State Space Representation of Gaussian Processes

Simo Särkkä

Department of Biomedical Engineering and Computational Science (BECS)
Aalto University, Espoo, Finland

June 12th, 2013

Contents

Definition of Gaussian Process: Spatial Case

- **Spatial** Gaussian process (GP) is a spatial random function $\mathbf{f}(\mathbf{x})$, such that joint distribution of $\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_n)$ is always Gaussian.
- Can be defined in terms of mean and covariance functions:

$$
\begin{aligned}
\mathbf{m}(\mathbf{x}) &= \mathrm{E}[\mathbf{f}(\mathbf{x})] \\
\mathbf{K}(\mathbf{x}, \mathbf{x}') &= \mathrm{E}[(\mathbf{f}(\mathbf{x}) - \mathbf{m}(\mathbf{x}))\,(\mathbf{f}(\mathbf{x}') - \mathbf{m}(\mathbf{x}'))^T].
\end{aligned}
$$

- The joint distribution of a collection of random variables $\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_n)$ is then given as

$$
\begin{pmatrix} \mathbf{f}(\mathbf{x}_1) \\ \vdots \\ \mathbf{f}(\mathbf{x}_n) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{m}(\mathbf{x}_1) \\ \vdots \\ \mathbf{m}(\mathbf{x}_n) \end{pmatrix}, \begin{pmatrix} \mathbf{K}(\mathbf{x}_1, \mathbf{x}_1) & \ldots & \mathbf{K}(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \\ \mathbf{K}(\mathbf{x}_n, \mathbf{x}_1) & & \mathbf{K}(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix} \right)
$$

Definition of Gaussian Process: Temporal and Spatial-Temporal Cases

- **Temporal** Gaussian process (GP) is a temporal random function $\mathbf{f}(t)$, such that joint distribution of $\mathbf{f}(t_1), \ldots, \mathbf{f}(t_n)$ is always Gaussian.
- Mean and covariance functions have the form:

$$
\begin{aligned}
\mathbf{m}(t) &= \mathrm{E}[\mathbf{f}(t)] \\
\mathbf{K}(t, t') &= \mathrm{E}[(\mathbf{f}(t) - \mathbf{m}(t))\,(\mathbf{f}(t') - \mathbf{m}(t'))^T].
\end{aligned}
$$

- **Spatio-temporal** Gaussian process (GP) is a space-time random function $\mathbf{f}(\mathbf{x}, t)$, such that joint distribution of $\mathbf{f}(\mathbf{x}_1, t_1), \ldots, \mathbf{f}(\mathbf{x}_n, t_n)$ is always Gaussian.
- Mean and covariance functions have the form:

$$
\begin{aligned}
\mathbf{m}(\mathbf{x}, t) &= \mathrm{E}[\mathbf{f}(\mathbf{x}, t)] \\
\mathbf{K}(\mathbf{x}, \mathbf{x}'; t, t') &= \mathrm{E}[(\mathbf{f}(\mathbf{x}, t) - \mathbf{m}(\mathbf{x}, t))\,(\mathbf{f}(\mathbf{x}', t') - \mathbf{m}(\mathbf{x}', t'))^T].
\end{aligned}
$$

Modeling with Gaussian processes

- **Gaussian process regression**:
  - GPs are used as non-parametric prior models for "learning" input-output $\mathbb{R}^d \mapsto \mathbb{R}^m$ mappings in form $\mathbf{y} = \mathbf{f}(\mathbf{x})$.
  - A set of noisy training samples $\mathcal{D} = \{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)\}$ given.
  - The values of function $\mathbf{f}(\mathbf{x})$ at measurement points and test points are of interest.
- **Spatial analysis and Kriging**:
  - The variable $\mathbf{x}$ (input) is the spatial location.
  - GP is used for modeling similarities in $\mathbf{f}(\mathbf{x})$ at different locations.
  - The interpolated/smoothed values of $\mathbf{f}(\mathbf{x})$ are of interest.
- **Signal processing and time series analysis**:
  - In signal processing the input is the time t.
  - Time series is modeled as Gaussian process $\mathbf{f}(t)$ with a known spectrum or correlation structure.
  - The filtered/smoothed values at the measurement points and in other points are of interest.

Modeling with Gaussian processes (cont.)

- **Mechanics and electronics**:
  - In stochastic mechanical and electrical models, which typically arise in stochastic control and optimal filtering context, the input is time t.
  - The Gaussian process $\mathbf{f}(t)$ arises when a physical law in form of differential equation contains a stochastic (unknown) term.
  - The filtered/smoothed values at the measurement points and in other time points are of interest.
- **Continuum mechanics**
  - In stochastic continuum mechanical models, e.g., in meteorology and hydrology, the input consists of time t and spatial location $\mathbf{x}$.
  - Spatio-temporal Gaussian processes arise when a physical law in form of partial differential equation contains a stochastic term.
  - The interpolated/smoothed values of $\mathbf{f}(\mathbf{x}, t)$ at the measurement points and other points at different times t are of interest.

