Machine Learning with Signal Processing

Part IV: Application Examples

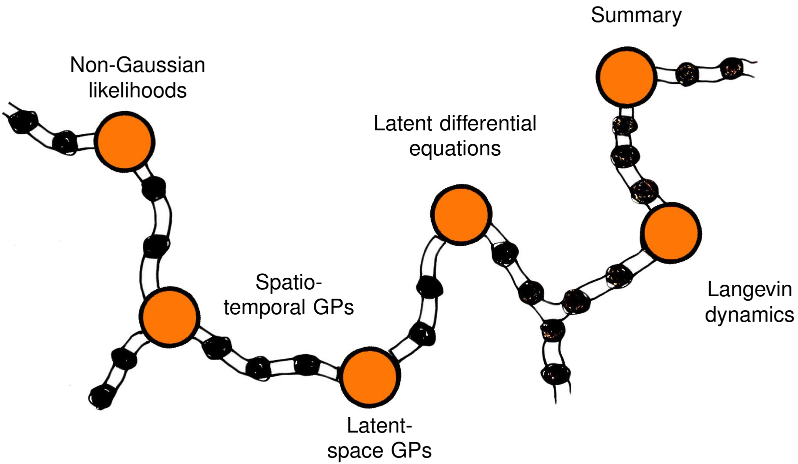
Arno Solin
Assistant Professor in Machine Learning
Department of Computer Science
Aalto University

ICML 2020 Tutorial

@arnosolin

arno.solin.fi

Outline



Summary
Non-Gaussian likelihoods
Latent differential equations
Spatio-temporal GPs
Langevin dynamics
Latent-space GPs

Gaussian processes ❤ SDEs

GPs under the kernel formalism

$$f(t) \sim \mathrm{GP}(0, \kappa(t, t'))$$

$$\mathbf{y} \mid \mathbf{f} \sim \prod_i p(y_i \mid f(t_i))$$

Flexible model specification

Inference / First-principles

Stochastic differential equations

$$\mathrm{d}\mathbf{f}(t) = \mathbf{F}\,\mathbf{f}(t) + \mathbf{L}\,\mathrm{d}\boldsymbol{\beta}(t)$$

$$y_i \sim p(y_i \mid \mathbf{h}^\mathsf{T}\mathbf{f}(t_i))$$

Non-Gaussian likelihoods

- The observation model might not be Gaussian

$$f(t) \sim \mathrm{GP}(0, \kappa(t, t'))$$
$$\mathbf{y} \mid \mathbf{f} \sim \prod_i p(y_i \mid f(t_i))$$

- There exists a multitude of great methods to tackle general likelihoods with approximations of the form

$$\mathbb{Q}(\mathbf{f} \mid \mathcal{D}) = \mathrm{N}(\mathbf{f} \mid \mathbf{m} + \mathbf{K}\boldsymbol{\alpha}, (\mathbf{K}^{-1} + \mathbf{W})^{-1})$$

- Use those methods, but deal with the latent using state space models

Inference

- Laplace approximation
- Variational Bayes
- Direct KL minimization
- EP or Assumed density filtering (Single-sweep EP)
- Can be evaluated in terms of a (Kalman) filter forward and backward pass, or by iterating them

Example

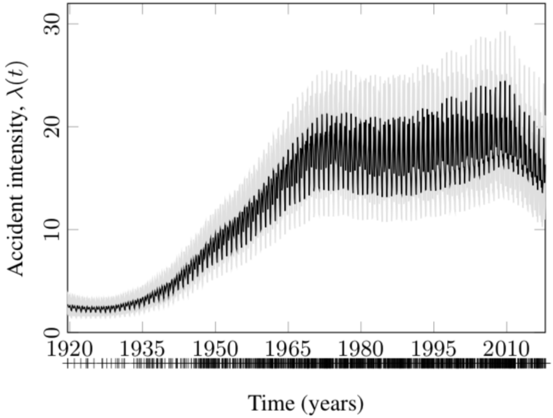
- Commercial aircraft accidents 1919–2017
- Log-Gaussian Cox process (Poisson likelihood) by ADF/EP
- Daily binning, $n = 35{,}959$
- GP prior with a covariance function:

$$\kappa(t,t') = \kappa_{\text{Mat.}}^{\nu=3/2}(t,t') + \kappa_{\text{Per.}}^{\text{year}}(t,t')\,\kappa_{\text{Mat.}}^{\nu=3/2}(t,t') + \kappa_{\text{Per.}}^{\text{week}}(t,t')\,\kappa_{\text{Mat.}}^{\nu=3/2}(t,t')$$

