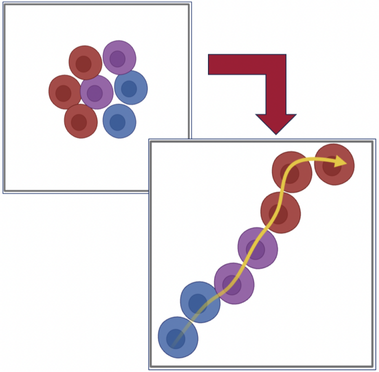
DGP-LVM: Derivative Gaussian process latent variable models

Soham Mukherjee

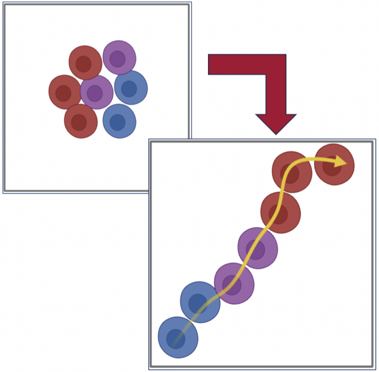
TU Dortmund, University of Tübingen

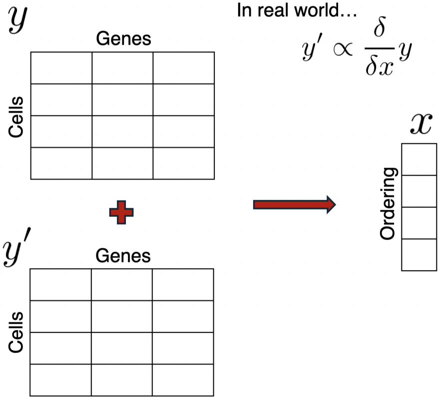
2nd September 2025

Motivation

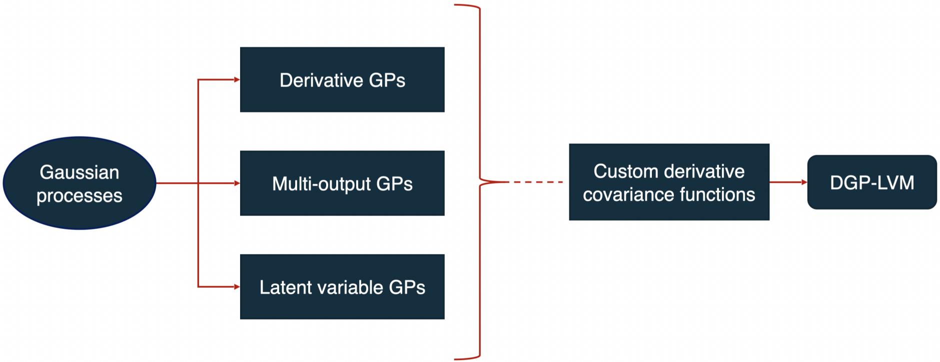


Motivation




y
Genes
Cells
y'
Genes
Cells
In real world...
$y' \propto \frac{\delta}{\delta x} y$
x
Ordering

Model structure



Derivative Gaussian process

For an input variable x, a GP $f(x) \sim \mathcal{GP}(m_f, K)$ and the derivative $f'(x) \sim \mathcal{GP}(m_{f'}, K'')$ is jointly specified as

$$\begin{pmatrix} f(x) \\ f'(x) \end{pmatrix} \sim \mathcal{GP}\left(\begin{pmatrix} m_f \\ m_{f'} \end{pmatrix}, \begin{pmatrix} K & K' \\ K'^T & K'' \end{pmatrix}\right)$$

Derivative covariance function

For a Squared Exponential covariance function with length-scale parameter ρ, marginal SDs α and α' for f and f' respectively,

$$K(x_i, x_j) = \alpha^2 \exp\left(-\frac{(x_i - x_j)^2}{2\rho^2}\right),$$

$$K'(x_i, x_j) = \alpha\alpha' \frac{(x_i - x_j)}{\rho^2} \exp\left(-\frac{(x_i - x_j)^2}{2\rho^2}\right),$$

$$K''(x_i, x_j) = \frac{\alpha'^2}{\rho^4}(\rho^2 - (x_i - x_j)^2) \exp\left(-\frac{(x_i - x_j)^2}{2\rho^2}\right).$$

Multi-output GPs

We model $D > 1$ output dimensions using dimension-specific individual GPs

$$\begin{pmatrix} f_d(x) \\ f'_d(x) \end{pmatrix} \sim \mathcal{GP}\left(\begin{pmatrix} m_{f_d} \\ m_{f'_d} \end{pmatrix}, \begin{pmatrix} K_d & K'_d \\ K'^{T}_d & K''_d \end{pmatrix} \right)$$

