Gaussian Processes for time-series

Vincent Adam

Secondmind

Senior Machine Learning researcher

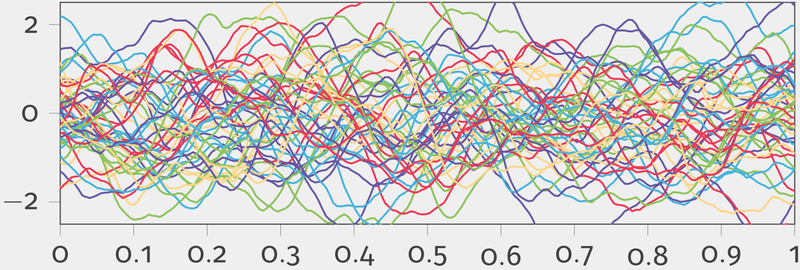
@vincentadam87

Cambridge Gaussian Process Meetup - 19/11/2020

Outline

- Recap on Gaussian Processes
- 3 views of Gaussian processes
- Gaussian processes as SDE
- Inference and learning
- Examples

Inspired by Arno Solin's ICML 2020 tutorial

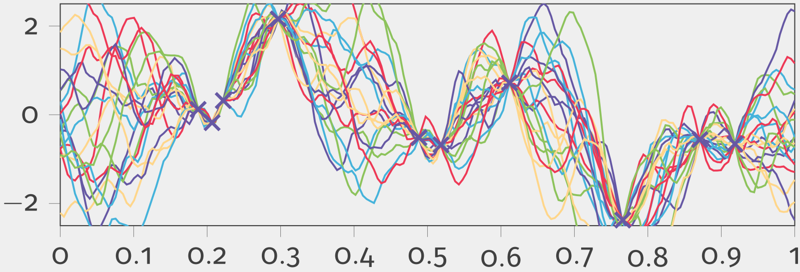


$$p(f(\cdot)) = \mathcal{GP}(0, k(\cdot, \cdot))$$

Gaussian Process Regression

$$p(f(\cdot)) = \mathcal{GP}(0, k(\cdot, \cdot))$$

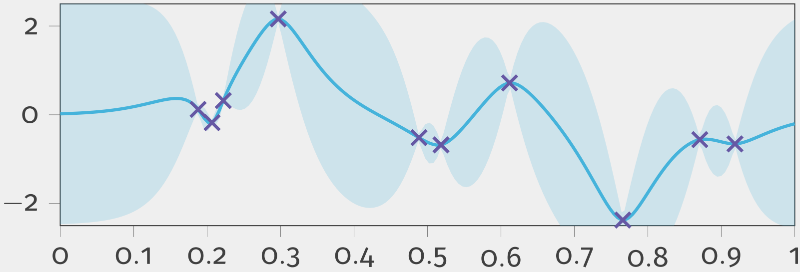
$$p(\mathbf{y}|f(\cdot)) = \prod_{i=1}^{n} p(y_i|f(t_i))$$



$$p(f(\cdot)) = \mathcal{GP}(0, k(\cdot,\cdot))$$

$$p(\mathbf{y}|f(\cdot)) = \prod_{i=1}^{n} p(y_i|f(t_i))$$

$$p(f(\cdot)|\mathbf{y}) = \frac{p(\mathbf{y}|f(\cdot))p(f(\cdot))}{p(\mathbf{y})}$$

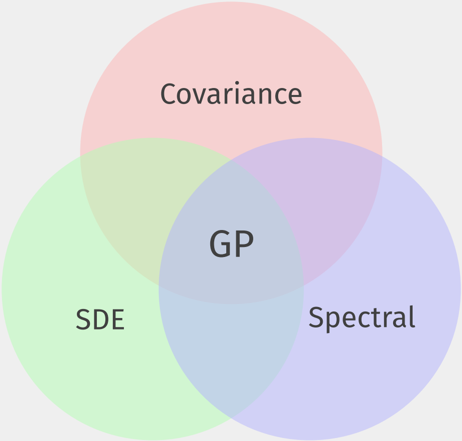


$$p(f(\cdot)) = \mathcal{GP}(0, k(\cdot,\cdot))$$

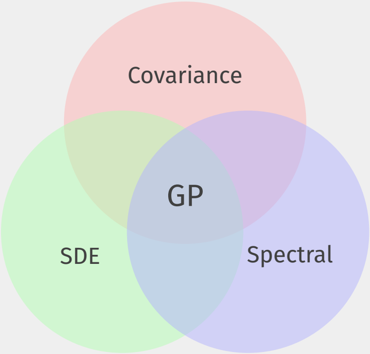
$$p(\mathbf{y}|f(\cdot)) = \prod_{i=1}^{n} p(y_i|f(t_i))$$

