Explicit Link Between Periodic Covariance Functions and State Space Models

Arno Solin Simo Särkkä
Department of Biomedical Engineering and Computational Science
Aalto University, Finland

Aalto University

Introduction

- Gaussian processes (GPs, [2]) are commonly used modeling tools in non-parametric machine learning.
- Prior assumptions are encoded into the covariance function (kernel).
- We show that periodic covariance functions in GP regression can be rewritten as state space models.
- Reduces the problematic $\mathcal{O}(n^3)$ computational complexity to $\mathcal{O}(n)$ in the number of observations n.
- The model is written in terms of a series of stochastic resonators.
- Generalizes to quasi-periodic (almost periodic) covariance functions.

Gaussian Process Regression

- Kernel representation: In GP regression the model functions f are assumed to be realizations from a GP prior, and the observations y_k, $k = 1, 2, \ldots, n$, corrupted by Gaussian noise:

$$f(t) \sim \mathcal{GP}(0, k(t,t'))$$
$$y_k = f(t_k) + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, \sigma_n^2)$$

- Certain classes of covariance functions allow to work with the mathematical dual, where the Gaussian process is constructed as a solution to a mth order linear stochastic differential equation (SDE).
- State space representation: The GP regression problem can also be given as:

$$\frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} = \mathbf{F}\mathbf{f}(t) + \mathbf{L}\mathbf{w}(t)$$
$$y_k = \mathbf{H}\mathbf{f}(t_k) + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, \sigma_n^2),$$

where $\mathbf{w}(t)$ is a multi-dimensional white noise process with spectral density $\mathbf{Q}_c$.
- The model is defined by the feedback matrix $\mathbf{F}$, the noise effect matrix $\mathbf{L}$, the spectral density $\mathbf{Q}_c$, the stationary covariance $\mathbf{P}_\infty$, and the observation model $\mathbf{H}$.
  - The inference problem can now be solved using Kalman filtering [3] in $\mathcal{O}(nm^3)$ time complexity.

Periodic Covariance Functions

- The canonical periodic covariance function:

$$k_p(t,t') = \exp\left(-\frac{2\sin^2\left(\omega_0 \frac{t-t'}{2}\right)}{\ell^2}\right),$$

where ℓ is the characteristic length-scale and ω_0 defines the angular velocity (period length).
- The covariance function can be expanded into a (almost everywhere) convergent Fourier series ($\tau = |t - t'|$)

$$k_p(\tau) = \sum_{j=0}^{\infty} q_j^2 \cos(j\,\omega_0\,\tau).$$

- The differential equation model is a superposition the following kind of models [1]:

$$\mathbf{F}_j^p = \begin{pmatrix} 0 & -\omega_0 j \\ \omega_0 j & 0 \end{pmatrix},$$

and the diffusion part is zero (*i.e.* the model is deterministic), $\mathbf{H}_j^p = (1 \;\; 0)$, and $\mathbf{P}_{\infty,j}^p = q_j^2 \mathbf{I}_2$.
- The spectral (variance) coefficients q_j^2 are given by

$$q_j^2 = \frac{2\,\mathrm{I}_j(\ell^{-2})}{\exp(\ell^{-2})}, \quad \text{for } j = 1, 2, \ldots,$$

and $q_0^2 = \mathrm{I}_0(\ell^{-2}) / \exp(\ell^{-2})$, where $\mathrm{I}_\alpha(z)$ is the modified Bessel function.
- Taking the J first terms in the series gives an approximation, and this approximation converges uniformly [1] to the actual covariance as $J \to \infty$.

Quasi-Periodic Covariance Functions

- It is often desirable to allow for seasonable periodic variation, allowing the shape of the periodic effect to change over time.
- A common way of constructing quasi-periodic covariances is to take the product of a periodic covariance function $k_p(t,t')$ with a covariance function $k_q(t,t')$ with rather long characteristic length-scale,

$$k(t,t') = k_p(t,t')\,k_q(t,t'),$$

allowing the covariance to decay away from exact periodicity.
- The joint model corresponding to the quasi-periodic product of the two covariance functions can then be given [1] in a block-form:

$$\begin{aligned}
\mathbf{F}_j &= \mathbf{F}^q \otimes \mathbf{I}_2 + \mathbf{I}_q \otimes \mathbf{F}_j^p, \\
\mathbf{L}_j &= \mathbf{L}^q \otimes \mathbf{L}_j^p, \\
\mathbf{Q}_{c,j} &= \mathbf{Q}_c^q \otimes q_j^2\,\mathbf{I}_2, \\
\mathbf{P}_{\infty,j} &= \mathbf{P}_\infty^q \otimes \mathbf{P}_{\infty,j}^p, \\
\mathbf{H}_j &= \mathbf{H}^q \otimes \mathbf{H}_j^p,
\end{aligned}$$

where '$\otimes$' denotes the Kronecker product of two matrices.