Multi-output GPs

We model $D > 1$ output dimensions using dimension-specific individual GPs

$$\begin{pmatrix} f_d(x) \\ f'_d(x) \end{pmatrix} \sim \mathcal{GP}\left(\begin{pmatrix} m_{f_d} \\ m_{f'_d} \end{pmatrix}, \begin{pmatrix} K_d & K'_d \\ K'^T_d & K''_d \end{pmatrix} \right)$$

Then we combine them using a between-output correlation $C = LL^T$ such that

$$\begin{pmatrix} \tilde{f}_1(x) \\ \dots \\ \tilde{f}_D(x) \end{pmatrix} = L \times \begin{pmatrix} f_1(x) \\ \dots \\ f_D(x) \end{pmatrix},$$

Latent variable input and likelihood

We assume a prior-like distribution for our latent variable x from an observed quantity $\tilde{x}$ such that

$$\tilde{x}_i \sim \mathcal{N}(x_i, s^2)$$

Latent variable input and likelihood

We assume a prior-like distribution for our latent variable x from an observed quantity $\tilde{x}$ such that

$$\tilde{x}_i \sim \mathcal{N}(x_i, s^2)$$

Finally, we specify our outputs y and their derivatives y' as

$$y_{di} \sim \mathcal{N}(\tilde{f}_d(x_i), \sigma_d^2),$$
$$y'_{di} \sim \mathcal{N}(\tilde{f}'_d(x_i), {\sigma'_d}^2).$$

Inference

We assume independent priors on model parameters θ_d such that

$$\theta_d \sim p(\theta_d) = p(\rho_d)\, p(\alpha_d)\, p(\alpha'_d)\, p(\sigma_d)\, p(\sigma'_d).$$

Inference

We assume independent priors on model parameters θ_d such that

$$\theta_d \sim p(\theta_d) = p(\rho_d)\, p(\alpha_d)\, p(\alpha'_d)\, p(\sigma_d)\, p(\sigma'_d).$$

The multi-output joint probability density thus factorizes as

$$p(y, y', x, \theta) = \prod_{d}^{D} p(y_d \mid x, \theta_d)\, p(y'_d \mid x, \theta_d)\, p(x)\, p(\theta_d).$$

Inference

We assume independent priors on model parameters θ_d such that

$$\theta_d \sim p(\theta_d) = p(\rho_d)\, p(\alpha_d)\, p(\alpha'_d)\, p(\sigma_d)\, p(\sigma'_d).$$

The multi-output joint probability density thus factorizes as

$$p(y, y', x, \theta) = \prod_{d}^{D} p(y_d \mid x, \theta_d)\, p(y'_d \mid x, \theta_d)\, p(x)\, p(\theta_d).$$

Using Bayes' rule, we obtain the joint posterior over x and θ as

$$p(x, \theta \mid y, y') = \frac{p(y, y', x, \theta)}{\int \int p(y, y', x, \theta)\, dx\, d\theta}.$$

Simulation study

We consider two different data generating processes.

- The derivative GP data scenario,
- A derivative periodic data scenario.

Simulation study

We consider two different data generating processes.

- The derivative GP data scenario,
- A derivative periodic data scenario.

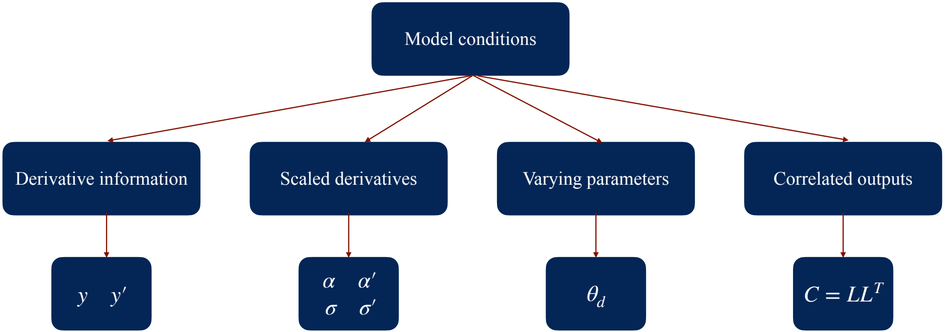
In the periodic data scenario, we generate data from

$$f_{id} = \alpha_d \sin\left(\frac{x_i}{\rho_d}\right)$$

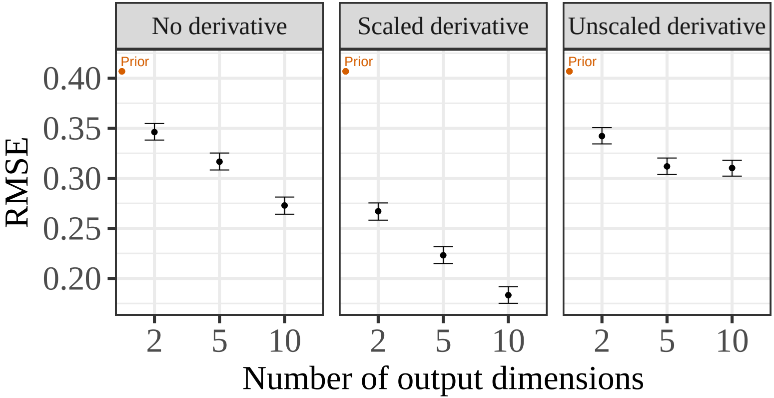
$$f'_{id} = \frac{\alpha'_d}{\rho_d} \cos\left(\frac{x_i}{\rho_d}\right)$$

* These are only two of the five simulation studies presented in the paper.

