Efficient computation for latent Gaussian spatio-temporal models

Simo Särkkä

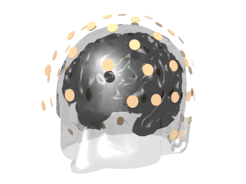
Aalto University, Finland
Visiting at the Dept. of Stats., **University of Oxford**, UK

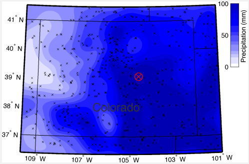
November, 2013

Contents

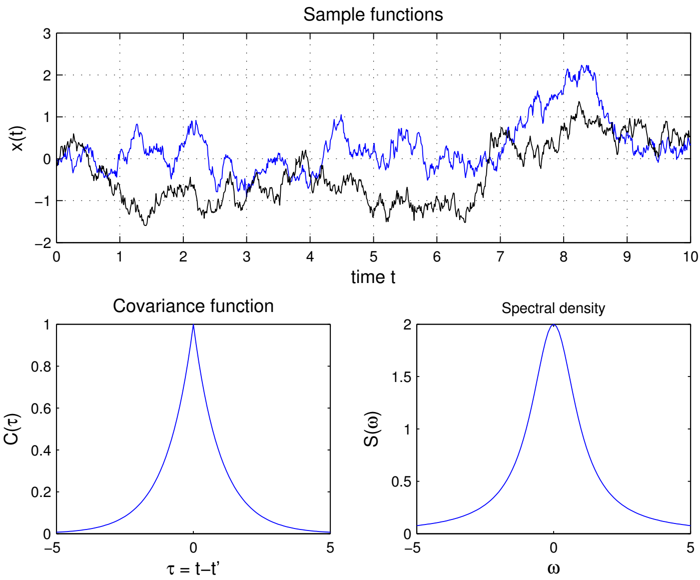
Motivating Applications

- Question arising in studies of brain activations: "Where are the activations located in brain?"
- Question arising in brain network analysis: "What is the spatio-temporal distribution of oscillations in brain?"
- Question arising in satellite positioning systems: "How can we predict the satellite position as accurately as possible?"
- Question arising in weather nowcasting: "How to predict/interpolate the evolution of precipitation?"
- Common characteristics: a huge number of data points, mainly in temporal direction.





Representations of Temporal Gaussian Processes



Representations of Temporal Gaussian Processes (cont.)

- Example: Ornstein-Uhlenbeck process $f(t)$ – path representation as a stochastic differential equation (SDE):

$$\frac{df(t)}{dt} = -\lambda\, f(t) + w(t).$$

  where $w(t)$ is a white noise process.
- The mean and covariance functions:

$$\begin{aligned} m(t) &= 0 \\ k(t, t') &= \exp(-\lambda |t - t'|) \end{aligned}$$

- Spectral density:

$$S(\omega) = \frac{2\lambda}{\omega^2 + \lambda^2}$$

- Ornstein-Uhlenbeck process $f(t)$ is Markovian – very efficient Bayesian inference possible with Kalman filter and smoother.

Rational Spectral Densities

- The key is the rational spectral density:

$$S(\omega) = \frac{(\text{polynomial in } \omega^2)}{(\text{polynomial in } \omega^2)}$$

- There always exists a state-space model with such a spectral density:

$$\begin{aligned} \frac{d\mathbf{f}(t)}{dt} &= \mathbf{A}\,\mathbf{f}(t) + \mathbf{L}\,w(t) \\ f(t) &= \mathbf{H}\,\mathbf{f}(t_i). \end{aligned}$$

$\hookrightarrow$ We can convert a large class of GPs into state-space models:

- ✓ The Matérn class has this form:

$$S(\omega) \propto (\lambda^2 + \omega^2)^{-(p+1)}.$$

- ✓ Non-rational spectral densities can be easily approximated:

$$S(\omega) = \sigma^2 \sqrt{\frac{\pi}{\kappa}} \exp\left(-\frac{\omega^2}{4\kappa}\right) \approx \frac{(\text{const})}{a_0 + a_1\,\omega^2 + \cdots + a_N\,\omega^{2N}}$$