Conclusions

- We have established the explicit connection between periodic covariance functions and state space models.
- This link enables the use of efficient sequential inference methods to solve periodic GP regression problems in $\mathcal{O}(n)$ time complexity.
- The approximation converges uniformly and a rough upper bound for the error can be given in closed-form.
- This is a 'best of both worlds' approach; it brings together the convenient model specification and hyperparametrization of GPs with the computational efficiency of state space models.

Example Implementation

- An example implementation is available on the author web page: `http://becs.aalto.fi/~asolin/`
- The method is also a part of the GPstuff toolbox for Matlab/Octave.

References

[1] A. Solin and S. Särkkä (2014). "Explicit link between periodic covariance functions and state space models." *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS).* JMLR W&CP vol. 33.
[2] C.E. Rasmussen and C.K.I. Williams (2006). "*Gaussian Processes for Machine Learning.*" The MIT Press.
[3] S. Särkkä (2013). "*Bayesin Filtering and Smoothing.*" Cambridge University Press.
[4] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. (2013). "*Bayesian Data Analysis.*" Third edition. Chapman & Hall/CRC Press.

These curves are random draws from a periodic GP prior and visualized in polar coordinates.

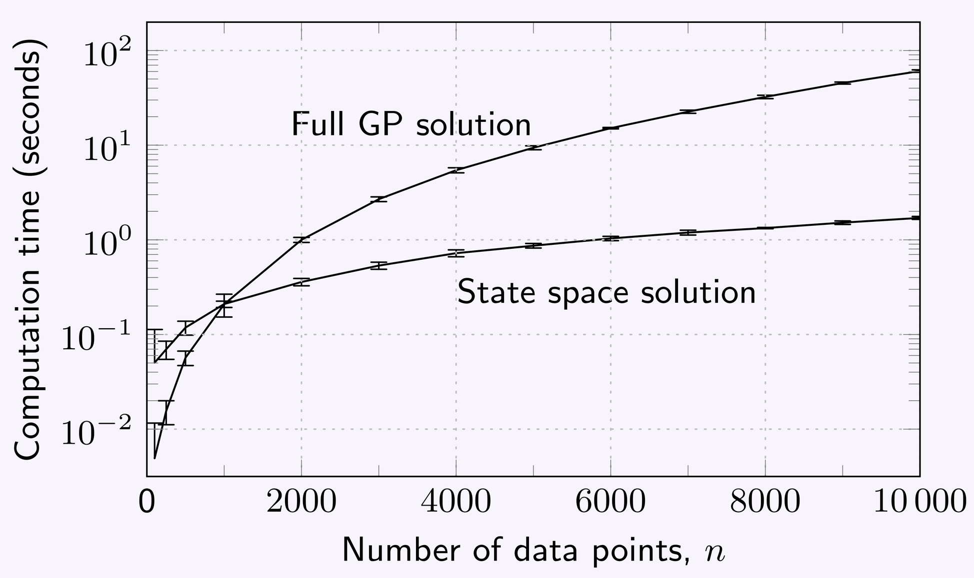
Contact details: arno.solin@aalto.fi, web address: http://arno.solin.fi

AISTATS 2014, Reykjavik, Iceland.

Demonstrations

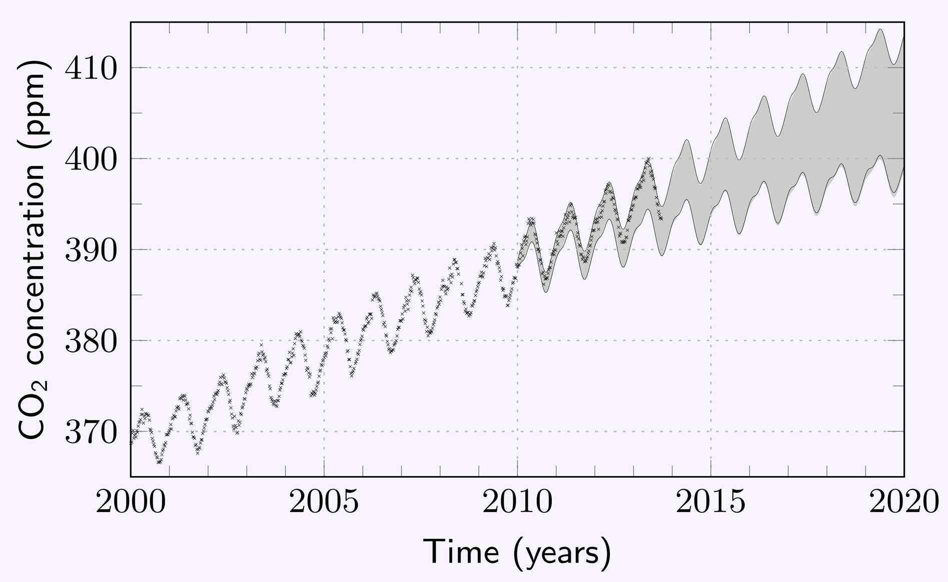
- A simulated example showing the computational efficiency.
- Prediction of CO_2 levels using weekly data (see [2]), where we compare the approximation to the full GP result.
- Explaining the periodic variation in the number of births per day in the US (see [4]).

