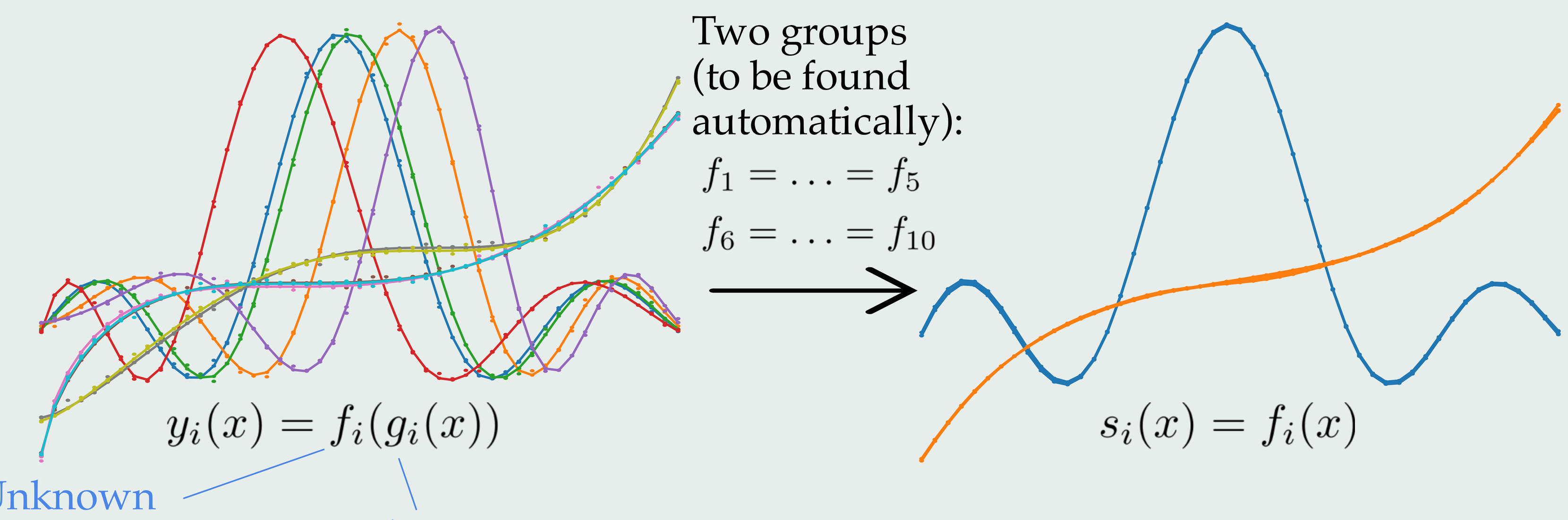
Gaussian Process Latent Variable Alignment Learning

Ieva Kazlauskaite[1], Carl Henrik Ek[2], Neill D. F. Campbell[1]
[1]University of Bath, UK [2]University of Bristol, UK

1. Introduction

Problem: aligning temporally warped noisy sequences.



Challenges:

- Unknown number of distinct latent functions f_j (or equivalently, unknown number of groups of sequences).
- Weak assumptions on the warps (smoothness, monotonicity) without a parametric description.
- Sequences of different lengths.

2. Overview

There are two parts to our model.