Fourier Transform

- The Fourier transform of function $f(\mathbf{x}) : \mathbb{R}^d \mapsto \mathbb{R}$ is

$$\mathcal{F}[f](\mathrm{i}\,\boldsymbol{\omega}) = \int_{\mathbb{R}^d} f(\mathbf{x}) \exp(-\mathrm{i}\,\boldsymbol{\omega}^T \mathbf{x}) \, d\mathbf{x}.$$

- The inverse Fourier transform of $\tilde{f}(\mathrm{i}\,\boldsymbol{\omega}) = \mathcal{F}[f](\mathrm{i}\,\boldsymbol{\omega})$ is

$$\mathcal{F}^{-1}[\tilde{f}](\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{f}(\mathrm{i}\,\boldsymbol{\omega}) \exp(\mathrm{i}\,\boldsymbol{\omega}^T \mathbf{x}) \, d\boldsymbol{\omega}.$$

- Properties of Fourier transform:
  - Linearity: For functions $f(\mathbf{x})$, $g(\mathbf{x})$ and constants $a, b \in \mathbb{R}$:

$$\mathcal{F}[a\,f + b\,g] = a\,\mathcal{F}[f] + b\,\mathcal{F}[g].$$

  - Derivative: If $f(\mathbf{x})$ is a k times differentiable function, then

$$\mathcal{F}[\partial^k f / \partial x_i^k] = (\mathrm{i}\,\omega_i)^k \, \mathcal{F}[f].$$

  - Convolution: The Fourier transform of the convolution is then the product of Fourier transforms of f and g:

$$\mathcal{F}[f * g] = \mathcal{F}[f] \, \mathcal{F}[g].$$

Covariance Functions and Spectral Densities

- Stationary GP: $k(\mathbf{x}, \mathbf{x}') \triangleq k(\mathbf{x} - \mathbf{x}')$ or just $k(\mathbf{x})$.
- Isotropic GP: $k(r)$ where $r = ||\mathbf{x} - \mathbf{x}'||$.
- The (power) spectral density of a function/process $f(\mathbf{x})$ is

$$S(\omega) = |\tilde{f}(\mathrm{i}\,\omega)|^2 = \tilde{f}(\mathrm{i}\,\omega)\,\tilde{f}(-\mathrm{i}\,\omega),$$

  where $\tilde{f}(\mathrm{i}\,\omega)$ is the Fourier transform of $f(\mathbf{x})$.
- Wiener-Khinchin: If $f(\mathbf{x})$ is a stationary Gaussian process with covariance function $k(\mathbf{x})$ then its spectral density is

$$S(\omega) = \mathcal{F}[k].$$

- Gaussian white noise is a zero-mean process with covariance function

$$k_w(\mathbf{x}) = q\,\delta(\mathbf{x}).$$

- The spectral density of the white noise is

$$S_w(\omega) = q.$$

Representations of Temporal Gaussian Processes

- Moment representation in terms of mean and covariance function

$$\mathbf{m}(t) = \mathrm{E}[\mathbf{f}(t)]$$
$$\mathbf{K}(t, t') = \mathrm{E}[(\mathbf{f}(t) - \mathbf{m}(t))\,(\mathbf{f}(t') - \mathbf{m}(t'))^{T}].$$

- Spectral representation in terms of spectral density function

$$\mathbf{S}(\omega_t) = \mathrm{E}[\tilde{\mathbf{f}}(\mathrm{i}\,\omega_t)\,\tilde{\mathbf{f}}^{T}(-\mathrm{i}\,\omega_t)].$$

- Path or state space representation as solution to a stochastic differential equation:

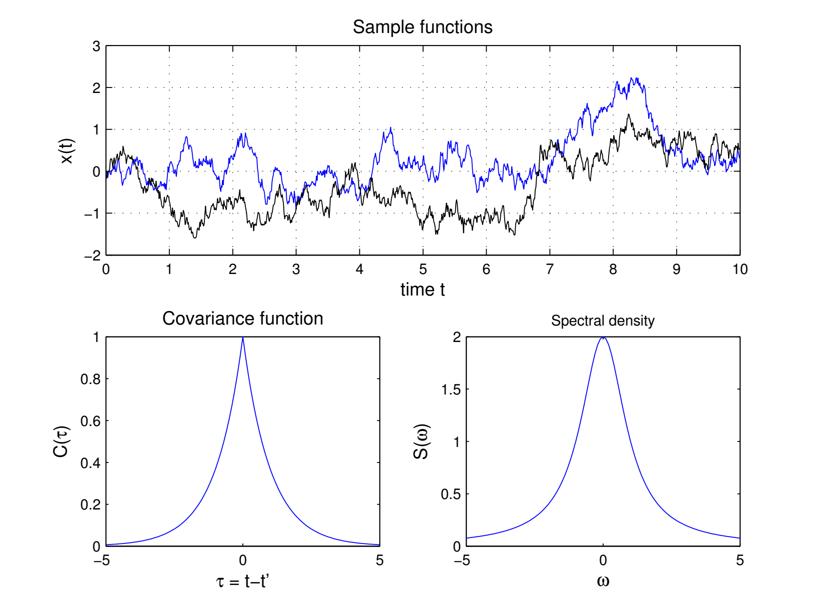
$$d\mathbf{f} = \mathbf{A}\,\mathbf{f}\,dt + \mathbf{L}\,d\boldsymbol{\beta},$$

where $\mathbf{f} \leftarrow (\mathbf{f}, d\mathbf{f}/dt, \ldots)$ and $\boldsymbol{\beta}(t)$ is a vector of Wiener processes, or equivalently, but more informally

$$\frac{d\mathbf{f}}{dt} = \mathbf{A}\,\mathbf{f} + \mathbf{L}\,\mathbf{w},$$

where $\mathbf{w}(t)$ is white noise.

