Explicit Link Between Periodic Covariance Functions and State Space Models

Arno Solin and Simo Särkkä

(AISTATS 2014)

Discussion by: Piyush Rai

December 12, 2014

Overview

- An explicit connection between GP regression with **periodic covariance functions** and state-space models
- Based on expanding the periodic covariance function into a series of stochastic resonators
- Allows scaling up GP regression with periodic covariance functions to large data sets
- Proposed method also extended to GPs **quasi-periodic** covariance functions

Gaussian Process Regression

- **The kernel view**

  Given: n training examples $\{(t_k, y_k)\}, k = 1, \ldots, n$

$$\begin{aligned} y_k &= f(t_k) + \epsilon_k \\ f(.) &\sim \mathcal{GP}(0, k(t, t')) \\ \epsilon_k &\sim \mathcal{N}or(0, \sigma_n^2) \end{aligned}$$

- Prior assumptions about f (e.g., smoothness, periodicity, etc.) encoded in the covariance function $k(t, t')$
- Can be solved in closed form but naïve solution is expensive: $\mathcal{O}(n^3)$ complexity at test time

- **The state-space view**

  Consider an m-th order SDE

$$\begin{aligned} \frac{d\mathbf{f}(t)}{dt} &= \mathbf{F}\mathbf{f}(t) + \mathbf{L}\mathbf{w}(t) \\ y_k &= \mathbf{H}\mathbf{f}(t_k) + \epsilon_k \end{aligned}$$

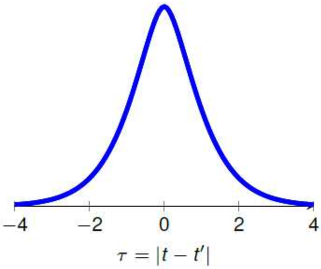
  where $\mathbf{f}(t)$ contains derivatives of $f(t)$ up to order $m-1$ and $\mathbf{w}(t)$ is the white noise process with spectral density $\mathbf{Q}_c$

- Model defined by $\mathbf{F}, \mathbf{L}, \mathbf{Q}_c$, stationary covariance $\mathbf{P}_\infty$, and the observation model $\mathbf{H}$
- Solved using Kalman filtering and has $\mathcal{O}(nm^3)$ time-complexity

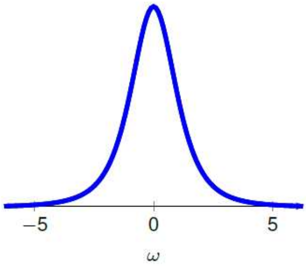
Covariance and Spectral Density

- Shown here: a stationary covariance function (Matérn) and its spectral density

$$k(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)\exp(-\mathrm{i}\,\omega\tau)\,\mathrm{d}\omega.$$



