Gaussian Process Regression Through Kalman Filtering

Yixuan Zou, Ye Zi

December 11, 2019

Outline

- Introduction
- Method
- Conclusion

Outline

- Introduction
- Method
- Conclusion

Background

- Spatiotemporal Gaussian processes, or Gaussian fields, arise in many disciplines such as spatial statistics and kriging, machine learning, physical inverse problems, and signal processing
- In these applications, we are interested in doing statistical inference on the dynamic state of the whole field based on a finite set of indirect measurements as well as estimating the properties of the underlying process
- Gaussian processes are used as nonparametric priors for regressor functions, and the space and time variables take the roles of input variables of the regressor function
- In classical signal processing and stochastic control, Gaussian processes are commonly used for modeling temporal phenomena in form of stochastic differential equations (SDEs) and the inference procedure is usually solved using Kalman filter type of methods

Computational Issue

- A central practical problem in the Gaussian process regression context as well as in more general statistical inverse problems is the cubic $\mathcal{O}(N^3)$ computational complexity in the number of measurements N
- In the spatiotemporal setting, when we obtain, say, M measurements per time step and the total number of time steps is T, the complexity is $\mathcal{O}(M^3 T^3)$
- SDE models and state-space models in signal processing—their inference problem can be solved with Kalman (or Bayesian) filters and smoothers, which have a linear $\mathcal{O}(T)$ time complexity

Goal

Due to the computational efficiency of Kalman filters and smoothers, it is beneficial to reformulate certain spatiotemporal Gaussian process regression problems as Kalman filtering and smoothing problems. The aim of this presentation is to show when and how this is possible

Outline

- Introduction
- Method
- Conclusion

Gaussian Process

- **Gaussian process** is a random function $f(\xi)$ with d-dimensional input ξ such that any finite collection of random variables $f(\xi_1), \ldots, f(\xi_n)$ has a multidimensional Gaussian distribution.
- A Gaussian process can be defined in terms of a mean $m(\xi)$ and covariance function $k(\xi, \xi')$.
- A Gaussian process is **stationary** if

$$m(\xi) = c \ (constant)$$

$$k(\xi, \xi') = C(\xi' - \xi),$$

where $C(\xi)$ is another function.

  - Spatial Gaussian process, $\xi = x$;
  - Temporal Gaussian process, $\xi = t$;
  - Spatiotemporal Gaussian process, $\xi = (t, x)$.

Gaussian Process Regression

- **Gaussian process regression problem**

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

$$y = \mathcal{H}f(x) + \epsilon, \ \epsilon \sim N(0, \Sigma)$$

where $\Sigma = \sigma^2 I$, and the linear operator $\mathcal{H}$ picks the training set inputs among the function values

$$\mathcal{H}f(x) = (f(x_1), \ldots, f(x_N)).$$

- Goal: predict the value of y at a certain test point x^*, based on a training set $\{(x_i, y_i); i = 1, \ldots, N\}$
- By the property of Gaussian distributions,

$$p(f(x^*)|y_1, \ldots, y_N) = N(f(x^*)|\mu(x^*), V(x^*))$$

Computation complexity of the equations is $O(N^3)$.

- **State-spacing models**

$$\frac{df(t)}{dt} = Af(t) + L\omega(t)$$
$$y_k = Hf(t_k) + \epsilon_t,$$

where $k = 1, \ldots, T$, and A, L, and H are given matrices, ϵ_k is a vector of Gaussian measurement noises, and $\omega(t)$ is a vector of white noise process.

- The solution of f(t) is a Gaussian process.
- Kalman filter and Rauch-Tung-Stribel (RTS) smoother algorithm can be used to compute the mean and covariance of the following distribution for any t in linear time complexity

$$p(f(t)|y_1, \ldots, y_t) = N(f(t)|m_s(t), P_s(t)).$$

Combining the Approaches

- Spatiotemporal Gaussian process regression model

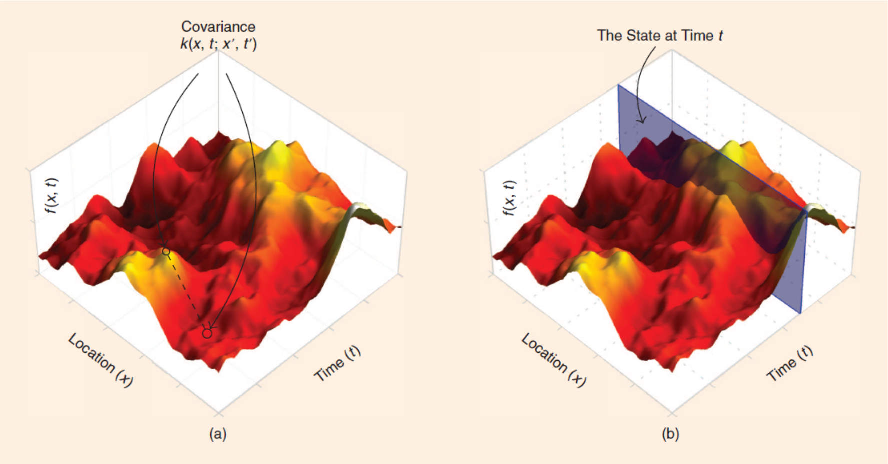
$$f(x,t) \sim \mathcal{GP}(0, k(x,t;x',t'))$$
$$y_k = \mathcal{H}_k f(x,t_k) + \epsilon_k$$