Representations of Temporal Gaussian Processes



Scalar 1d Gaussian Processes

- Example of Gaussian process: Ornstein-Uhlenbeck process

$$m(t) = 0$$
$$k(t, t') = \exp(-\lambda |t - t'|)$$

- Path representation: stochastic differential equation (SDE)

$$df(t) = -\lambda\, f(t)\, dt + d\beta(t),$$

where $\beta(t)$ is a Brownian motion, or more informally

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

where $w(t)$ is a formal white noise process.

- The equation has the solution

$$f(t) = \exp(-\lambda\, t)\, f(0) + \int_0^t \exp(-\lambda(t - s))\, w(s)\, ds.$$

- Has the covariance $k(t, t')$ at stationary state.

Markov Property of Scalar 1d Gaussian Process

- Ornstein-Uhlenbeck process $f(t)$ is Markovian in the sense that given $f(t)$ the past $\{f(s), s < t\}$ does not affect the distribution of the future $\{f(s'), s' > t\}$.
- The marginal mean $m(t)$ and covariance $P(t) = k(t,t)$ satisfy the differential equations

$$
\begin{aligned}
\frac{dm(t)}{dt} &= -\lambda\, m(t) \\
\frac{dP(t)}{dt} &= -2\lambda\, P(t) + 2\lambda.
\end{aligned}
$$

- Due to the Markov property, these statistics are sufficient for computations, i.e., the full covariance $k(t, t')$ is not needed.
- The related inference algorithms, which utilize the Markov property are the Kalman filter and Rauch-Tung-Striebel smoother.

Aalto University

Spectral Representation of 1d Gaussian Process

- The spectral density of f(t) can be obtained by computing Fourier transform of the covariance $k(t, t') = k(t - t')$:

$$S(\omega) = \int_{-\infty}^{\infty} \exp(-\lambda|\tau|) \exp(-\mathrm{i}\,\omega\,\tau)\, d\tau = \frac{2\lambda}{\omega^2 + \lambda^2}.$$

- Alternatively, we can take Fourier transform of the original equation $df(t)/dt = -\lambda f(t) + w(t)$, which yields

$$(\mathrm{i}\,\omega)\,\tilde{f}(\mathrm{i}\,\omega) = -\lambda\,\tilde{f}(\mathrm{i}\,\omega) + \tilde{w}(\mathrm{i}\,\omega),$$

  and further

$$\tilde{f}(\mathrm{i}\,\omega) = \frac{\tilde{w}(\omega)}{(\mathrm{i}\,\omega) + \lambda}$$

- The spectral density is by definition $S(\omega) = |\tilde{f}(\mathrm{i}\,\omega)|^2$:

$$S(\omega) = \frac{|\tilde{w}(\mathrm{i}\,\omega)|^2}{((\mathrm{i}\,\omega) + \lambda)((-\mathrm{i}\,\omega) + \lambda)} = \frac{2\lambda}{\omega^2 + \lambda^2}$$

Vector Valued 1d Gaussian Processes

- Vector valued 1d Gaussian processes correspond to linear time-invariant state space models with white noise input $\mathbf{w}(t)$:

$$\frac{d\mathbf{f}(t)}{dt} = \mathbf{A}\,\mathbf{f}(t) + \mathbf{L}\,\mathbf{w}(t).$$

- Can be solved in terms of the matrix exponential $\exp(t\,\mathbf{A})$:

$$\mathbf{f}(t) = \exp(t\,\mathbf{A})\,\mathbf{f}(0) + \int_0^t \exp((t-s)\,\mathbf{A})\,\mathbf{L}\,\mathbf{w}(s)\,ds.$$

- The covariance function and spectral density at stationary state are

$$\mathbf{K}(t,t') = \begin{cases} \mathbf{P}_\infty \exp((t'-t)\,\mathbf{A})^T & , \quad \text{if } t' \geq t \\ \exp((t-t')\,\mathbf{A})\,\mathbf{P}_\infty & , \quad \text{if } t' < t. \end{cases}$$

$$\mathbf{S}(\omega) = (\mathbf{A} + \mathrm{i}\,\omega\,\mathbf{I})^{-1}\mathbf{L}\,\mathbf{Q}\,\mathbf{L}^T(\mathbf{A} - \mathrm{i}\,\omega\,\mathbf{I})^{-T}.$$

where $\mathbf{P}_\infty$ is the solution to the equation

$$\mathbf{A}\,\mathbf{P}_\infty + \mathbf{P}_\infty\,\mathbf{A}^T + \mathbf{L}\,\mathbf{Q}\,\mathbf{L}^T = 0.$$