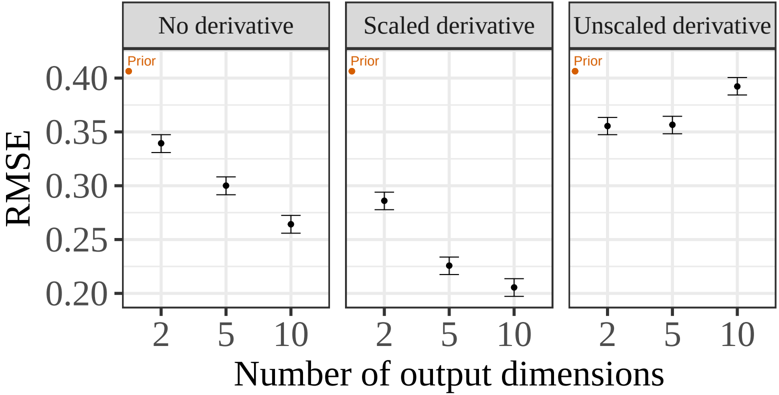
Model setup


Model conditions
Derivative information
Scaled derivatives
Varying parameters
Correlated outputs
$y \quad y'$
$\alpha \quad \alpha'$
$\sigma \quad \sigma'$
θ_d
$C = LL^T$

Results (GP scenario)

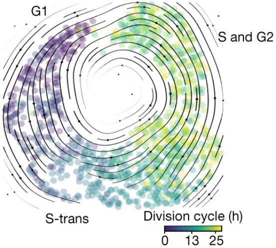


Results (Periodic scenario)



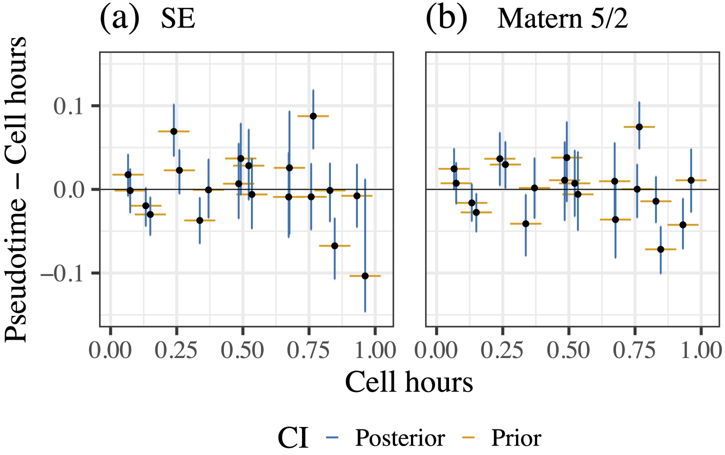
Case study

- We demonstrate our method on a reduced Cell Cycle RNA expression data.
- DGP-LVM with SE and Matern 5/2 are used.



Ref: Mahdessian, D., Cesnik, A.J., Gnann, C. et al. Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. Nature 590, 649–654 (2021). https://doi.org/10.1038/s41586-021-03232-9

Case study


(a) SE
(b) Matern 5/2
Pseudotime − Cell hours
0.1
0.0
−0.1
0.00
0.25
0.50
0.75
1.00
Cell hours
CI
Posterior
Prior

Summary

- We introduce latent variable GPs to jointly model multi-dimensional data with scaled derivatives.
- Increased accuracy and high quality posterior estimates of latent variables.
- Data generating process that resembles the complexities of real single-cell RNA expression data.
- Limitation: This is an exact model which has high computational complexity.

Full paper: Mukherjee, S., Claassen, M. & Bürkner, PC. DGP-LVM: Derivative Gaussian process latent variable models. Stat Comput 35, 120 (2025). https://doi.org/10.1007/s11222-025-10644-4
Preprint: https://doi.org/10.48550/arXiv.2404.04074

Outlook

- We overcome the scalability limitations in our following works.
- We generalise Hilbert space approximation methods for a scalable multi-output latent variable GPs (`https://doi.org/10.48550/arXiv.2505.16919`).
- Our method provides superior quality of latent variable estimates as compared to other GP approximation methods.
- We are currently working on a practical approximation for DGP-LVM.

Collaborators

My supervisors and co-authors of this paper:

- Paul-Christian Bürkner
- Manfred Claassen

You can take a look at my other works here:
https://soham6298.github.io/
Feel free to reach out for collaborations!