- Learn hyperparameters by optimizing the marginal likelihood

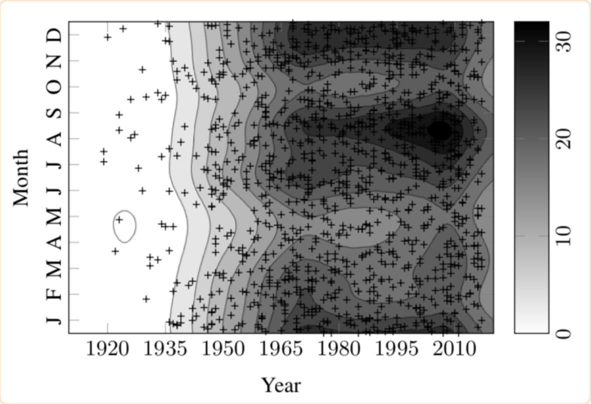
Nickisch *et al.* (2018). ICML.

Example



Nickisch *et al.* (2018). ICML.

Example



Nickisch *et al.* (2018). ICML.

Spatio-temporal GPs

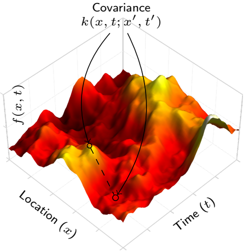
$$f(\mathbf{r}) \sim \mathrm{GP}(0, \kappa(\mathbf{r}, \mathbf{r}'))$$
$$\mathbf{y} \mid \mathbf{f} \sim \prod_i p(y_i \mid f(\mathbf{r}_i))$$

$$f(\mathbf{x}, t) \sim \mathrm{GP}(0, \kappa(\mathbf{x}, t; \mathbf{x}', t'))$$
$$\mathbf{y} \mid \mathbf{f} \sim \prod_i p(y_i \mid f(\mathbf{x}_i, t_i))$$

Spatio-temporal Gaussian processes

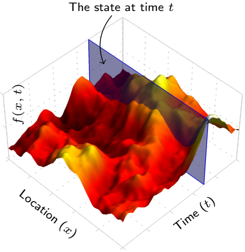
GPs under the kernel formalism

$$f(\mathbf{x}, t) \sim \mathrm{GP}(0, k(\mathbf{x}, t; \mathbf{x}', t'))$$
$$y_i = f(\mathbf{x}_i, t_i) + \varepsilon_i$$

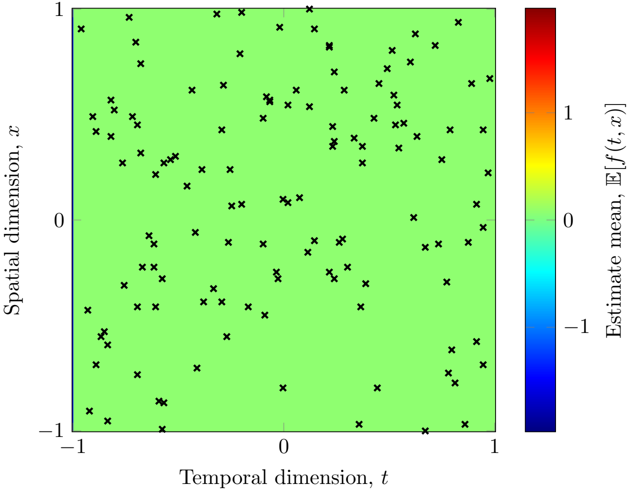


Stochastic partial differential equations

$$\frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial t} = \mathcal{F}\,\mathbf{f}(\mathbf{x}, t) + \mathcal{L}\,w(\mathbf{x}, t)$$
$$y_i = \mathcal{H}_i\,\mathbf{f}(\mathbf{x}, t) + \varepsilon_i$$