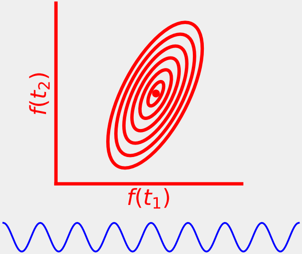
$$p(f(\cdot)|\mathbf{y}) = \frac{p(\mathbf{y}|f(\cdot))p(f(\cdot))}{p(\mathbf{y})}$$

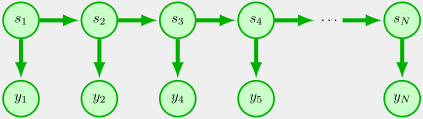
Three views of GPs [1]



Three views of GPs







Parameterization through mean μ and covariance function k

$$
\begin{aligned}
f &\sim \mathcal{GP}(\mu(\cdot), k(\cdot,\cdot)) \\
y \,|\, f &\sim \prod_n p(y_n \,|\, f(t_n))
\end{aligned}
$$

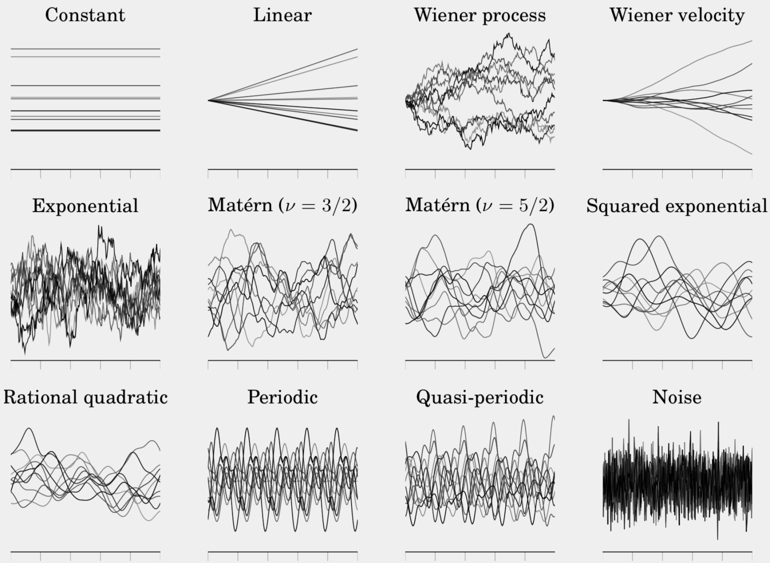
Parameterization through mean μ and covariance function k

$$\begin{aligned} f &\sim \mathcal{GP}(\mu(\cdot), k(\cdot, \cdot)) \\ y \,|\, f &\sim \prod_n p(y_n \,|\, f(t_n)) \end{aligned}$$

Noting $\mathbf{t} = (t_1, \ldots, t_N)$ and $\mathbf{f} = f(\mathbf{t})$, we have $\mathbf{f} \sim \mathcal{N}(\mu(\mathbf{t}), k(\mathbf{t}, \mathbf{t}))$

$$\begin{aligned} \mu(\mathbf{t}) &= \mathbb{E}[f(\mathbf{t})] \\ k(\mathbf{t}, \mathbf{t}) &= \mathbb{E}[(f(\mathbf{t}) - \mu(\mathbf{t}))(f(\mathbf{t}) - \mu(\mathbf{t}))^\top] \end{aligned}$$

Covariance view



Bochner theorem for stationary kernels

$$S(\omega) = \int e^{-i\omega\tau} k(\tau) d\tau$$

Power of Fourier components

$$s(i\omega) = \mathcal{F}[\omega](f) = \int e^{-i\omega t} f(t) dt$$
$$S(\omega) = \mathbb{E}[s(i\omega)\bar{s}(i\omega)^\top]$$

Example: Design through spectral density [2]

$$S(\omega) = \sum_i a_i \mathcal{N}(\omega \mid \mu_i, \sigma_i^2)$$

$$k(\tau) = \sum_i a_i \exp(-2\pi^2 \sigma_i^2 \tau^2) \cos(2\pi\tau\mu_i)$$

Gaussian Processes as SDE [3]

For Gaussian Process on the line (1d input), one can rewrite

$$f \sim GP(0, k(t,t')) \leftrightarrow \begin{cases} \frac{d\mathbf{s}(t)}{dt} = \mathbf{F}\,\mathbf{s}(t) + \mathbf{L}w(t) \\ f(t) = \mathbf{H}\,\mathbf{s}(t) \end{cases} \qquad (SDE)$$