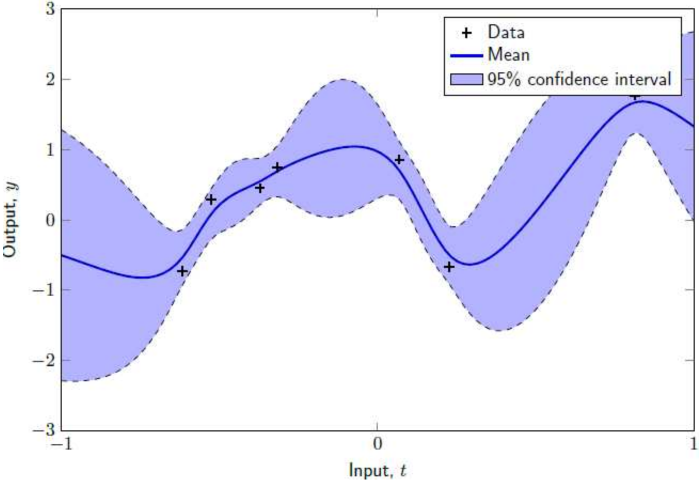
Covariance function $k(\tau)$



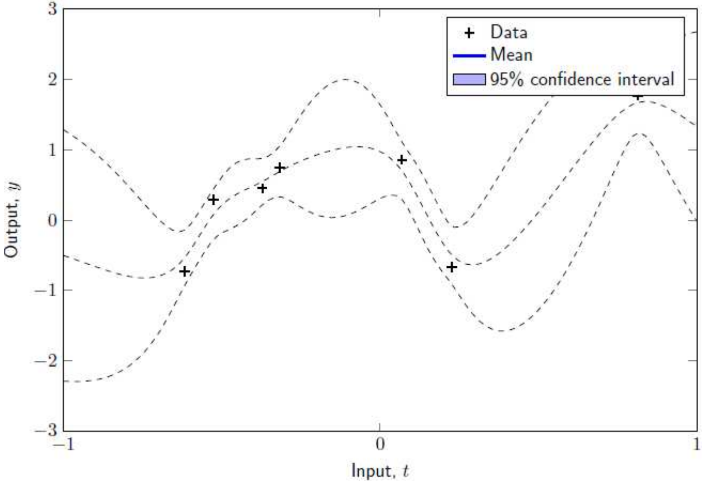
Spectral density $S(\omega)$

- This equivalence enables transforming the GP into a state-space model and solve the problem more efficiently in $\mathcal{O}(n)$ time

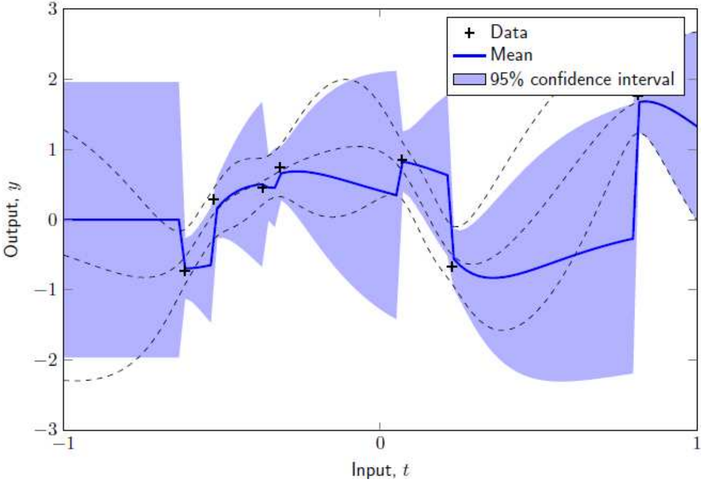
GP Regression (the naïve way)



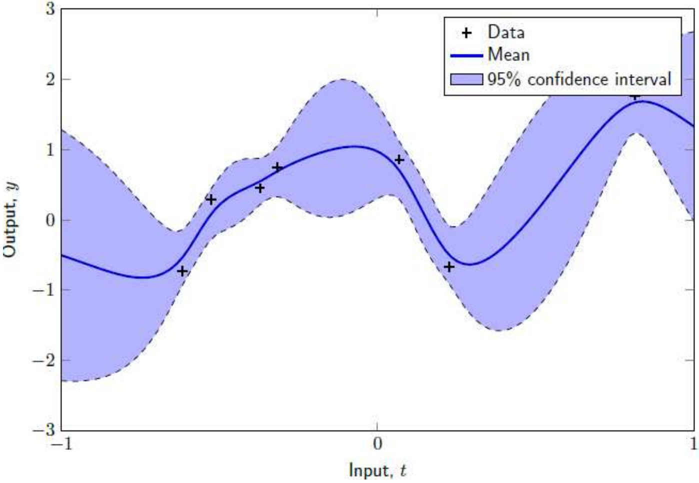
GP Regression (the naïve way)



GP Regression (via filtering and smoothing)

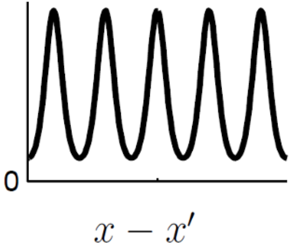


GP Regression (via filtering and smoothing)

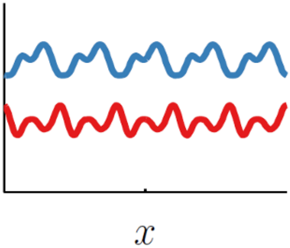


This Paper

How to establish the GP vs state-space model equivalence when the GP covariance function is periodic?



Periodic covariance function



Draws from a GP with this covariance function

Periodic Covariance Functions

Start off with the squared exponential:

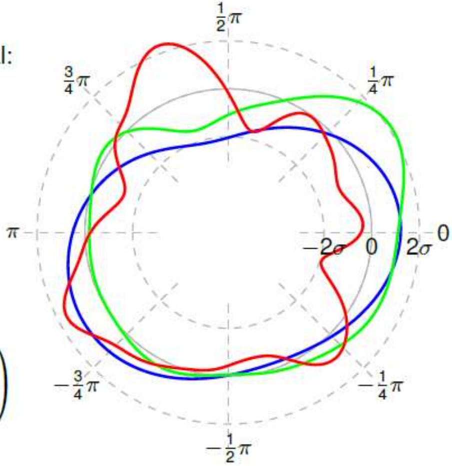
$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2}\right)$$