- Indefinite-dimensional state-space model

$$\frac{\partial f(x,t)}{\partial t} = \mathcal{A} f(x,t) + L\omega(x,t)$$
$$y_k = \mathcal{H}_k f(x,t_k) + \epsilon_k.$$

This model is an infinite-dimensional Markovian type of model that allows for linear-time inference with infinite dimensional Kalman filter and RTS smoother.

Combining the Approaches



[FIG1] (a) Covariance function view. In the covariance function-based representation, the spatiotemporal field is considered "frozen," and we postulate the covariance between two space–time points. (b) State-space model view. In the state-space model-based description, we construct a differential equation for the temporal behavior of a sequence of "snapshots" of the spatial field.

Converting Covariance Functions to State-Space Models

- Gaussian processes can be constructed as a solution to

$$a_n \frac{d^n f(t)}{dt^n} + \cdots + a_1 \frac{df(t)}{dt} + a_0 f(t) = \omega(t) \tag{1}$$

where $\omega(t)$ is a zero-mean continuous-time Gaussian white noise process.

- By Fourier transformation, we get

$$F(i\omega) = \underbrace{\left(\frac{1}{a_n (i\omega)^n + \cdots + a_1 (i\omega) + a_0} \right)}_{G(i\omega)} W(i\omega), \tag{17}$$

where $W(i\omega)$ is the (formal) Fourier transform of the white noise.

Covariance Functions of Stochastic Differential Equations

- The spectral density of the process is

$$S(\omega) = q_c |G(i\omega)|^2,$$

where $q_c = |W(i\omega)|^2$ the spectral density of white noice.

- **Wiener-Khinchin theorem:** the stationary covariance function of the process is given by the inverse Fourier transform of the spectral density

$$C(t) = \mathcal{F}^{-1}[S(\omega)] = \frac{1}{2\pi} \int S(\omega) \exp(i\omega t) d\omega, \tag{2}$$

and thus the corresponding covariance function is

$$k(t, t') = C(t - t').$$

Converting Covariance Functions to State-Space Models

- Equation (1) can be represented in the following state space form.
- Let $f = (f, df/dt, \ldots, d^{n-1}f/dt^{n-1})$, then we have

$$\frac{\mathrm{df}(t)}{\mathrm{d}t} = \underbrace{\begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-1} \end{pmatrix}}_{\mathrm{A}} \mathrm{f}(t) + \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}}_{\mathrm{L}} w(t). \tag{22}$$

$$y_k = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}}_{\mathrm{H}} \mathrm{f}(t) + \varepsilon_k, \tag{23}$$

- If we apply the Kalman filter and smoother to the state-space model described by (22) an (23), we will get the same result as if we applied Gaussian process regression equations to the covariance function defined by (2).
- The computational complexity of the Kalman filter and smoother is $\mathcal{O}(T)$, while the Gaussian process regression is $\mathcal{O}(T^3)$.

From Temporal Covariance Functions to State-Space Models

- **Goal**: find a state-space model with a component having a stationary covariance function $C(t)$ such that $k(t, t') = C(t - t')$.
- **Procedure:**
  1) Compute the corresponding spectral density $S(\omega)$ by computing the Fourier transform of $C(t)$.
  2) If $S(\omega)$ is not a rational function of the form

$$S(\omega) = \frac{m\text{th order polynomial in } \omega^2}{n\text{th order polynomial in } \omega^2}$$

  where $m < n$ with $a_n \neq 0$, then approximate (Taylor series expansions or Padé approximations) it with such a function.

From Temporal Covariance Functions to State-Space Models

- **Procedure Cont'**

  3) Find a stable rational transfer function $G(i\omega)$ of the form

  $$G(i\omega) = \frac{b_m(i\omega)^m + \cdots + b_1(i\omega) + b_0}{a_n(i\omega)^n + \cdots + a_1(i\omega) + a_0}$$

  and constant q_c such that

  $$S(\omega) = G(i\omega) q_c G(-i\omega).$$

  The procedure for finding a stable transfer function is called spectral factorization, one simple way is the following:

  - Compute the roots of the numerator and denominator polynomials of $S(\omega)$.
  - Construct the numerator and denominator polynomials of the transfer function $G(i\omega)$ from the positive-imaginary-part roots only.