Converting Covariance Functions to State Space Models

- Consider a Nth order LTI SDE of the form

$$\frac{d^N f}{dt^N} + a_{N-1} \frac{d^{N-1} f}{dt^{N-1}} + \cdots + a_0 f = w(t).$$

- This can be expressed as state space model as

$$\frac{d\mathbf{f}}{dt} = \underbrace{\begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & -a_1 & \dots & -a_{N-1} \end{pmatrix}}_{\mathbf{A}} \mathbf{f} + \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}}_{\mathbf{L}} w(t)$$

$$f(t) = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}}_{\mathbf{H}} \mathbf{f},$$

where $\mathbf{f} = (f, \dots, d^{N-1} f / dt^{N-1})$.

Converting Covariance Functions to State Space Models (cont.)

- By taking the Fourier transform of the equation, we get a spectral density of the following form for $f(t)$:

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

- $\Rightarrow$ We can convert covariance functions into state space models by writing or approximating the spectral density in the above form:
  - With certain parameter values, the Matérn has this form:

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

  - The exponentiated quadratic can be easily approximated:

$$S(\omega) = \sigma^2 \sqrt{\frac{\pi}{\kappa}} \exp\left(-\frac{\omega^2}{4\kappa}\right) \approx \frac{\text{(const)}}{N!/0!(4\kappa)^N + \cdots + \omega^{2N}}$$

- In conversion of the spectral density to differential equation, we need to do so called spectral factorization.

Converting Covariance Functions to State Space Models (cont.)

- The Gaussian process regression problem of the form

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

- or actually

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

- can be thus converted into state estimation problem 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

- Kalman filter and RTS smoother can be used for efficiently computing posteriors of models in form

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

  where $\mathbf{y}_i$ is the measurement and $\mathbf{e}_i \sim N(0, \mathbf{R}_i)$.

- Can be equivalently (in weak sense) expressed as discrete-time model

$$\mathbf{f}_i = \mathbf{U}_i\,\mathbf{f}_{i-1} + \mathbf{v}_i$$
$$\mathbf{y}_i = \mathbf{H}\,\mathbf{f}_i + \mathbf{e}_i$$

- Many Gaussian process regression problems with differentiable covariance function can be efficiently solved with KF & RTS
- With n measurements, complexity of KF/RTS is $O(n)$, when the brute-force GP solution is $O(n^3)$.

Aalto University

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.
- The spectral density is of the form

$$S(\omega) = \frac{q}{(\lambda^2 + \omega^2)^{(\nu+1/2)}},$$

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

Example: Matérn Covariance Function (cont.)

Example (1D Matérn covariance function (cont.))

- The spectral density can be factored as

$$S(\omega) = \frac{q}{(\lambda + i\,\omega)^{(p+1)}\,(\lambda - i\,\omega)^{(p+1)}},$$

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

- The transfer function of the corresponding stable part is

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

- For integer values of p ($\nu = 1/2, 3/2, \ldots$), we can expand this. For example, if $p = 0$ ($\nu = 1/2$), we get the Ornstein–Uhlenbeck process

$$\frac{\mathrm{d}f(t)}{\mathrm{d}t} = -\lambda\, f(t) + w(t)$$

Example: Matérn Covariance Function (cont.)

Example (1D Matérn covariance function (cont.))

- $p = 1$ $(\nu = 3/2)$ gives

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

  where $\mathbf{f}(t) = (f(t), \mathrm{d}f(t)/\,\mathrm{d}t)$.

- $p = 2$ in turn gives

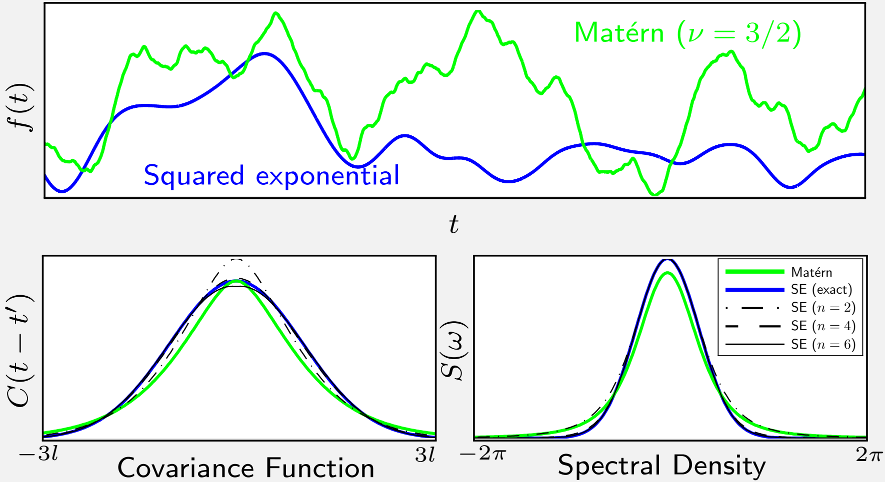
$$\frac{\mathrm{d}\mathbf{f}(t)}{dt} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\lambda^3 & -3\lambda^2 & -3\lambda \end{pmatrix} \mathbf{f}(t) + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} w(t),$$