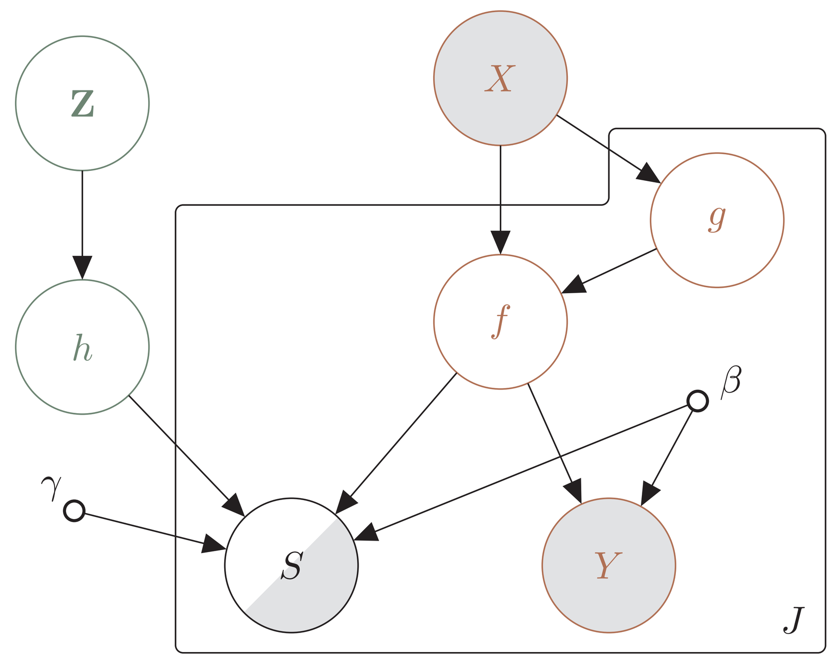
- We place GP priors on f_j and g_j.
- Fit these GPs on the observed sequences y_j.

} Model over time

This describes each observed sequence in isolation without aligning them. Moreover, fitting both f_j and g_j to a single sequence y_j is an ill-posed problem. To address this, we:

- Evaluate estimated f_i at fixed inputs x: $S_j = f_j(x) + \epsilon_j$.
- Impose a constraint encouraging $\{S_j\}$ to split into a small number of clusters.
- Define this constraint using GP-LVM.

} Alignment model



Goal: find the aligned sequences $\{S_j\}$ which have high likelihood under both parts of the model.

3. Model

Inputs:
$Y_j \in \mathbb{R}^N$ - observed sequences
$X \in \mathbb{R}^N$ - observed (fixed) uniform sampling of time

Warpings:
The time warpings g_j need to be monotonic, which we ensure by parametrising them using auxiliary variables $U_j \in \mathbb{R}^N$ such that $[G_j]_n := 2\sum_{k=1}^{n}[\mathrm{softmax}(U_j)]_k - 1$.

Model over time:

Warps evaluated at fixed inputs $(g_j(X))$ — Fixed inputs — Kernel hyperparams

Prior over warps: $p(G_j \mid X, \omega) \sim \mathcal{N}(0, k_{\omega_j}(X, X))$

Prior over sequences: $p\left(\begin{bmatrix} F_j^{X} \\ F_j^{G} \end{bmatrix} \mid G_j, X_j, \theta_j\right) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} k_{\theta_j}(X,X) & k_{\theta_j}(X,G_j) \\ k_{\theta_j}(G_j,X) & k_{\theta_j}(G_j,G_j) \end{bmatrix}\right)$

Latent function GP evaluated at fixed inputs $f_j(X)$ — Latent function GP evaluated at warped inputs $f_j(G_j)$

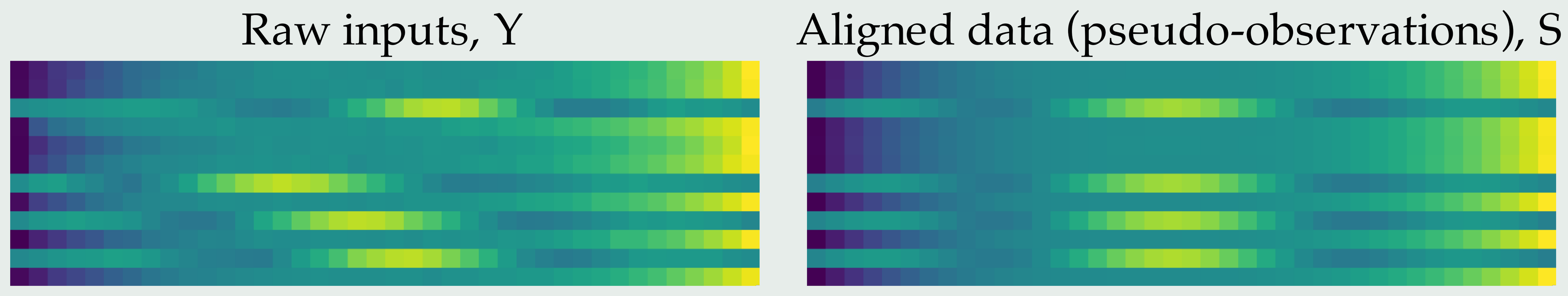
Alignment model:
We use a GP-LVM that places independent GPs over the data features.