From Temporal Covariance Functions to State-Space Models

- **Procedure Cont'**
  4) Use the methods from the control theory to convert the transfer function model into an equivalent state-space model. The constant q_c will then be the spectral density of the driving white noise process.

Example 1 (1-D Matern Covariance Function)

The one-dimentiaonal (1-D) isotropic and stationary covariance function of the Matern family can be given as

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

where ν, σ, l are are the smoothness, magnitude, and length scale parameters, and $K_{\nu}()$ is the modified Bessel function. The spectral density is of the form

$$S(\omega) \propto (\lambda^2 + \omega^2)^{-(\nu+1/2)},$$

where $\lambda = \sqrt{2}/l$. Hence,

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

where $\nu = p + 1/2$.

Example 1 (1-D Matern Covariance Function) Cont

The transfer function of the corresponding stable part is

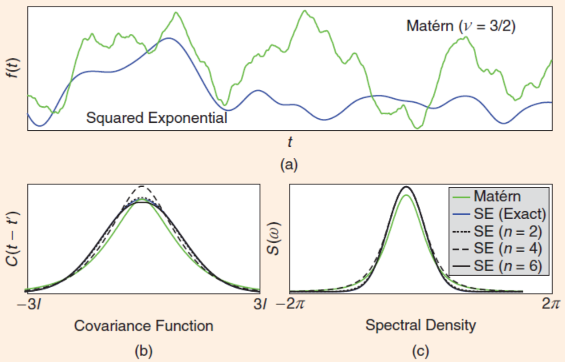
$$G(i\omega) = (\lambda + i\omega)^{-(p+1)}.$$

For integer values of p ($\nu = 1/2, 3/2, \cdots$), we can expand this expression using the binomial formula.
For example, if $p = 1$ ($\nu = 3/2$), the corresponding LTI SDE is

$$\frac{d\boldsymbol{f}(t)}{dt} = \begin{pmatrix} 0 & 1 \\ -\lambda & -2\lambda \end{pmatrix} \boldsymbol{f}(t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(t),$$

where $\boldsymbol{f}(t) = (f(t), \frac{f(t)}{dt})$.

Covariance, Spectral Density, and an Example Realization



[FIGS1] Part (a) shows random realizations drawn using the state-space models in "Example 1 (1-D Matérn Covariance Function)" and "Example 2 (1-D Squared Exponential Covariance Function)." The processes can be characterized through (b) their covariance functions or (c) using their spectral densities. The representation for the Matérn covariance function is exact, whereas the squared exponential needs to be approximated with a finite-order model (illustrated with $n = 2, 4, 6$ above). The errors in the tails of the spectral density transform into bias at the origin of the covariance function. With order $n = 6$, which was also used for drawing the random realization, both the approximations are already almost indistinguishable from the exact ones.

State Space Representation of Spatiotemporal Gaussian Process

- Goal: representing a spatiotemporal covariance function $k(x, t; x', t')$ in the state space form.
- nth-order spatial Fourier domain stochastic differential equation

$$a_n(i\omega_x)\frac{\partial^n \tilde{f}(i\omega_x, t)}{\partial t^n} + \cdots + a_1(i\omega_x)\frac{\partial \tilde{f}(i\omega_x, t)}{\partial t} + a_0(i\omega_x)\tilde{f}(i\omega_x, t) = \tilde{\omega}(i\omega_x, t),$$

- We can express the following equivalent state space form:

$$\frac{\partial \tilde{\mathrm{f}}(i\omega_x, t)}{\partial t} = \mathrm{A}(i\omega_x)\tilde{\mathrm{f}}(i\omega_x, t) + \mathrm{L}\tilde{w}(i\omega_x, t),$$

$$\mathrm{A}(i\omega_x) = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0(i\omega_x) & -a_1(i\omega_x) & \cdots & -a_{n-1}(i\omega_x) \end{pmatrix},$$

State Space Representation of Spatiotemporal Gaussian Process

- This leads to the equation:

$$\frac{\partial f(x,t)}{\partial t} = \mathcal{A} f(x,t) + L w(x,t),$$

where $\mathcal{A}$ is a matrix of linear operators as follows:

$$\mathcal{A} = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -\mathcal{A}_0 & -\mathcal{A}_1 & \cdots & -\mathcal{A}_{n-1} \end{pmatrix}$$

and the spatial covariance of the white noise is given by the inverse Fourier transform of $\tilde{q}_c(\omega_x)$. The operators $\mathcal{A}_j$ are defined as

$$\begin{aligned} \mathcal{A}_0 &= \mathcal{F}_x^{-1}[a_0(i\omega_x)], \\ \mathcal{A}_1 &= \mathcal{F}_x^{-1}[a_1(i\omega_x)], \\ &\vdots \\ \mathcal{A}_{n-1} &= \mathcal{F}_x^{-1}[a_{n-1}(i\omega_x)]. \end{aligned}$$