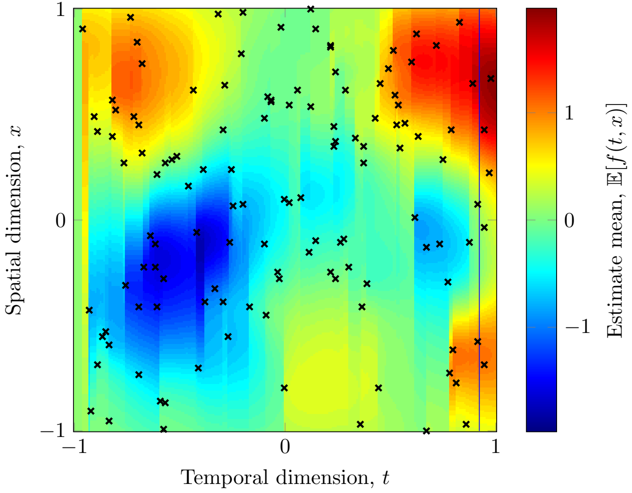


Spatio-temporal GP regression



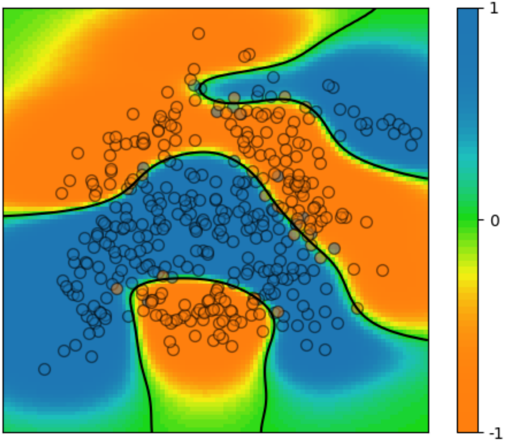
Särkkä *et al.* (2013). IEEE Signal Processing Magazine.

Spatio-temporal GP regression



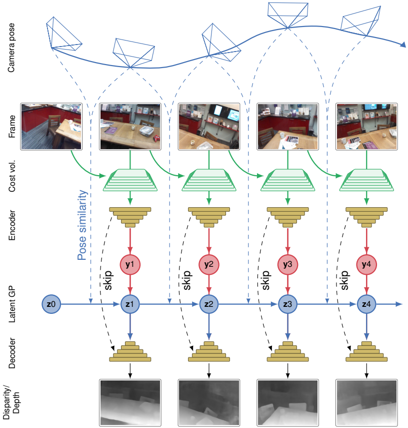
Särkkä *et al.* (2013). IEEE Signal Processing Magazine.

Spatio-temporal GP classification



Wilkinson *et al.* (2020). ICML.

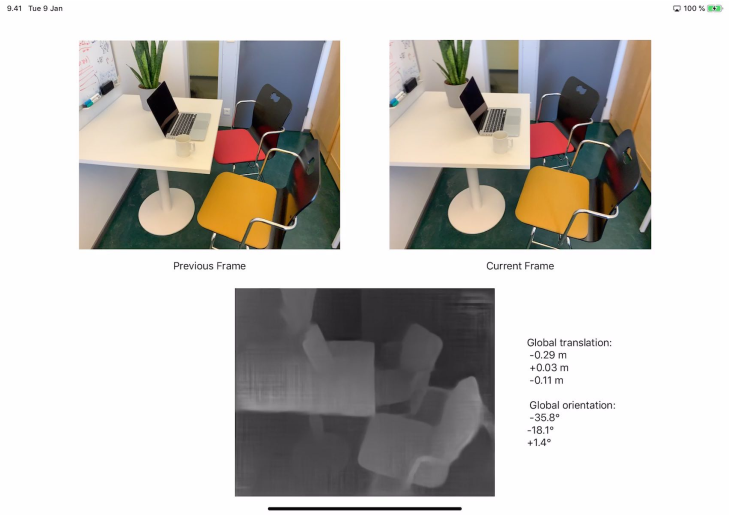
Multi-view stereo as a temporal fusion problem



- Inputs: Frame pairs and relative camera poses
- Encoder–decoder network for depth estimation
- Treat the encoder as a feature extractor, and do GP regression in the latent space
- The GP prior encodes the similarity of the camera views

Hou *et al.* (2019). ICCV.

Multi-view stereo as a temporal fusion problem



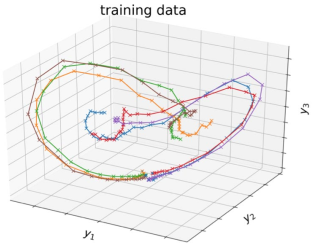
Hou *et al.* (2019). ICCV. Video: `https://youtu.be/iellGrlNW7k`

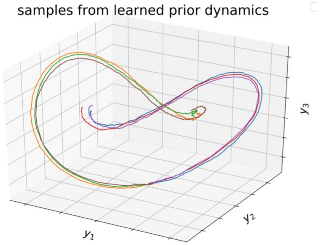
Sequential priors in GAN latent space

Hou *et al.* Submitted.