Example: Exponentiated Quadratic Covariance Function

Example (1D EQ covariance function)

- The one-dimensional exponentiated quadratic (EQ or SE) covariance function:
$$k(\tau) = \sigma^2 \exp(-\tau^2/(2l^2))$$
- The spectral density is not a rational function:
$$S(\omega) = \sigma^2 \sqrt{2\pi}\, l \,\exp\left(-\frac{l^2\,\omega^2}{2}\right).$$
- By using the Taylor series we get
$$S(\omega) \approx \frac{\text{constant}}{1 + l^2\,\omega^2 + \cdots + \frac{1}{n!}\, l^{2n}\,\omega^{2n}}$$
- We can factor the result into stable and unstable parts, and further convert into an n-dimensional state space model.

Example: Comparison of Exponentiated Quadratic and Matérn



Inference in Practice

- Conventional GP regression:
  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.

→ ▸ Matern 5/2 animation

→ ▸ EQ animation

Aalto University

Benefits of SDE Representation

- The computational complexity is $O(n)$, where n is the number of measurements.
- The representation can be naturally combined with physical models (leading to LFMs).
- It is straightforward to form integrated GPs, superpositions of GPs and many other linearly transformed GPs.
- Can be extended to non-stationary processes.
- Can be extended to non-Gaussian processes.

Representations of Spatial Gaussian Processes

- Moment representation in terms of mean and covariance function

$$\begin{aligned} \mathbf{m}(\mathbf{x}) &= \mathrm{E}[\mathbf{f}(\mathbf{x})] \\ \mathbf{K}(\mathbf{x}, \mathbf{x}') &= \mathrm{E}[(\mathbf{f}(\mathbf{x}) - \mathbf{m}(\mathbf{x}))\,(\mathbf{f}(\mathbf{x}') - \mathbf{m}(\mathbf{x}'))^T]. \end{aligned}$$

- Spectral representation in terms of spectral density function

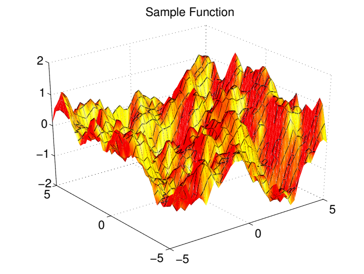
$$\mathbf{S}(\boldsymbol{\omega}_x) = \mathrm{E}[\tilde{\mathbf{f}}(\mathrm{i}\,\boldsymbol{\omega}_x)\,\tilde{\mathbf{f}}^T(-\mathrm{i}\,\boldsymbol{\omega}_x)]$$

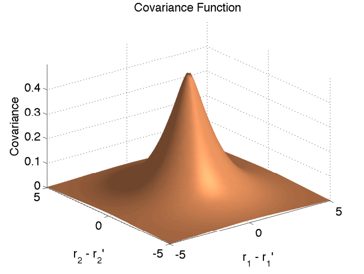
- Representation as stochastic partial differential equation

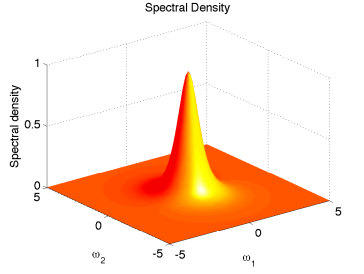
$$\mathcal{A}\,\mathbf{f}(\mathbf{x}) = w(\mathbf{x}),$$

where $\mathcal{A}$ is a linear operator (e.g., matrix of differential operators).

Representations of Spatial Gaussian Processes







Stochastic Partial Differential Equation Representation of Spatial Gaussian Process

- The origins of the Matérn covariance function are in stochastic partial differential equations (SPDEs).
- Consider the following SPDE:

$$\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} + \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} - \lambda^2 f(x_1, x_2) = w(x_1, x_2),$$

  where $w(x_1, x_2)$ is a Gaussian white noise process.
- Because f and w appears linearly in the equation, the classical PDE theory tells that the solution is a linear operation on w.
- Because w is Gaussian, f is a Gaussian process.
- But what are the spectral density and covariance function of f?

Spectral and Covariance Representation of Spatial Gaussian Process

- By taking Fourier transform of the example SPDE we get

$$\tilde{f}(\mathrm{i}\,\omega_1, \mathrm{i}\,\omega_2) = \frac{\tilde{w}(\mathrm{i}\,\omega_1, \mathrm{i}\,\omega_2)}{\omega_1^2 + \omega_2^2 + \lambda^2}.$$

- The spectral density is then

$$S(\omega_1, \omega_2) = \frac{1}{(\omega_1^2 + \omega_2^2 + \lambda^2)^2}$$

- By inverse Fourier transform we get the covariance function

$$k(\mathbf{x}, \mathbf{x}') = \frac{|\mathbf{x} - \mathbf{x}'|}{2\lambda} K_1(\lambda|\mathbf{x} - \mathbf{x}'|)$$

  where K_1 is the modified Bessel function.
- A special case of Matern class covariance functions - so called Whittle covariance function.