Gaussian Processes as SDE [3]

For Gaussian Process on the line (1d input), one can rewrite

$$f \sim GP(0, k(t,t')) \leftrightarrow \begin{cases} \frac{d\mathbf{s}(t)}{dt} = \mathbf{F}\,\mathbf{s}(t) + \mathbf{L}w(t) \\ f(t) = \mathbf{H}\,\mathbf{s}(t) \end{cases} \qquad (SDE)$$

This can be discretized on $\mathbf{t} = [t_1, ..., t_T]$

$$\mathbf{f} \sim GP(0, K_{\mathbf{tt}}) \leftrightarrow \begin{cases} \mathbf{s}_{t+1} = \mathbf{A}_t\mathbf{s}_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \mathbf{Q}_t) \\ f_t = \mathbf{H}\,\mathbf{s}_t \end{cases} \qquad (SSM)$$

Gaussian Processes as SDE [3]

For Gaussian Process on the line (1d input), one can rewrite

$$f \sim GP(0, k(t,t')) \leftrightarrow \begin{cases} \frac{d\mathbf{s}(t)}{dt} = \mathbf{F}\,\mathbf{s}(t) + \mathbf{L}w(t) \\ f(t) = \mathbf{H}\,\mathbf{s}(t) \end{cases} \qquad (SDE)$$

This can be discretized on $\mathbf{t} = [t_1, ..., t_T]$

$$\mathbf{f} \sim GP(0, K_{\mathbf{tt}}) \leftrightarrow \begin{cases} \mathbf{s}_{t+1} = \mathbf{A}_t \mathbf{s}_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \mathbf{Q}_t) \\ f_t = \mathbf{H}\,\mathbf{s}_t \end{cases} \qquad (SSM)$$

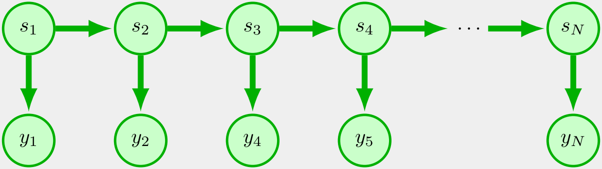
with

$$\mathbf{A}_i = e^{\Delta t_i \mathbf{F}} \quad \text{and} \quad \mathbf{Q}_i = \int_0^{\Delta t_i} e^{(\Delta t_i - \tau)\mathbf{F}}\,\mathbf{L}\,\mathbf{Q}_c\,\mathbf{L}^T\,e^{(\Delta t_i - \tau)\mathbf{F}^T} d\tau$$

$$\begin{cases} \mathbf{s}_{t+1} = \mathbf{A}_t \mathbf{s}_t + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, \mathbf{Q}_t) \\ f_t = \mathbf{H}\,\mathbf{s}_t \end{cases} \qquad (SSM)$$

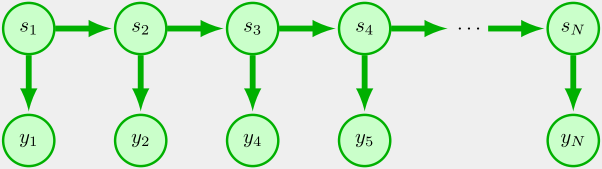
Markov property

$$\begin{cases} \mathbf{s}_{t+1} = \mathbf{A}_t \mathbf{s}_t + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, \mathbf{Q}_t) \\ f_t = \mathbf{H}\,\mathbf{s}_t \end{cases} \qquad (SSM)$$



Markov property

$$
\begin{cases} \mathbf{s}_{t+1} = \mathbf{A}_t \mathbf{s}_t + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, \mathbf{Q}_t) \\ f_t = \mathbf{H}\, \mathbf{s}_t \end{cases} \qquad (SSM)
$$

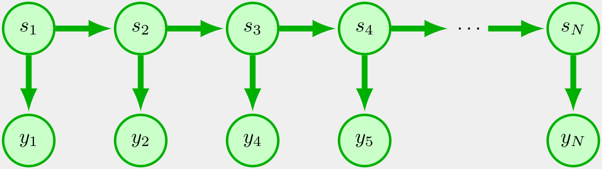


Markov Property

$$
\mathbf{s}_{t+1} \,\big|\, \mathbf{s}_0, \ldots, \mathbf{s}_t = \mathbf{s}_{t+1} \,\big|\, \mathbf{s}_t
$$

Markov property

$$\begin{cases} \mathbf{s}_{t+1} = \mathbf{A}_t \mathbf{s}_t + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, \mathbf{Q}_t) \\ f_t = \mathbf{H}\, \mathbf{s}_t \end{cases} \qquad (SSM)$$