Likelihoods:
We treat $\mathbf{S}$ as if they are observed (see Limitations) calling them *pseudo-observations* with the likelihood defined as an equal mixture:

Latent functions evaluated at fixed inputs — GP-LVM latent space — Vector of n-th samples of aligned sequences — j-th aligned sequence

$$p(\mathbf{S} \mid \mathbf{H}, \mathbf{F}^X, \mathbf{Z}) = \frac{1}{2}\left(\prod_n \mathcal{N}(S_{:,n} \mid \mathbf{H}_n, \gamma^{-1} I_J) + \prod_j \mathcal{N}(S_{j,:} \mid \mathbf{F}_j^X, \beta_j^{-1} I_N)\right)$$

Aligned sequences — Noiseless GP-LVM evaluations — Factorisation over columns of S — Factorisation over rows of S



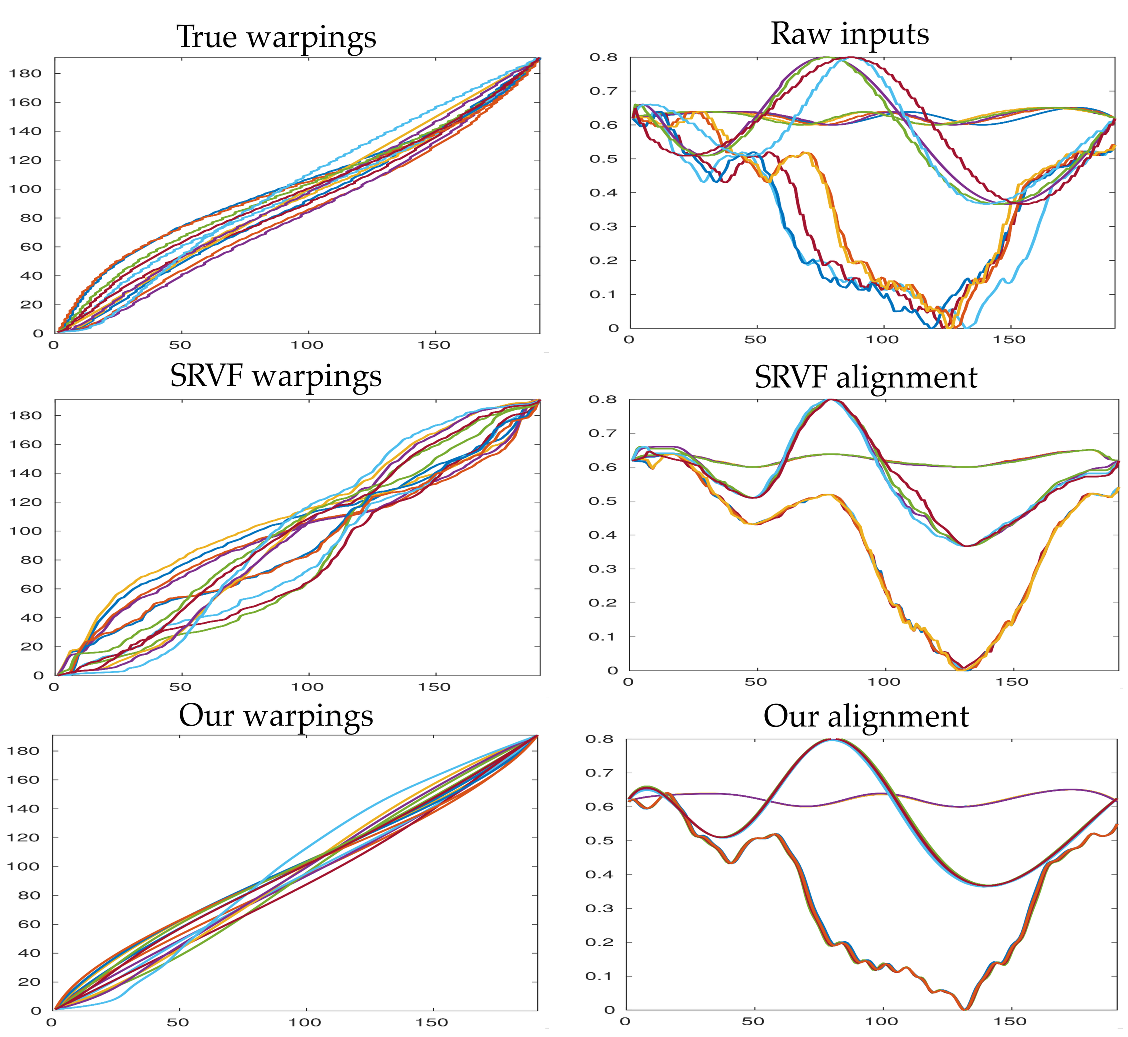
4. Inference

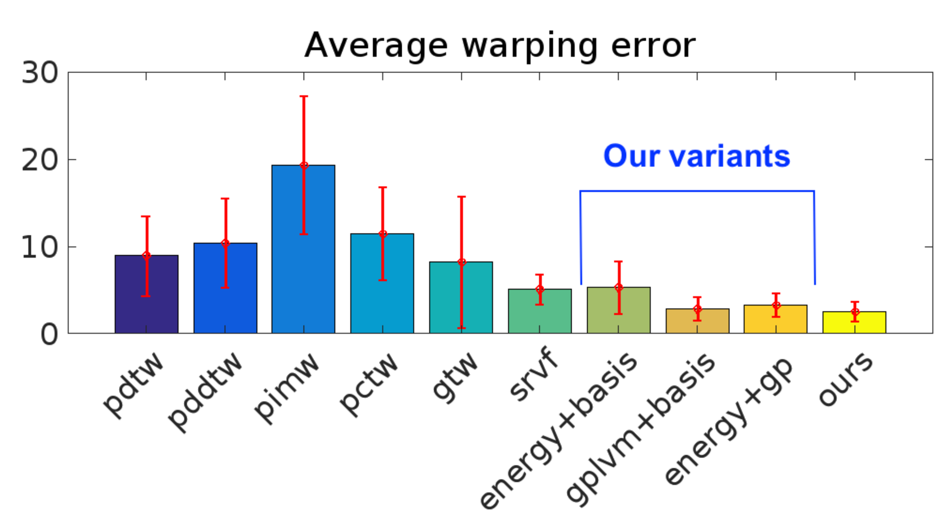
We optimise marginal log-likelihood:

$$\log p(\mathbf{S}, \mathbf{Y}, \mathbf{Z}, G, \mid X) = \log p(\mathbf{S}, \mathbf{Y} \mid \mathbf{G}, X) + \log p(\mathbf{S} \mid \mathbf{Z}) + \log p(\mathbf{Z}) + \log p(\mathbf{G} \mid X)$$