Converting Covariance Functions to SPDEs

- Finding an SPDE such that it has a given stationary covariance function $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}')$ can be, in principle, done as follows:
  - Compute the Fourier transform of $k(\mathbf{x})$, i.e., the spectral density $S(\boldsymbol{\omega})$.
  - Find a function $A(\mathrm{i}\,\boldsymbol{\omega})$ such that

$$S(\boldsymbol{\omega}) = A(\mathrm{i}\,\boldsymbol{\omega})\,A(-\mathrm{i}\,\boldsymbol{\omega}).$$

  - Note that because $S(\boldsymbol{\omega})$ is a symmetric and positive function, one (but not maybe the best) choice is the self-adjoint $A(\mathrm{i}\,\boldsymbol{\omega}) = \sqrt{S(\boldsymbol{\omega})}$.
  - Next, form the linear operator corresponding to the function $A(\mathrm{i}\,\boldsymbol{\omega})$:

$$\mathcal{A}_x = \mathcal{F}^{-1}[A(\mathrm{i}\,\boldsymbol{\omega})].$$

  - The SPDE is then given as

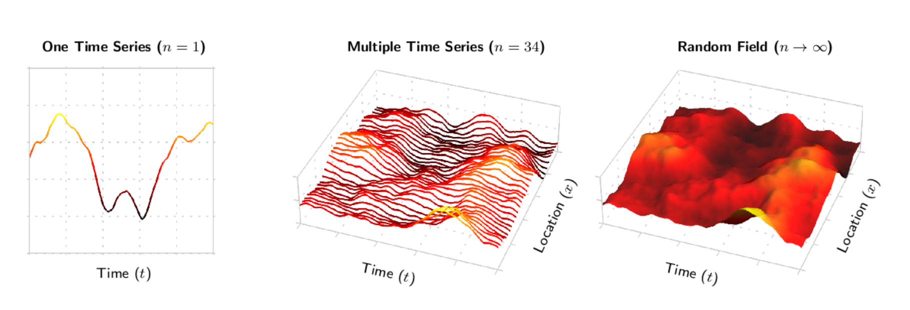
$$\mathcal{A}_x\, f(\mathbf{x}) = w(\mathbf{x}),$$

    where $w(\mathbf{x})$ is a Gaussian white noise process.

Benefits of SPDE Representation

- The SPDE representation allows the use of partial differential equation (PDE) methods to approximate the solutions.
- For large data, PDE methods can be used to form computationally efficient sparse and reduced-rank approximations.
- Finite element method (FEM) leads to sparse approximations of the process.
- Eigenbasis of the Laplace operator leads to reduced-rank approximations.
- SPDEs also allow the combination of GPs with physical models.
- It is also possible to construct non-stationary processes by altering the coefficients of the SPDE.

From Temporal to Spatio-Temporal Processes



Representations of Spatio-Temporal Gaussian Processes

- Moment representation in terms of mean and covariance function

$$\mathbf{m}(\mathbf{x}, t) = \mathrm{E}[\mathbf{f}(\mathbf{x}, t)]$$
$$\mathbf{K}(\mathbf{x}, \mathbf{x}'; t, t') = \mathrm{E}[(\mathbf{f}(\mathbf{x}, t) - \mathbf{m}(\mathbf{x}, t))\,(\mathbf{f}(\mathbf{x}, t) - \mathbf{m}(\mathbf{x}, t'))^T].$$

- Spectral representation in terms of spectral density function

$$\mathbf{S}(\boldsymbol{\omega}_x, \omega_t) = \mathrm{E}[\tilde{\mathbf{f}}(\mathrm{i}\,\boldsymbol{\omega}_x, \mathrm{i}\,\omega_t)\,\tilde{\mathbf{f}}^T(-\mathrm{i}\,\boldsymbol{\omega}_x, -\mathrm{i}\,\omega_t)].$$

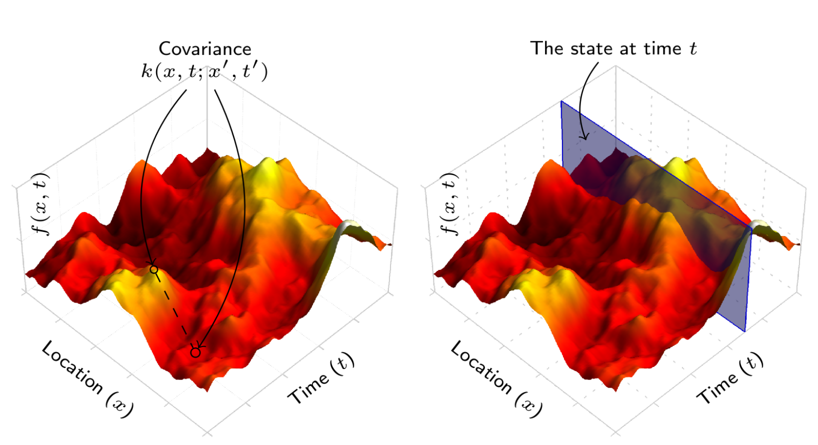
- As infinite-dimensional state space model or stochastic differential equation (SDE):

$$d\mathbf{f}(\mathbf{x}, t) = \mathcal{A}\,\mathbf{f}(\mathbf{x}, t)\,dt + \mathcal{L}\,d\boldsymbol{\beta}(\mathbf{x}, t).$$

  or more informally

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

Spatio-Temporal Gaussian SPDEs



Aalto University

GP Regression as Infinite Linear Model

- Consider the following finite-dimensional linear model

$$\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_0)$$
$$\mathbf{y} = \mathbf{H}\,\mathbf{f} + \mathbf{e},$$

where $\mathbf{f} \in \mathbb{R}^s$, $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{H} \in \mathbb{R}^{n \times s}$, and $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \Sigma)$.