Application to Gaussian Process Regression / Kriging

- Gaussian process regression (or Kriging) problems have the form

$$
\begin{aligned}
f(x) &\sim \mathcal{GP}(0, k(x, x')) \\
y_i &= f(x_i) + e_i, \qquad e_i \sim \mathcal{N}(0, \sigma^2_{\text{noise}}).
\end{aligned}
$$

- The computational complexity is awful $O(n^3)$, where n is the number of measurements.
- Renaming x into time t gives us a temporal model:

$$
\begin{aligned}
f(t) &\sim \mathcal{GP}(0, k(t, t')) \\
y_i &= f(t_i) + e_i, \qquad e_i \sim \mathcal{N}(0, \sigma^2_{\text{noise}}).
\end{aligned}
$$

- By converting into state-space form we get a model of the form:

$$
\begin{aligned}
\frac{d\mathbf{f}(t)}{dt} &= \mathbf{A}\,\mathbf{f}(t) + \mathbf{L}\,w(t) \\
y_i &= \mathbf{H}\,\mathbf{f}(t_i) + e_i.
\end{aligned}
$$

Kalman Filter and Rauch-Tung-Striebel Smoother

- Recall that Kalman filter and RTS smoother are algorithms for Bayesian inference in linear state-space models:

$$\frac{d\mathbf{f}(t)}{dt} = \mathbf{A}\,\mathbf{f}(t) + \mathbf{L}\,\mathbf{w}(t)$$
$$\mathbf{y}_i = \mathbf{H}\,\mathbf{f}(t_i) + \mathbf{e}_i$$

- Computational efficiency by utilizing Markov properties of Itô stochastic differential equations (SDEs).

$\hookrightarrow$ Many Gaussian process regression and Kriging problems can be efficiently solved with KF & RTS:

  - ✓ With n measurements, complexity of KF/RTS is $O(n)$, when the brute-force GP regression solution is $O(n^3)$.
  - ✓ Straight-forward generalization to non-linear/non-Gaussian models via non-linear Kalman and particle filtering/smoothing .

Example: Matérn Covariance Function

Example (1D Matérn covariance function)

- 1D Matérn family is ($\tau = |t - t'|$):

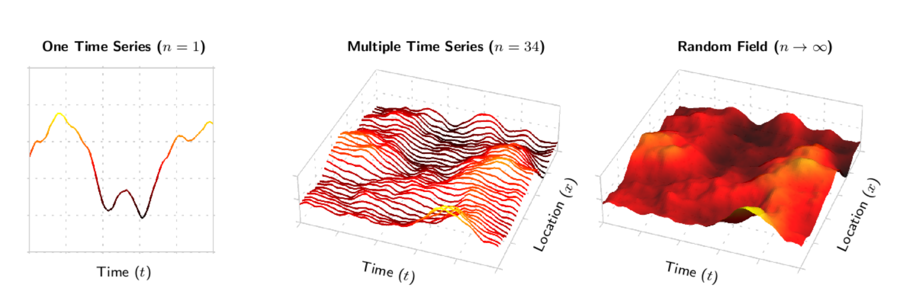
$$k(\tau) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\,\frac{\tau}{l}\right)^{\nu} K_\nu \left(\sqrt{2\nu}\,\frac{\tau}{l}\right),$$

  where $\nu, \sigma, l > 0$ are the smoothness, magnitude and length scale parameters, and $K_\nu(\cdot)$ the modified Bessel function.