Latent differential equations

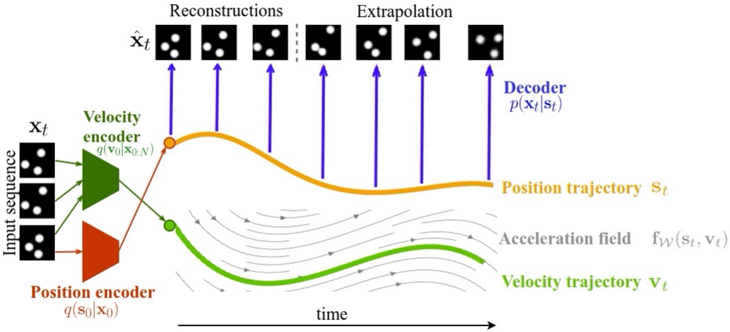
- Define an implicit prior over functions through dynamics
- Define observation likelihoods. Anything differentiable w.r.t. latent state (*e.g.* text models!)
- Train everything jointly with automatic differentiation + variational inference





Adopted form David Duvenaud.

Latent ODEs

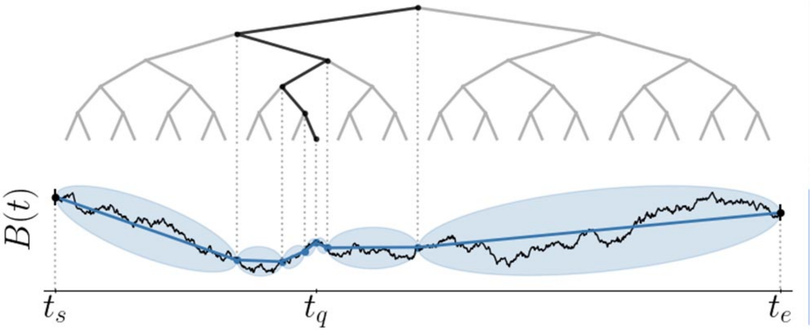


- Learning low-rank latent representations of possibly high-dimensional sequential data trajectories
- Combination of variational auto-encoders (VAEs) with sequential data with a latent space governed by a continuous-time ODE

Yildiz *et al.* (2019). NeurIPS.

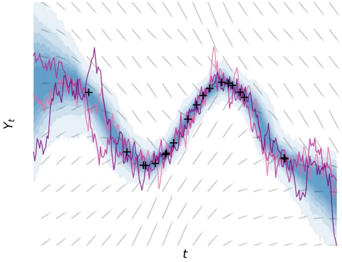
In SDEs: Need to store noise

- Infinite reparameterization trick: Use same Brownian motion sample on forward and reverse passes

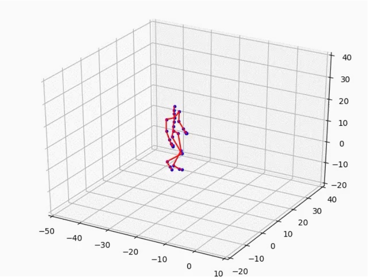


Li *et al.* (2020). AISTATS.

Latent SDE models



Ornstein–Uhlenbeck prior
Laplace likelihood

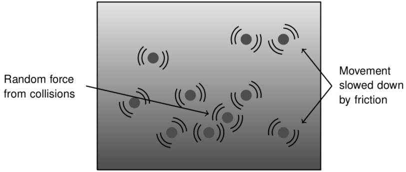


MOCAP example
50D data, 6D latent space,
11000 params

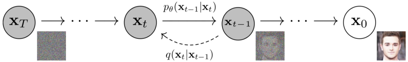
Li *et al.* (2020). AISTATS.

Langevin dynamics

- Molecular simulation
- Sampling and parameter inference (*e.g.*, Metropolis-adjusted Langevin algorithm)
- Diffusion models (even for images!)



Langevin's model for brownian motion



Denoising diffusion probabilistic model with Langevin dynamics

Ho *et al.* Submitted.