Markov Property

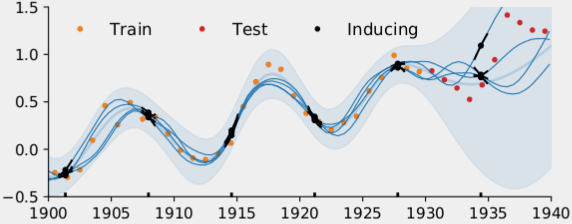
$$\mathbf{s}_{t+1} \,\big|\, \mathbf{s}_0, \ldots, \mathbf{s}_t = \mathbf{s}_{t+1} \,\big|\, \mathbf{s}_t$$

$$\mathbf{s}_t \,\big|\, \mathbf{s}_{\neg t} = \mathbf{s}_t \,\big|\, \mathbf{s}_{t-1}, \mathbf{s}_{t+1}$$

Gaussian Processes as SDE: Example

$$f \sim GP(0, k(t,t')) \leftrightarrow \begin{cases} \frac{d\mathbf{s}(t)}{dt} = \mathbf{F}\,\mathbf{s}(t) + \mathbf{L}w(t) \\ f(t) = \mathbf{H}\,\mathbf{s}(t) \end{cases} \qquad (SDE)$$

with $\mathbf{s} = [f, f', \ldots, f^{(d)}]$



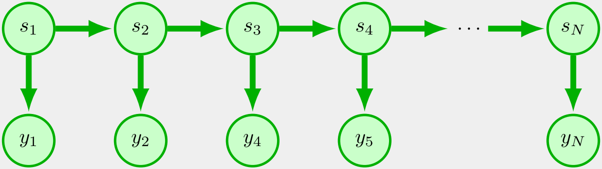
Inference in State Space Models [4]

$$\mathbf{f} \sim GP(0, K_{\mathbf{tt}}) \leftrightarrow \begin{cases} \mathbf{s}_{t+1} = \mathbf{A}_t \mathbf{s}_t + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, \mathbf{Q}_t) \\ f_t = \mathbf{H}\, \mathbf{s}_t \end{cases} \tag{SSM}$$

$$\mathbf{y} \,|\, \mathbf{f} \sim \prod_n p(\mathbf{y}_n \,|\, \mathbf{f}_n)$$

Inference and learning $\rightarrow$ **Kalman** filtering and smoothing

- both $p(\mathbf{f} \,|\, \mathbf{y})$ and $p(\mathbf{y})$ available in time $\mathcal{O}(Nd^3)$
- Sequential computation



Approximate Filtering $\mathcal{O}(N)$

- Sampling [5]
- Expectation Propagation (EP) [6, 7]
- Variational Inference (VI) [7, 8]
- ...
- (Sparse $\mathcal{GP}$) $\cap$ ($\mathcal{GP}$ as SDE) [9]

Kernel Combination

- Sum of kernels: $(d_1, d_2) \rightarrow d_1 + d_2$
- Product of kernels: $(d_1, d_2) \rightarrow d_1 \times d_2$

Model design [3]

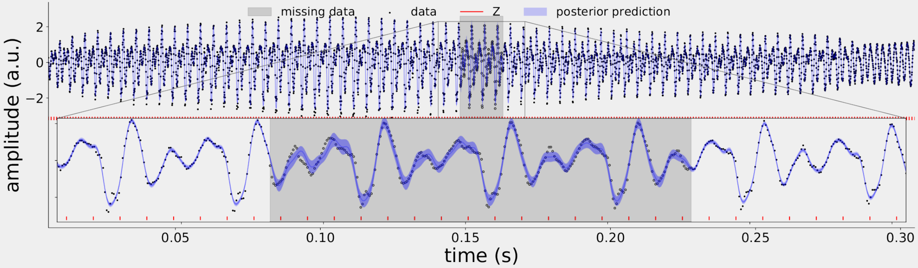
Kernel Combination

- Sum of kernels: $(d_1, d_2) \to d_1 + d_2$
- Product of kernels: $(d_1, d_2) \to d_1 \times d_2$

Multi-output Gaussian Processes

- o outputs: $\mathbf{f}_t = \mathbf{H}\,\mathbf{s}_t$, with $\mathbf{H} \in \mathbb{R}^{o \times d}$

Imputing missing segment in speech signal

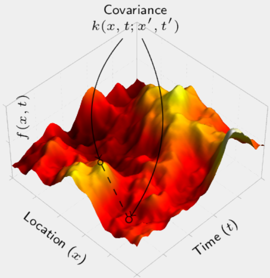


Harmonic kernel, Gaussian noise.