- The posterior distribution is Gaussian with mean and covariance

$$\hat{\mathbf{m}} = \mathbf{K}_0\,\mathbf{H}^T(\mathbf{H}\,\mathbf{K}_0\,\mathbf{H}^T + \Sigma)^{-1}\mathbf{y}$$
$$\hat{\mathbf{K}} = \mathbf{K}_0 - \mathbf{K}_0\,\mathbf{H}^T(\mathbf{H}\,\mathbf{K}_0\,\mathbf{H}^T + \Sigma)^{-1}\mathbf{H}\,\mathbf{K}_0.$$

- Note that Kalman filtering model is an extension to this, where $\mathbf{f}$ can depend on time (but does not need to).

GP Regression as Infinite Linear Model (cont.)

- In the infinite-dimensional limit $\mathbf{f}$ becomes a member of Hilbert space of functions $f(\mathbf{x}) \in \mathcal{H}(\mathbb{R}^d)$.
- The corresponding linear model now becomes

$$\begin{aligned} f(\mathbf{x}) &\sim \mathcal{GP}(0, C_0(\mathbf{x}, \mathbf{x}')) \\ \mathbf{y} &= \mathfrak{H}\, f(\mathbf{x}) + \mathbf{e}, \end{aligned}$$

  where $\mathfrak{H} \;:\; \mathcal{H}(\mathbb{R}^d) \mapsto \mathbb{R}^n$ is a vector of functionals.
- The posterior mean and covariance become

$$\begin{aligned} \hat{m}(\mathbf{x}) &= C_0(\mathbf{x}, \mathbf{x}') \mathfrak{H}^* \left[\mathfrak{H}\, C_0(\mathbf{x}, \mathbf{x}')\, \mathfrak{H}^* + \Sigma \right]^{-1} \mathbf{y} \\ \hat{C}(\mathbf{x}, \mathbf{x}') &= C_0(\mathbf{x}, \mathbf{x}') - C_0(\mathbf{x}, \mathbf{x}')\, \mathfrak{H}^* \left[\mathfrak{H}\, C_0(\mathbf{x}, \mathbf{x}')\, \mathfrak{H}^* + \Sigma \right]^{-1} \\ &\quad \times \mathfrak{H}\, C_0(\mathbf{x}, \mathbf{x}'), \end{aligned}$$

- With $\mathfrak{H}\, f(\mathbf{x}) = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))$ we get GP regression.

State Space Form of Spatio-Temporal Gaussian Processes

- Infinite dimensional generalization of state space model is the stochastic evolution equation

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

  where $\mathcal{A}$ and $\mathcal{L}$ are linear operators in $\mathbf{x}$-variable and $\mathbf{w}(\cdot)$ is a time-space white noise.

- The mild solution to the equation is:

$$\mathbf{f}(\mathbf{x}, t) = \mathfrak{U}(t)\, \mathbf{f}(\mathbf{x}, 0) + \int_0^t \mathfrak{U}(t - s)\, \mathcal{L}\, \mathbf{w}(\mathbf{x}, s)\, ds.$$

- $\mathfrak{U}(t) = \exp(t\, \mathcal{A})$ is the evolution operator – corresponds to propagator in quantum mechanics.

Infinite-Dimensional Kalman Filtering and Smoothing

- The infinite-dimensional Kalman filter and RTS smoother can be used in models of the form:

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

- GP regression is a special case: $\partial \mathbf{f}(\mathbf{x}, t) / \partial t = 0$.
- Weakly equivalent discrete-time model has the form

$$\mathbf{f}_i(\mathbf{x}) = \mathcal{U}_i\, \mathbf{f}_{i-1}(\mathbf{x}) + \mathbf{n}_i(\mathbf{x})$$
$$\mathbf{y}_i = \mathcal{H}_i\, \mathbf{f}_i(\mathbf{x}) + \mathbf{e}_i$$

- If $\mathcal{A}$ and $\mathcal{H}$ are "diagonal" in the sense that they only involve point-wise evaluation in $\mathbf{x}$, we get a finite-dimensional algorithm.
- We can approximate with basis function expansions, Galerkin approximations, FEM, finite-differences, spectral methods, etc.