- For example, when $\nu = 5/2$, we get

$$\frac{d\mathbf{f}(t)}{dt} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\lambda^3 & -3\lambda^2 & -3\lambda \end{pmatrix} \mathbf{f}(t) + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} w(t).$$

From Temporal to Spatio-Temporal Processes



The temporal vector-valued process becomes an infinite-dimensional function (Hilbert space) -valued process:

$$\mathbf{f}(t) = \begin{pmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{pmatrix} \to \begin{pmatrix} f(\mathbf{x}_1, t) \\ \vdots \\ f(\mathbf{x}_n, t) \end{pmatrix} \to f(\mathbf{x}, t), \quad \mathbf{x} \in \mathbb{R}^d.$$

Infinite-Dimensional Kalman Filtering and Smoothing

- Spatio-temporal Gaussian process regression (or Kriging) problem:

$$f(\mathbf{x}, t) \sim \mathcal{GP}(0, k(\mathbf{x}, t; \mathbf{x}', t'))$$
$$y_i = f(\mathbf{x}_i, t_i) + e_i, \qquad e_i \sim \mathcal{N}(0, \sigma^2_{\text{noise}}).$$

- Leads to an infinite-dimensional model with operators $\mathcal{A}$ and $\mathcal{H}_i$:

$$\frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial t} = \mathcal{A}\,\mathbf{f}(\mathbf{x}, t) + \mathbf{L}\,\mathbf{w}(\mathbf{x}, t)$$
$$\mathbf{y}_i = \mathcal{H}_i\,\mathbf{f}(\mathbf{x}, t_i) + \mathbf{e}_i$$

- We can use the infinite-dimensional Kalman filter and RTS smoother – scale linearly in time dimension.
- Solution with PDE methods such as basis function expansions, FEM, finite-differences, spectral methods, etc.

Example: 2D Matérn Covariance Function

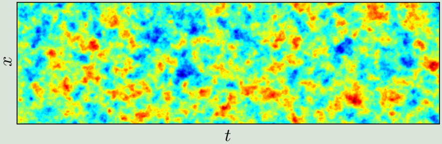
Example (2D Matérn covariance function)

- The multidimensional Matérn covariance function is the following ($r = ||\boldsymbol{\xi} - \boldsymbol{\xi}'||$, for $\boldsymbol{\xi} = (x_1, x_2, \ldots, x_{d-1}, t) \in \mathbb{R}^d$):

$$k(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\,\frac{r}{l}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\,\frac{r}{l}\right).$$

- For example, if $\nu = 1$ and $d = 2$, we get the following:

$$\frac{\partial \mathbf{f}(x,t)}{\partial t} = \begin{pmatrix} 0 & 1 \\ \partial^2/\partial x^2 - \lambda^2 & -2\sqrt{\lambda^2 - \partial^2/\partial x^2} \end{pmatrix} \mathbf{f}(x,t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(x,t).$$



The Basic Idea of State-Space Representation

- A latent force model is of the form

$$\frac{dz(t)}{dt} = g(z(t)) + u(t),$$

where $u(t)$ is the latent force.

- We measure the system at discrete instants of time:

$$y_k = z(t_k) + r_k$$

- Let's now model $u(t)$ as a Gaussian process of Matern type

$$k(\tau) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \, \frac{\tau}{l} \right)^{\nu} K_{\nu} \left(\sqrt{2\nu} \, \frac{\tau}{l} \right)$$

- Recall that if, for example, $\nu = 1/2$ then the GP can be expressed as the solution of the stochastic differential equation (SDE)

$$\frac{du(t)}{dt} = -\lambda \, u(t) + w(t)$$

The Basic Idea of State-Space Representation (cont.)