Spatio-Temporal Gaussian Processes [10]

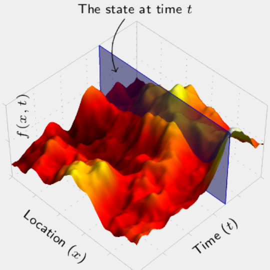
Kernel Formalism

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



Stochastic *partial* differential equation

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

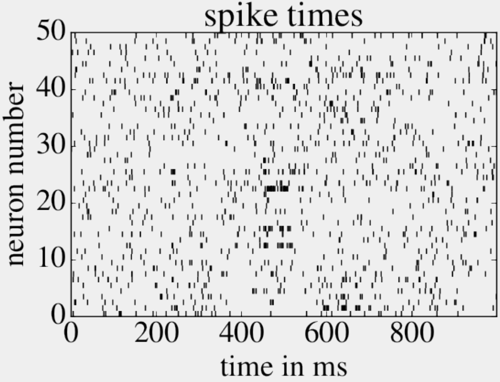


Scalar

- Classic Regression $y \mid \mathbf{f} = \mathcal{N}(y \mid \mathbf{f}_1, \sigma^2)$
- Heteroskedastic $y \mid \mathbf{f} = \mathcal{N}(y \mid \mathbf{f}_1, e^{\mathbf{f}_2})$

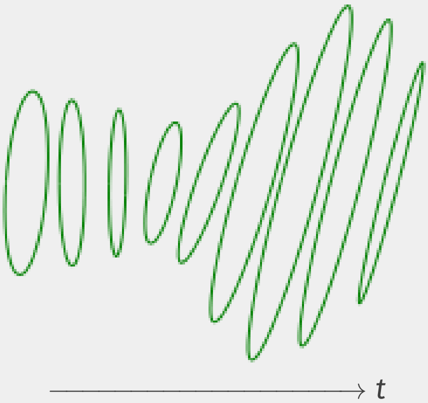
Multi-output (shared signal)

- $\mathbf{f}_t = \mathbf{H}\,\mathbf{s}_t$, with $\mathbf{H} \in \mathbb{R}^{o \times d}$



Multi-output (shared noise)

- Wishart: $\mathbf{y} \mid \mathbf{f} = \mathcal{N}(\mathbf{y} \mid 0, \mathbf{W}(\mathbf{f}))$



Summary

- 3 views of Gaussian processes
- Gaussian processes as SDE
- Inference and learning
- Examples

Thanks !

References I

Carl Edward Rasmussen and Christopher K. I. Williams.
Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning).
The MIT Press, 2005.

Andrew Wilson and Ryan Adams.
Gaussian process kernels for pattern discovery and extrapolation.
In *International conference on machine learning*, pages 1067–1075, 2013.

Simo Särkkä and Arno Solin.
***Applied stochastic differential equations*, volume 10.**
Cambridge University Press, 2019.

Simo Särkkä.
***Bayesian filtering and smoothing*, volume 3.**
Cambridge University Press, 2013.

References II

Thomas B Schön, Fredrik Gustafsson, and Rickard Karlsson.
The particle filter in practice.
In *The Oxford Handbook of Nonlinear Filtering*, pages 741–767.
Oxford University Press, 2011.

William J Wilkinson, Paul E Chang, Michael Riis Andersen, and Arno Solin.
State space expectation propagation: Efficient inference schemes for temporal gaussian processes.
2020.

Hannes Nickisch, Arno Solin, Alexander Grigorevskiy, et al.
State space gaussian processes with non-gaussian likelihood.
In *International Conference on Machine Learning*. PMLR, 2018.

- Nicolas Durrande, Vincent Adam, Lucas Bordeaux, Stefanos Eleftheriadis, and James Hensman.
  Banded matrix operators for gaussian markov models in the automatic differentiation era.
  In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2780–2789, 2019.

- Vincent Adam, Stefanos Eleftheriadis, Artem Artemev, Nicolas Durrande, and James Hensman.
  Doubly sparse variational gaussian processes.
  In *International Conference on Artificial Intelligence and Statistics*, pages 2874–2884, 2020.

- Simo Sarkka, Arno Solin, and Jouni Hartikainen.
  Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: A look at gaussian process regression through kalman filtering.
  IEEE Signal Processing Magazine, 30(4):51–61, 2013.

References IV

Paul E Chang, William J Wilkinson, Mohammad Emtiyaz Khan, and Arno Solin.
Fast variational learning in state-space gaussian process models.
In *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2020.