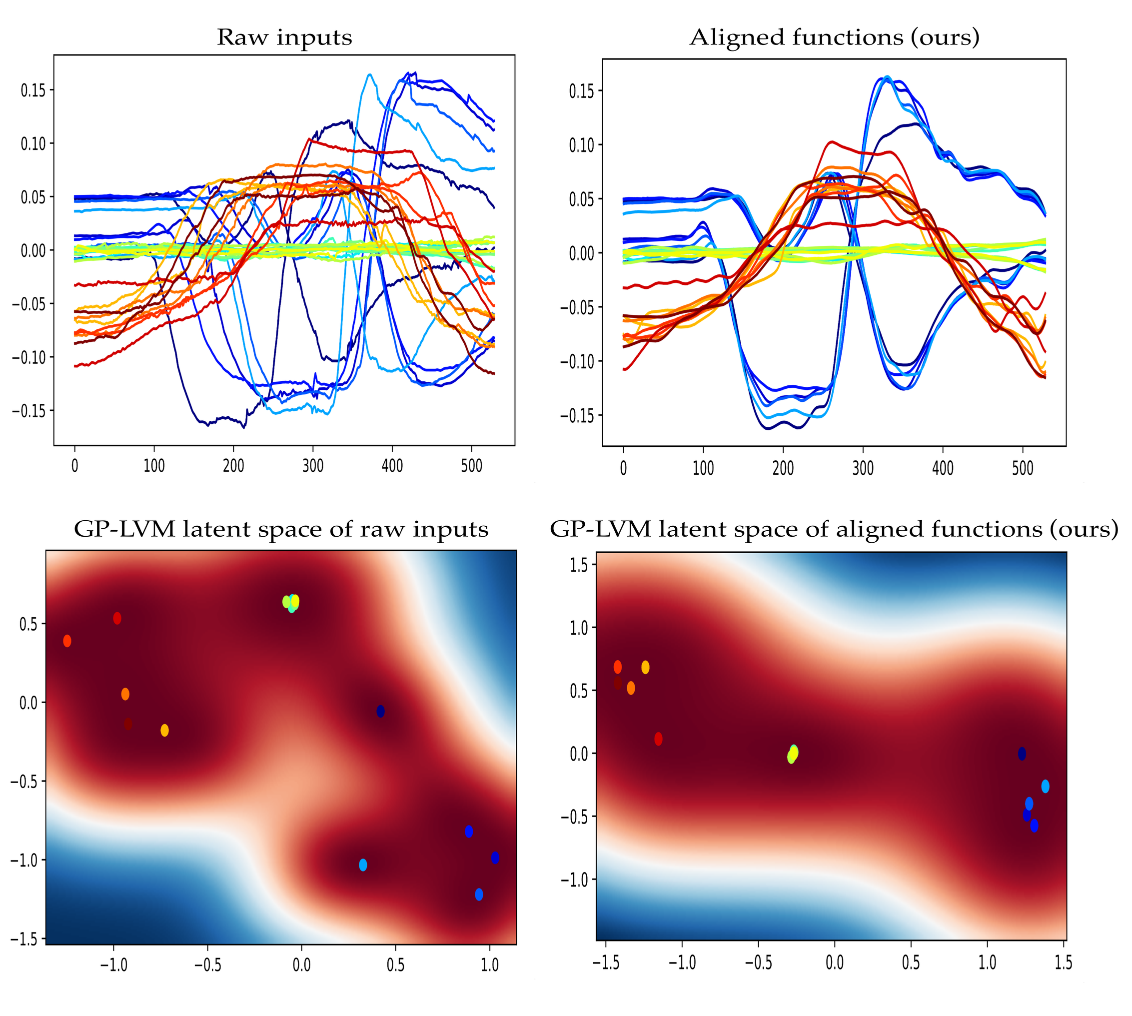
w.r.t. the pseudo observations $\mathbf{S}$, the latent variables $\mathbf{Z}$, warps auxiliary variables U_j, and the kernel hyperparameters.