Polar coordinates:

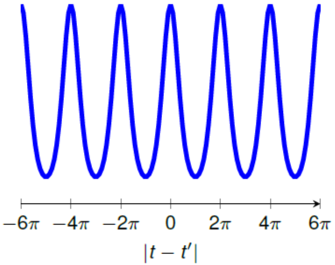
$$\mathbf{x}(t) = \begin{pmatrix} \cos(\omega_0\, t) \\ \sin(\omega_0\, t) \end{pmatrix}$$

The *canonical* periodic covariance:

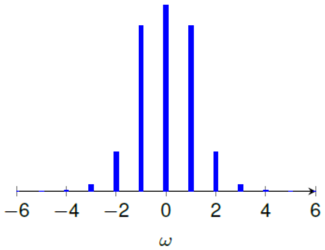
$$k_\mathrm{p}(t, t') = \sigma^2 \exp\left(-\frac{2\sin^2\left(\omega_0 \frac{t-t'}{2}\right)}{\ell^2}\right)$$



Periodic Covariance Functions



Covariance function



Spectral density

- Not amenable to the state-space transformation which requires the spectral density to be approximated by rational functions

State-Space Formulation

- Fourier series representation

$$k_p(\tau) = \sum_{j=1}^{\infty} q_j^2 \cos(j\omega_0\tau)$$

- Each periodic term j can be constructed as solution of a second-order ODE

$$\begin{pmatrix} \dot{x}_j(t) \\ \dot{y}_j(t) \end{pmatrix} = \begin{pmatrix} 0 & -\omega_0 j \\ \omega_0 j & 0 \end{pmatrix} \begin{pmatrix} x_j(t) \\ y_j(t) \end{pmatrix}$$

with $(x_j(0), y_j(0))^\top \sim \mathcal{N}or(0, q_j^2 \mathbf{I})$

- One way to determine the coefficients q_j^2 is via projection to the cosine basis

$$q_j^2 = \frac{\omega_0}{\pi} \int_{-\pi/\omega_0}^{+\pi/\omega_0} k_p(\tau) \cos(j\omega_0\tau) d\tau$$

but there are other way too

State-Space Formulation

- The state-space model $\frac{d\mathbf{f}(t)}{dt} = \mathbf{F}\mathbf{f}(t) + \mathbf{L}\mathbf{w}(t)$ will have block-diagonal matrices $\mathbf{F}$, $\mathbf{L}$, and $\mathbf{P}_\infty$ with blocks being (for $j = 1, \ldots, J$)

$$\mathbf{F}_j^p = \begin{pmatrix} 0 & -\omega_0 j \\ \omega_0 j & 0 \end{pmatrix}, \quad \mathbf{L}_j^p = \mathbf{I}_2, \quad \mathbf{P}_{\infty,j}^p = q_j^2 \mathbf{I}_2$$

  and the measurement model matrix $\mathbf{H}$ in $y_k = \mathbf{H}\mathbf{f}(t_k) + \epsilon_k$ is a block-row vector of $\mathbf{H}_j^p = (1 \quad 0)$. The diffusion part is zero (deterministic model).

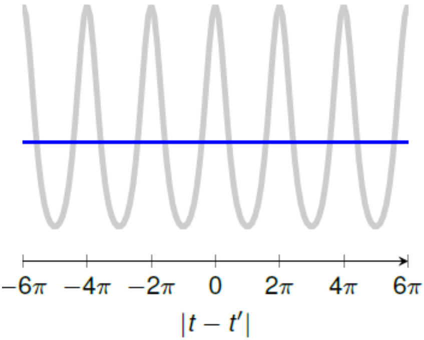
- The spectral (variance) coefficients

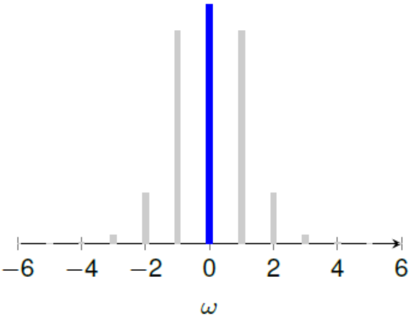
$$q_j^2 = \frac{2\mathrm{I}_j(l^{-2})}{\exp(l^{-2})}, \quad \text{for} \quad j = 1, 2, \ldots$$

  and $q_0^2 = \mathrm{I}_0(l^{-2})/\exp(l^{-2})$ where $\mathrm{I}_\alpha(z)$ is the modified Bessel function