From Spatiotemporal Covariance Functions to State-Space Models

- **Goal:** find a infinite-dimensional state-space model with a component having a covariance function $k(x, t, x', t') = C(x' - x, t' - t)$.
- **Procedure:**
  1) Compute the corresponding spectral density $S(\omega_x, \omega_t)$ as the spatiotemporal Fourier transform of $C(x, t)$.
  2) Approximate the function $\omega_t \to S(\omega_x, \omega_t)$ with a rational function in variable ω_t^2.
  3) Find a stable ω_t-rational transfer function $G(i\omega_x, i\omega_t)$ and function $\tilde{q}_c(\omega_x)$ such that

  $$S(\omega_x, \omega_t) = G(i\omega_x, i\omega_t)\tilde{q}_c(\omega_x)G(-i\omega_x, -i\omega_t).$$

  The transfer function needs to have all its roots and zeros with respect to the ω_t variable in upper half plane, for all values of ω_x.

From Spatiotemporal Covariance Functions to State-Space Models

- **Procedure (Cont'):**
  4) Use the methods from control theory to convert the transfer function model into an equivalent spatial Fourier domian state-space model.
  5) Transform each of the coefficients $a_j(i\omega_x)$ and $b_j(i\omega_x)$ into the corresponding pseudodifferential operators and set the spatial stationary covariance function of the white noise process to the inverse Fourier transform of $\tilde{q}_c(\omega_x)$.

Example 3 (2-D Matern Covariance Function)

The multidimensional equivalent of the Matern covariance function given in "Example 1 (1-D Matern Covariance Function)" is the following:

$$C(\tau) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \big(\sqrt{2\nu}\frac{\tau}{l}\big)^{\nu} K_{\nu}\big(\sqrt{2\nu}\frac{\tau}{l}\big).$$

The corresponding spectral density is of the form

$$S(\boldsymbol{\omega_r}) = S(\boldsymbol{\omega_x}, \omega_t) \propto \frac{1}{(\lambda^2 + ||\boldsymbol{\omega_x}||^2 + \omega_t^2)^{\nu + d/2}},$$

where $\lambda = \sqrt{2\nu}/l$.

Example 3 (2-D Matern Covariance Function) Cont

The transfer function of the stable Markov process

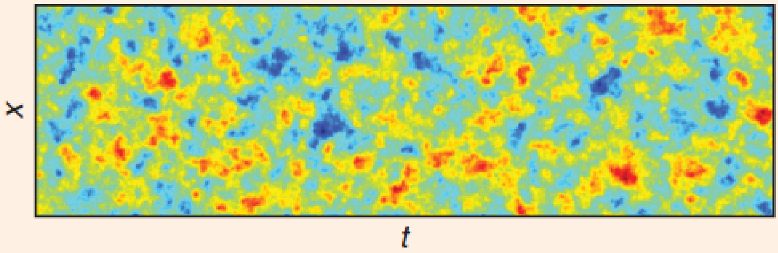
$$G(i\boldsymbol{\omega}_{\mathbf{x}}, i\omega_t) = (i\omega_t + \sqrt{\lambda^2 - ||i\boldsymbol{\omega}_{\mathbf{x}}||^2})^{\nu + d/2}.$$

The expansion of the denominator depends on the value of $p = \nu + d/2$. If p is an integer, the expansion can be easily done by the binomial theorem. For example, if $\nu = 1$ and $d = 2$, we get

$$\frac{d\boldsymbol{f}(x,t)}{dt} = \begin{pmatrix} 0 & 1 \\ \Delta^2 - \lambda^2 & -2\sqrt{\lambda^2 - \Delta^2} \end{pmatrix} \boldsymbol{f}(x,t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(x,t),$$

where $\boldsymbol{f}(x,t) = (f(x,t), \frac{f(x,t)}{dt})$.

An Example Realization



[FIGS2] A random realization simulated by the state-space model in (S1).

Outline

- Introduction
- Method
- Conclusion

Conclusion

- We have discussed the connection of Gaussian process regression and Kalman filtering and showed how certain classes of temporal or spatiotemporal Gaussian process regression problems can be converted into finite or infinitedimensional state-space models
- The advantage of this formulation is that it allows the use of computationally efficient linear-time-complexity Kalman filtering and smoothing methods
- One limitation in the present methodology is that it can only be used with stationary covariance functions
- Although the present method solves the problem of temporal time complexity, the space complexity is still cubic in the number of measurements, computationally efficient sparse approximations could be used