5. Comparisons [3, 4]

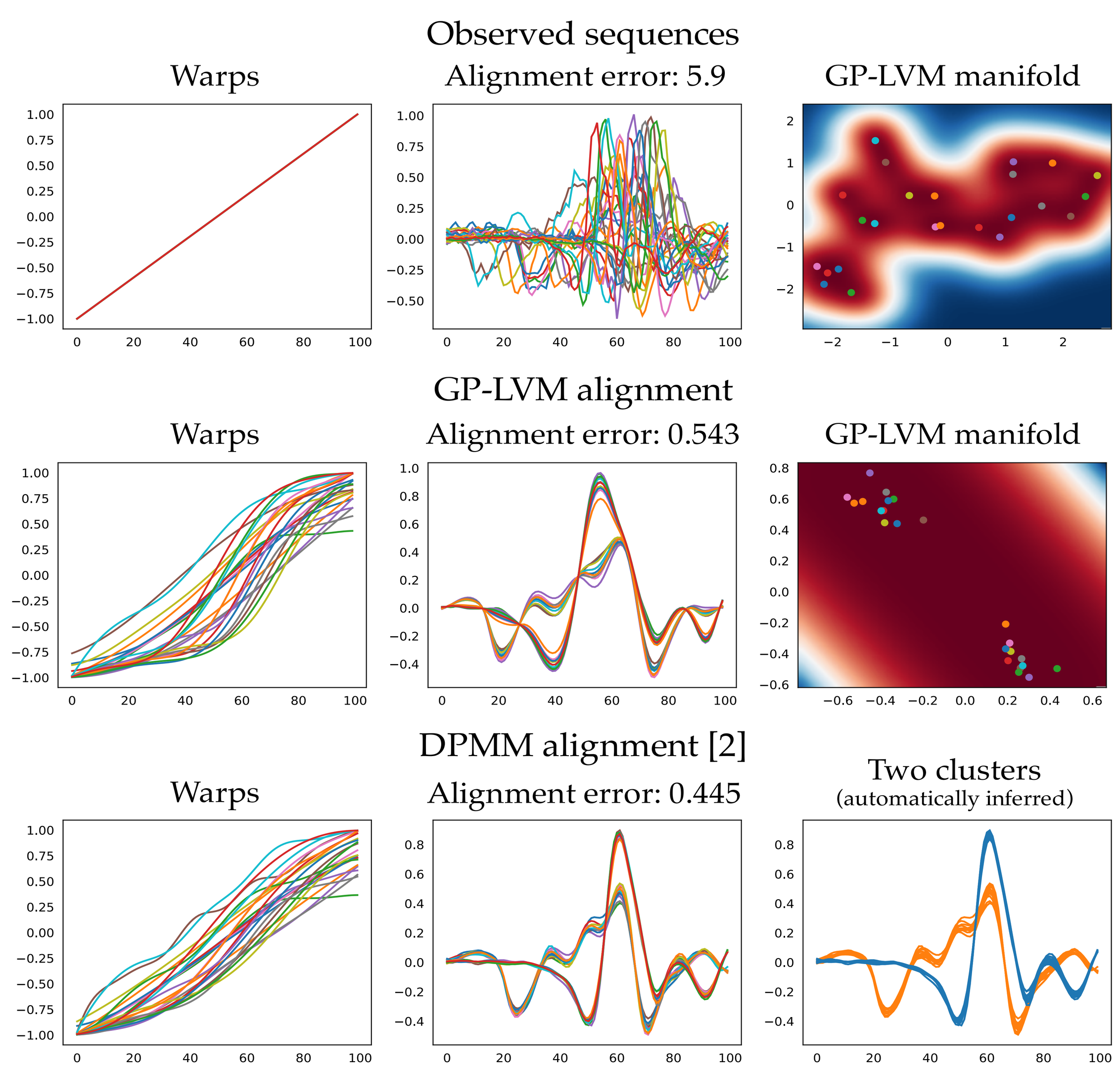




6. CMU Motion capture data



7. Heartbeats data [1]



8. Limitations

- $\mathbf{S}$ needs to be observed for the two parts of the model to be conditionally dependent, thus, we directly optimise $\mathbf{S}$ obtaining a point estimate of the aligned sequences. How to make the model fully generative?
- Scalability (beyond sparse GPs)

References

[1] P. Bentley, G. Nordehn, M. Coimbra, S. Mannor. The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011).
[2] I. Kazlauskaite, I. Ustyuzhaninov, C.H. Ek, N.D.F. Campbell. Sequence Alignment with Dirichlet Process Mixtures. BNP@NeurIPS, 2018.
[3] S. Kurtek, A. Srivastava, E. Klassen, Z. Ding. Statistical Modeling of Curves using Shapes and Related Features. Journal of the American Statistical Association, 2012.
[4] F. Zhou, F. de al Torre. Generalized Canonical Time Warping. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2016.