- Taking the first J term of the series gives an approximation and this approximation converges uniformly to the actual covariance as $J \to \infty$

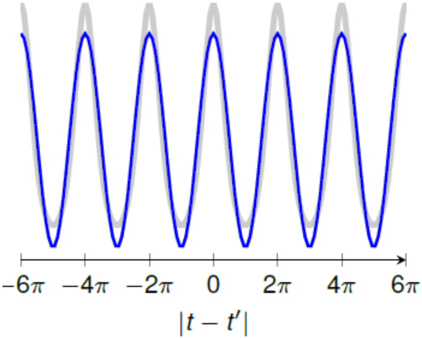
Approximated Covariance Functions (J=0)



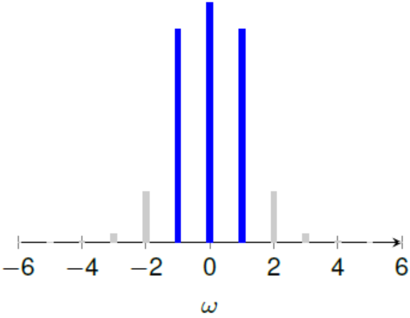


Covariance function **Spectral density**

Approximated Covariance Functions (J=1)

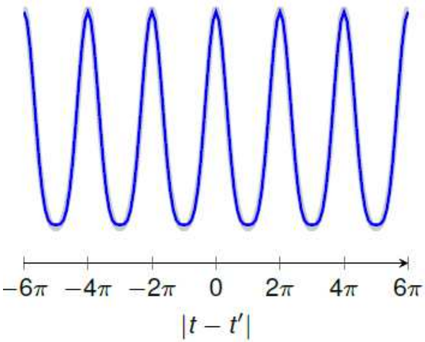


Covariance function

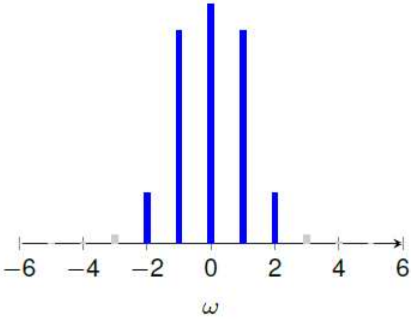


Spectral density

Approximated Covariance Functions (J=2)

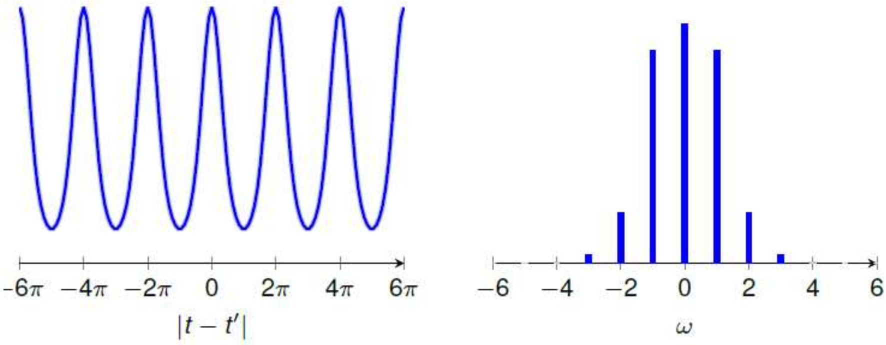


Covariance function



Spectral density

Approximated Covariance Functions (J=3)

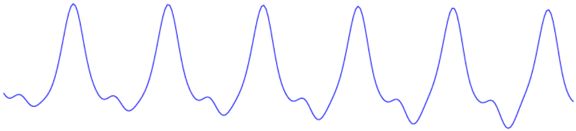


Covariance function

Spectral density

Quasi-Periodic Covariance Functions

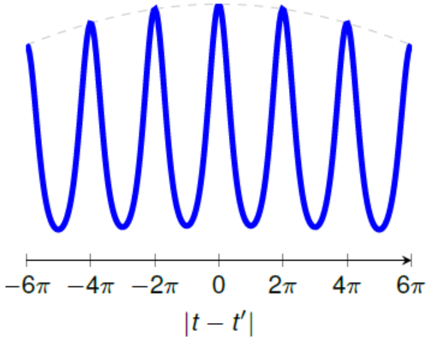
- The shape of the periodic effect may change with time



