Aligned Multi-task Gaussian Process

Olga Mikheeva [1], Ieva Kazlauskaite [2], Adam Hartshorne [3], Hedvig Kjellström [1],
Carl Henrik Ek [2], Neill D. F. Campbell [3]

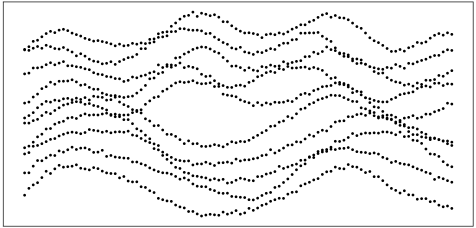
[1] KTH Royal Institute of Technology, [2] University of Cambridge, [3] University of Bath

Motivation: Multi-task learning on misaligned data

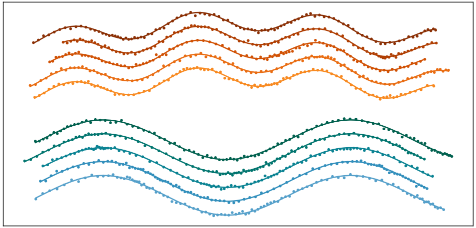
Why? Misalignment hinders learning correlations between tasks.

Downsides of existing models:

- cannot model flexible misalignment
- use a-priori known correlation between tasks
- either probabilistic or monotonic alignment, but not both



Observed data



Aligned data

Modelling monotonic functions

Misaligned := Input (time or space) is warped with a monotonic function.

Possible model: Monotonic GP flow [2]. SDE-based. Prohibitively expensive.

Our proposal: ODE-based Monotonic GP flow.

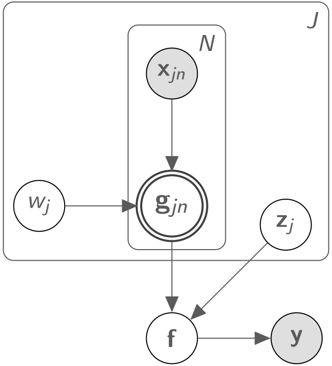
$$\boxed{g(x) := u(\tau = T; x) = \int_0^T w\big(u(\tau)\big)\, \mathrm{d}\tau}$$

$$\begin{aligned} \text{ODE:} \quad & \mathrm{d}u = w(u)\,\mathrm{d}\tau, \\ \text{Uncertain drift function:} \quad & w(u) \sim \mathcal{GP}\big(\mathbf{0}, K_\omega(u,u)\big) \end{aligned}$$

Solution g(x) of the ODE is monotonic as a function of initial condition $u(\tau = 0) := x$.

Aligned Multi-task Gaussian Process

Our model: Fully Bayesian multi-task learning
for misaligned data

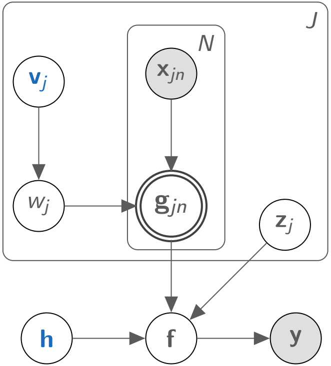


Latent corr.	$\mathbf{z}_j \sim \mathcal{N}(\mathbf{z}_j \mid \mathbf{0}, \mathbf{I}_Q)$
ODE Drift	$w_j \sim \mathcal{GP}\big(w_j \mid \mathbf{0}, K_\omega(u_j, u_j)\big)$
Warp	$\mathbf{g}_j \mid \mathbf{x}_j, w_j \sim \text{Monotonic Process}\big(\mathbf{g}_j \mid \mathbf{x}_j, w_j\big)$
Function	$\mathbf{f} \mid \mathbf{z}, \mathbf{g} \sim \mathcal{GP}\big(\mathbf{f} \mid \mathbf{0}, K_\psi(\mathbf{z}_j, \mathbf{z}_{j'}) \odot K_\theta(\mathbf{g}_{j,n}, \mathbf{g}_{j',n'})\big)$
Noisy data	$\mathbf{y} \mid \mathbf{f} \sim \mathcal{N}(\mathbf{y} \mid \mathbf{f}, \beta^{-1}\mathbf{I}_{JN})$

Joint prob.: $p(\mathbf{y}, \mathbf{f}, \mathbf{z}, \mathbf{g}, w \mid \mathbf{X}) = p(\mathbf{f} \mid \mathbf{z}, \mathbf{g}) \prod_{j=1}^{J} p(\mathbf{g}_j \mid \mathbf{x}_j, w_j)\, p(w_j)\, p(\mathbf{z}_j) \prod_{n=1}^{N} p(y_{jn} \mid f_{jn})$

Sparse Stochastic Variational Inference

- Sparse GP posteriors (inducing variables $\mathbf{v}_j$ and $\mathbf{h}$)
- ELBO is factorized over tasks j
- Efficient sampling of monotonic functions with pathwise approach [3]

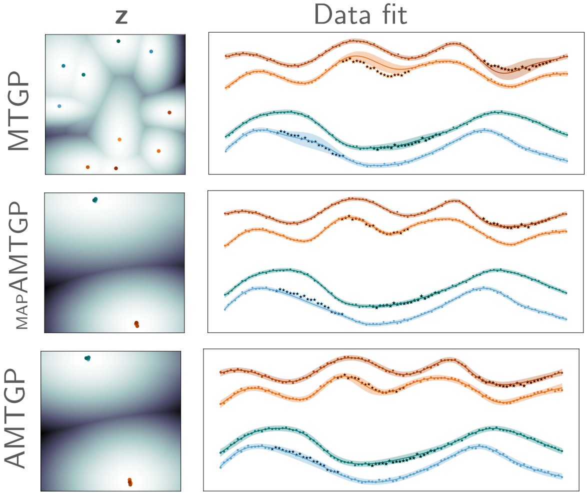


Experiments

Synthetic data: 2 functions, 5 warps each

Evaluation scenario: missing segments of data at random locations

	Predictive quality: Mean	Likelihood
MTGP	✗	✓
GP-LVA [1]	✗	✗
MAPAMTGP	✓	✗
AMTGP	✓	✓



*See our paper for more experiments.

Summary

Contributions:

- a fully Bayesian GP model for multi-task learning on misaligned data
- an efficient inference scheme based on sparse SVI
- a ODE reformulation of monotonic GP flow [2] with efficient training

Thank you!

Github: https://github.com/OlgaMikheeva/aligned_mtgp E-mail: olgamik@kth.se

[1] Ieva Kazlauskaite, Carl Henrik Ek, and Neill Campbell. Gaussian process latent variable alignment learning. In *The International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR, 2019.

[2] Ivan Ustyuzhaninov, Ieva Kazlauskaite, Carl Henrik Ek, and Neill Campbell. Monotonic Gaussian process flows. In *The International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

[3] James T. Wilson, Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, and Marc P. Deisenroth. Efficiently sampling functions from gaussian process posteriors. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2020.