- If we define $\mathbf{f} = (z, u)$, we get a two-dimensional SDE

$$\frac{d\mathbf{f}}{dt} = \underbrace{\begin{pmatrix} g(f_1(t)) + f_2(t) \\ -\lambda\, f_2(t) \end{pmatrix}}_{\mathbf{a}(\mathbf{f})} + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{\mathbf{L}} w(t)$$

- We can now rewrite the measurement model as

$$y_k = \underbrace{\begin{pmatrix} 1 & 0 \end{pmatrix}}_{\mathbf{H}} \mathbf{f}(t_k) + r_k$$

- Thus the result is a model of the generic form

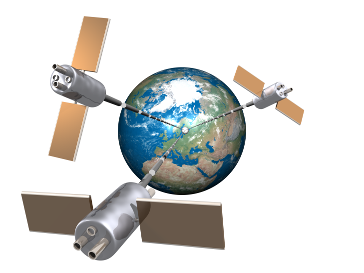
$$\begin{aligned} \frac{d\mathbf{f}}{dt} &= \mathbf{a}(\mathbf{f}) + \mathbf{L}\,\mathbf{w}(t) \\ \mathbf{y}_k &= \mathbf{H}\,\mathbf{f}(t_k) + \mathbf{r}_k. \end{aligned}$$

✓ Solution via non-linear Kalman/particle filtering and smoothing.
✓ Extends to spatio-temporal systems (PDEs).

GPS Satellite Orbit Prediction: Model

- Accurate orbit prediction improves Time To First Fix (TTFF) when network is not available for A-GPS.
- The equation of motion for the satellite can be written as

$$\frac{d}{dt}\begin{bmatrix}\mathbf{r}\\ \mathbf{v}\end{bmatrix} = \begin{bmatrix}\mathbf{v}\\ \mathbf{a}(\mathbf{r},t)+\mathbf{u}(\mathbf{r},\mathbf{v},t)\end{bmatrix}.$$

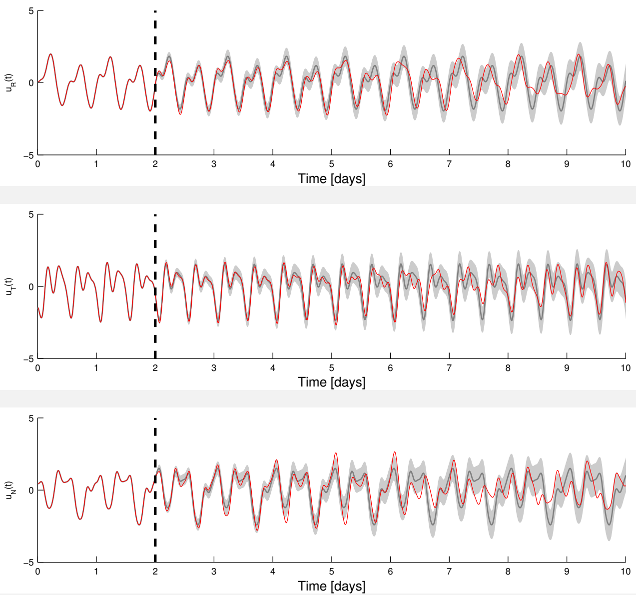


- The model for acceleration is

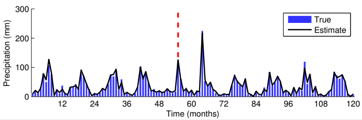
$$\mathbf{a}(\mathbf{r},t) = \mathbf{a}_{\mathrm{g}} + \mathbf{a}_{\mathrm{moon}} + \mathbf{a}_{\mathrm{sun}} + \mathbf{a}_{\mathrm{srp}}.$$

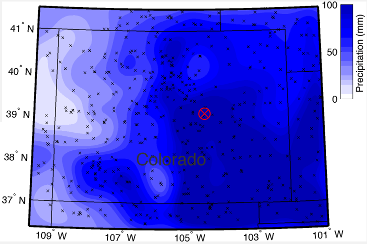
- Most modeling errors reside in the solar radiation pressure $\mathbf{a}_{\mathrm{srp}}$
- Unknown forces $\mathbf{u}(\mathbf{r},\mathbf{v},t)$ modeled as state-space GPs.