Computational Efficiency



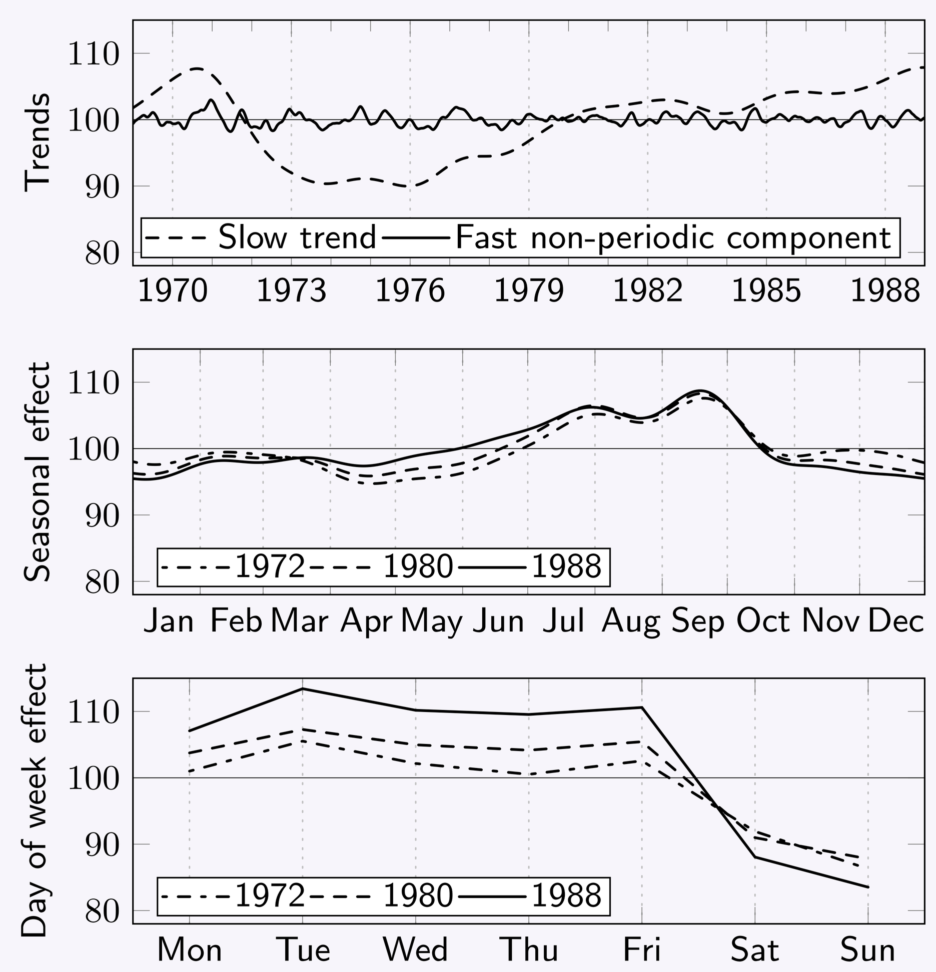
Demonstration of the computational benefits of the state space model in solving a GP regression problem for a number of data points up to 10 000 and with ten repetitions. The state space model execution times grow exactly linearly.

Consistency with the Full GP



CO_2 concentration observations ($n = 2227$, values for years 1958–2000 not shown in figure) together with the 95% predictive confidence region (the shaded patch is from the state space model, and the thin lines from the exact GP solution).

Example Study



Relative number of births in the US based on daily data between 1969–1988 ($n = 7305$). The first plot shows the non-periodic long-term effects, the two latter the quasi-periodic seasonal and weekly effects.