Conversion of Spatio-Temporal Covariance into Infinite-Dimensional State Space Model

- First compute the spectral density $S(\boldsymbol{\omega}_x, \omega_t)$ by Fourier transforming the covariance function.
- Form rational approximation in variable $\mathrm{i}\,\omega_t$:

$$S(\boldsymbol{\omega}_x, \omega_t) = \frac{q(\mathrm{i}\,\boldsymbol{\omega}_x)}{b_0(\mathrm{i}\,\boldsymbol{\omega}_x) + b_1(\mathrm{i}\,\boldsymbol{\omega}_x)\,(\mathrm{i}\,\omega_t) + \cdots + (\mathrm{i}\,\omega_t)^N}.$$

- Form the corresponding Fourier domain SDE:

$$\frac{\partial^N \tilde{f}(\boldsymbol{\omega}_x, t)}{\partial t^N} + a_{N-1}(\mathrm{i}\,\boldsymbol{\omega}_x)\,\frac{\partial^{N-1} \tilde{f}(\boldsymbol{\omega}_x, t)}{\partial t^{N-1}} + \cdots + a_0(\mathrm{i}\,\boldsymbol{\omega}_x)\,\tilde{f}(\boldsymbol{\omega}_x, t) = \tilde{w}(\boldsymbol{\omega}_x, t).$$

Conversion of Spatio-Temporal Covariance into Infinite-Dimensional State Space Model (cont.)

- By converting this to state space form and by taking spatial inverse Fourier transform, we get stochastic evolution equation

$$\frac{\partial \mathbf{f}(\mathbf{x},t)}{\partial t} = \underbrace{\begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -\mathcal{A}_0 & -\mathcal{A}_1 & \dots & -\mathcal{A}_{N-1} \end{pmatrix}}_{\mathcal{A}} \mathbf{f}(\mathbf{x},t) + \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}}_{\mathcal{L}} w(\mathbf{x},t)$$

where $\mathcal{A}_j$ are pseudo-differential operators.

- We can now use infinite-dimensional Kalman filter and RTS smoother for efficient estimation of the state $\mathbf{f}(\mathbf{x},t)$.

Example: 2D Matérn covariance function

Example (2D Matérn covariance function)

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

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

- The corresponding spectral density is of the form

$$S(\boldsymbol{\omega}_r) = S(\boldsymbol{\omega}_x, \omega_t) \propto \frac{C}{\left(\lambda^2 + ||\boldsymbol{\omega}_x||^2 + \omega_t^2 \right)^{\nu + d/2}}.$$

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

Example: 2D Matérn covariance function (cont.)

Example (2D Matérn covariance function (cont.))

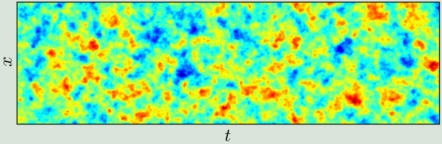
- The denominator roots are $(i\omega_t) = \pm\sqrt{\lambda^2 - ||i\boldsymbol{\omega}_x||^2}$, which gives

$$G(i\boldsymbol{\omega}_x, i\omega_t) = \frac{1}{\left(i\omega_t + \sqrt{\lambda^2 - ||i\boldsymbol{\omega}_x||^2}\right)^{(\nu+d/2)}}.$$

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

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

- Example realization:



Benefits of Stochastic Evolution Equation Representation

- Linear time complexity in time direction
- Easy to combine with physical models (i.e., partial differential equations)
- Linear operations on GPs easy to implement
- Can be extended to non-stationary processes.
- Can be extended to non-Gaussian processes.

Conclusion

- Gaussian processes have different representations:
  - Covariance function.
  - Spectral density.
  - Stochastic (partial) differential equation – a state space model.
- Temporal (single-input) Gaussian processes
  $\Longleftrightarrow$ stochastic differential equations (SDEs) (state space models).
- Spatial (multiple-input) Gaussian processes
  $\Longleftrightarrow$ stochastic partial differential equations (SPDEs).
- Spatio-temporal Gaussian processes
  $\Longleftrightarrow$ stochastic evolution equations (inf.-dim. state space models).
- Kalman filter and RTS smoother are computationally efficient algorithms for Bayesian inference in temporal Gaussian processes.
- Infinite-dimensional Kalman filters and RTS smoothers can be used for efficient inference in spatio-temporal Gaussian process models.

References

- Hartikainen, J. and Särkkä, S. (2010). *Kalman Filtering and Smoothing Solutions to Temporal Gaussian Process Regression Models*. Proc. of MLSP.
- Lindgren, F., Rue, H., and Lindström, J. (2011). *An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach*. JRSS B, 73(4):423–498.
- Särkkä , S. (2011). *Linear operators and stochastic partial differential equations in Gaussian process regression*. Proc. of ICANN.
- Särkkä, S. and Hartikainen, J. (2012). *Infinite-Dimensional Kalman Filtering Approach to Spatio-Temporal Gaussian Process Regression*. AISTATS 2012.
- Särkkä, S., Solin, A. and Hartikainen, J. (2013). *Spatio-Temporal Learning via Infinite-Dimensional Bayesian Filtering and Smoothing*. IEEE Signal Processing Magazine (to appear).
- Solin, A., Särkkä, S. (2013). *Hilbert Space Methods for Reduced-Rank Gaussian Process Regression*. Submitted.
- S. Särkkä (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press. (Forthcoming).