GPS Satellite Orbit Prediction: Prediction Results

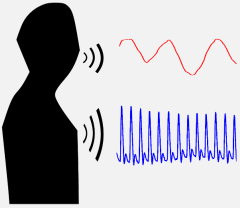


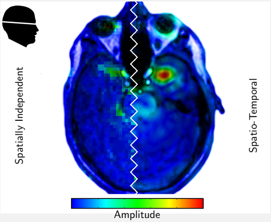
Spatio-Temporal Modeling of Precipitation





- Spatio-temporal interpolation of precipitation levels based on monthly data, years 1895–1997, Colorado, US.
- We used an infinite-dimensional state-space GP model with the non-separable spatio-temporal Matérn covariance function.
- Truncated eigenfunction expansion of the Laplace operator with 384 eigenfunctions.





- Spatio-temporal estimation of heart beat induced oscillations in fMRI brain data (measured at AMI centre, Finland).
- Superposition of spatio-temporal oscillators (GPs as well):

$$\frac{\partial^2 f_j(\mathbf{x},t)}{\partial t^2} + \mathcal{A}_j \frac{\partial f_j(\mathbf{x},t)}{\partial t} + \mathcal{B}_j f_j(\mathbf{x},t) = \xi_j(\mathbf{x},t).$$

- Spatial smoothness controlled by the spectral density kernel of $\xi_j(\cdot, t)$.

Conclusion

- Gaussian processes (and fields) have different representations:
  - Covariance function.
  - Spectral density.
  - Stochastic (partial) differential equation – a state space model.
- We can often convert between the representations:
  - Rational spectral densities $\Leftrightarrow$ state-space models.
  - Spatio-temporal models $\Leftrightarrow$ infinite-dimensional state-space models.
- Kalman filter and RTS smoother are algorithms for linear-time inference in state-space models ($O(n)$ vs. $O(n^3)$).
- State-space methods for latent force models:
  - GP models can be combined with differential equations.
  - The resulting model can be expressed as a state-space model.
  - Efficient Bayesian computation with non-linear filtering and smoothing.

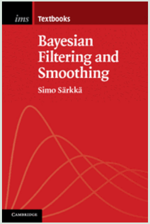
Key References

State-Space Representation of Temporal & Spatio-Temporal GPs

Simo Särkkä, Arno Solin, and Jouni Hartikainen (2013). *Spatio-Temporal Learning via Infinite-Dimensional Bayesian Filtering and Smoothing*. IEEE Signal Processing Magazine, 30(5):51–61.

Bayesian Methods for State-Space Models (Kalman filters and stuff)

Simo Särkkä (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press.



Note on Non-Markovian SPDE Models

- Let's consider the following Whittle SPDE

$$\frac{\partial^2 f(x,t)}{\partial x^2} + \frac{\partial^2 f(x,t)}{\partial t^2} - \lambda^2 f(x,t) = w(x,t),$$

  where $w(x,t)$ is a space–time white Gaussian random field.
- The covariance function is the same Matérn covariance function as in the previous example and with $\nu = 1$ and $d = 2$
- We can now form a state-space model by defining $\mathbf{f} = (f, \partial f / \partial t)$:

$$\frac{\partial \mathbf{f}(x,t)}{\partial t} = \begin{pmatrix} 0 & 1 \\ \lambda^2 - \partial^2/\partial x^2 & 0 \end{pmatrix} \mathbf{f}(x,t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(x,t),$$

  – but this state-space representation is different from the example!
- The catch is that the above model is non-Markovian whereas the one in the example is Markovian – we still need the spectral factorization.

Inference in Practice

- Conventional GP regression / Kriging:
  1. Evaluate the covariance function at the training and test set points.
  2. Use GP regression formulas to compute the posterior process statistics.
  3. Use the mean function as the prediction.
- State-space GP regression:
  1. Form the state space model.
  2. Run Kalman filter through the measurement sequence.
  3. Run RTS smoother through the filter results.
  4. Use the smoother mean function as the prediction.
- With both GP regression and state-space formulation we have the corresponding parameter estimation methods – see, e.g., Rasmussen & Williams (2006) and Särkkä (2013), respectively.