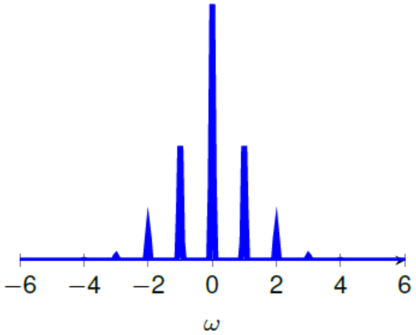
- Modeled using a product of a truly periodic covariance function $k_p(t, t')$ and another covariance function $k_q(t, t')$ with long characteristic length-scale

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

Quasi-Periodic Covariance Functions



Covariance function



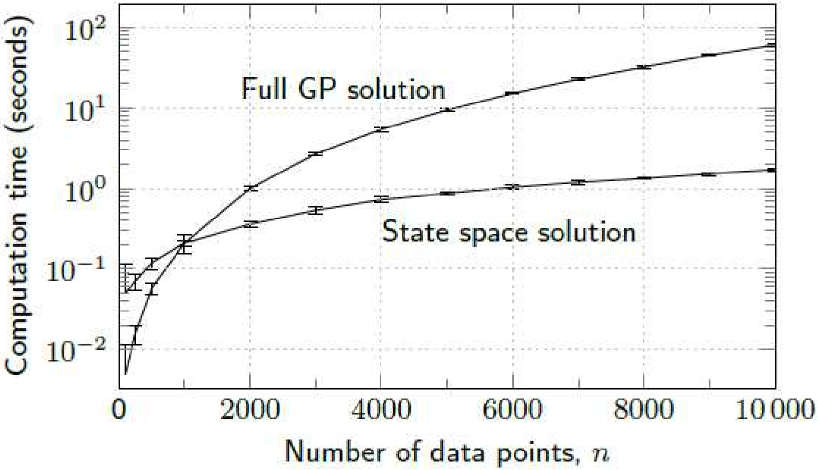
Spectral density

Quasi-Periodic Covariances: State-Space Form

- Have state-space representations for both quasi and periodic parts
- Set up the state-space such as the feedback matrices of both parts commute (i.e., $\mathbf{F}_p \mathbf{F}_q = \mathbf{F}_q \mathbf{F}_p$)
- Properties of Kronecker product help accomplish this
- The joint model for the quasi-periodic product of two covariance functions can be written in a block-form

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

Experiments: Computational Complexity

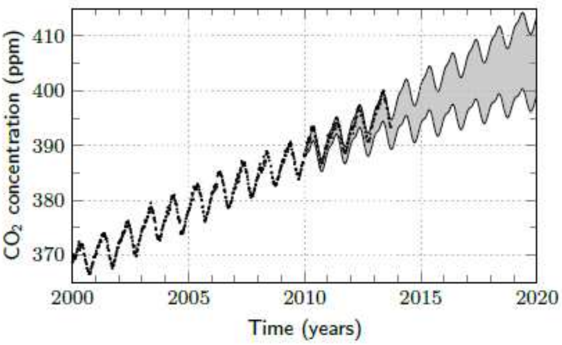


Experiments: CO_2 Concentration

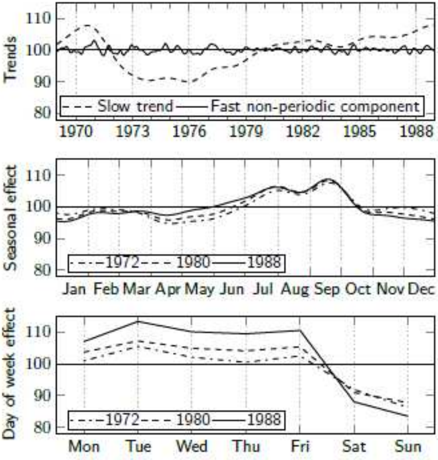
- Observations are CO_2 concentrations across time ($n = 2227$)
- GP covariance function

$$k(t, t') = k_{SE}(t, t') + k_p(t, t')k_{\nu=3/2}(t, t') + k_{\nu=3/2}(t, t')$$

- Converted to state-space, hyperparams optimized w.r.t. marginal likelihood



Experiments: Daily Births between 1969-1988



- Observations are number of daily births between 1969-1988 ($n = 7305$)
- The model: Matern ($\nu = 7/2$) for the slow trend, Matern ($\nu = 3/2$) for faster variation, Quasi-periodic (yearly) with Matern ($\nu = 3/2$) damping, Quasi-periodic (weekly) with Matern ($\nu = 3/2$) damping
- Converted to state-space, hyperparams optimized w.r.t. marginal likelihood

Conclusions

- Established connections between periodic covariance functions and state-space models
- The connection allows using efficient sequential inference methods (from state-space modeling) to solve periodic GP regression problem
- Approximation error due to truncation available in closed-form