event-related potential analysis using dynamic time warping. *Biomed. Sig. Proc. and Control*, 87, 2024.

[14] A. Napolitano. Time-warped almost-cyclostationary signals: Characterization and statistical function measurements. *IEEE Transactions on Signal Processing*, 65(20):5526–5541, 2017.

[15] U. Ozertem and D. Erdogmus. Signal denoising using principal curves: Application to timewarping. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3709–3712, 2008.

[16] A.-G. Pielmus, M. Klum, T. Tigges, R. Orglmeister, and . Urban. Progressive dynamic time warping for noninvasive blood pressure estimation. *Current Directions in Biomed. Engg.*, 6(3):579–582, 2020.

[17] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer series in statistics. Springer, 1997.

[18] M. Roser, H. Ritchie, E. Ortiz-Ospina, and Lucas Rodés-Guirao. World population growth. *Our World in Data*, 2013.

[19] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, 1978.

[20] I. Simpson, J. A. Schnabel, A. R. Groves, J. Andersson, and M. W. Woolrich. Probabilistic inference of regularisation in non-rigid registration. *NeuroImage*, 59(3):2438–2451, 2012.

[21] H. Skutkova, M. Vitek, P. Babula, R. Kizek, and I. Provaznik. Classification of genomic signals using dynamic time warping. *BMC Bioinformatics*, 14(10):S1, 2013. S1.

[22] X. Song, Q. Wen, Y. Li, and Li. Sun. Robust time series dissimilarity measure for outlier detection and periodicity detection. In *31st ACM Conf. on Info. & Know. Management*, pages 4510–4514, 2022.

[23] A. Srivastava and E.P. Klassen. *Functional and Shape Data Analysis*. Springer Series in Statistics. Springer New York, 2016.

[24] S. Yang and A. C. Singer. Hb-dtw: Hyperdimensional bayesian dynamic time warping for non-uniform doppler. In *2021 29th European Signal Processing Conference (EUSIPCO)*, pages 2020–2024, 2021.

[25] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:249–274, 02 2005.