Summary

Part I

Tools and discrete-time models

Part II

SDEs (continuous-time models)

Part III

Gaussian processes

Part IV

Application examples

Acknowledgements

Simo Särkkä
(Aalto)

Manon Kok
(TU Delft)

Thomas Schön
(Uppsala)

Richard Turner
(Cambridge)

James Hensman
(Amazon)

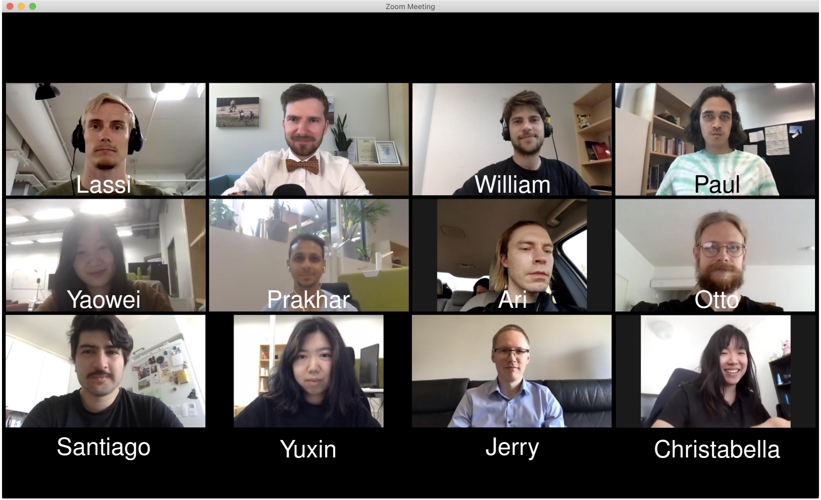
Nicolas Durrande
(prowler.io)

Michael Riis Andersen
(DTU)

Emtiyaz Khan
(RIKEN)

The following researchers kindly provided material/examples:
David Duvenaud, Markus Heinonen, Cagatay Yildiz, Jonathan Ho

Acknowledgements



My group at Aalto University

References

The examples in this tutorial were related to the following publications:

- Solin, A., Särkkä, S., Kannala, J., and Rahtu, E. (2016). Terrain navigation in the magnetic landscape: Particle filtering for indoor positioning. *European Navigation Conference (ENC)*.
- Solin, A., Cortés, S., Rahtu, E., and Kannala, J. (2018). PIVO: Probabilistic inertial-visual odometry for occlusion-robust navigation. *IEEE Winter Conference on Applications of Computer Vision (WACV)*.
- Särkkä, S., Solin, A., and Hartikainen, J. (2013). Spatio-temporal learning via infinite-dimensional Bayesian filtering and smoothing. *IEEE Signal Processing Magazine*, 30(4):51–61.
- Särkkä, S., and Solin, A. (2019). *Applied Stochastic Differential Equations*. Cambridge University Press. Cambridge, UK.
- Solin, A. (2016). *Stochastic Differential Equation Methods for Spatio-Temporal Gaussian Process Regression*. Doctoral dissertation, Aalto University.
- Solin, A., Hensman, J., and Turner, R.E. (2018). Infinite-horizon Gaussian processes. *Advances in Neural Information Processing Systems (NeurIPS)*.
- Durrande, N., Adam, V., Bordeaux, L., Eleftheriadis, E., Hensman, J. (2019). Banded matrix operators for Gaussian Markov models in the automatic differentiation era. *International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR 89:2780–2789.

References

The examples in this tutorial were related to the following publications:

- Nickisch, H., Solin, A., and Grigorievskiy, A. (2018). State apace Gaussian processes with non-Gaussian likelihood. *International Conference on Machine Learning (ICML)*. PMLR 80:3789–3798.
- Wilkinson, W. J., Chang, P., Riis Andersen, M., and Solin, A. (2020). State space expectation propagation: Efficient inference schemes for temporal Gaussian processes. *International Conference on Machine Learning (ICML)*.
- Hou, Y., Kannala, J. and Solin, A. (2019). Multi-view stereo by temporal nonparametric fusion. *International Conference on Computer Vision (ICCV)*.
- Hou, Y., Heljakka, A. and Solin, A. (submitted). Gaussian process priors for view-aware inference. arXiv:1912.03249.
- Yildiz, C., Heinonen, M., and Lähdesmäki, H. (2019). ODE2VAE: Deep generative second order ODEs with Bayesian neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*.
- Li, X., Wong, T.-K. L., Chen, R. T. Q., and Duvenaud, D. (2020). Scalable gradients for stochastic differential equations. *International Conference on Artificial Intelligence and Statistics (AISTATS)*.
- Ho, J., Jain, A., and Abbeel, P. (submitted). Denoising diffusion probabilistic models. arXiv:2006.11239.

Machine learning with signal processing

Part I

Tools and discrete-time models

Part II

SDEs (continuous-time models)

Part III

Gaussian processes

Part IV

Application examples