Spatial temporal Gaussian process

포항공과대학교 산업경영공학과

Stochastic Systems Lab

Jongwon Kim

Contents

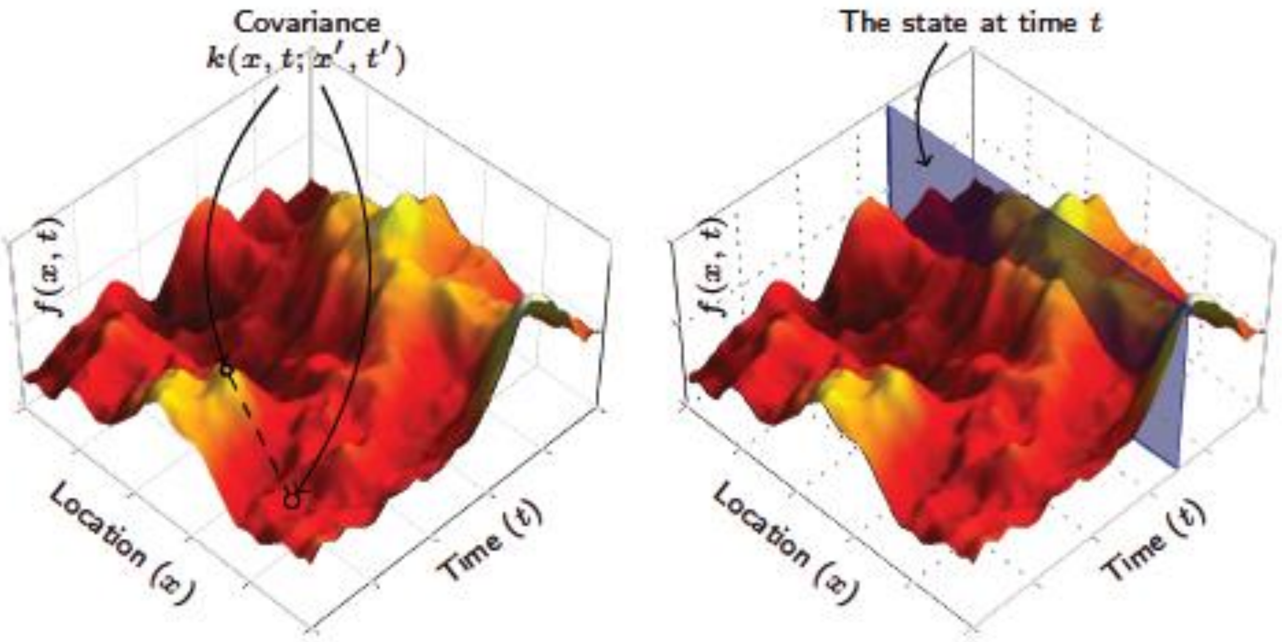
Problem statement

What is stochastic process?

Stochastic process $\{X_i, i \in I\}$ is a family of random variable X_t that is indexed by some mathematical set I.

If the index set I can be represented one-dimension, we call the stochastic process random process.

If the index set I can be represented multi-dimension, we call the stochastic process random field.



(a) Covariance function view (b) State space model view

Särkkä, Simo, Arno Solin, and Jouni Hartikainen. "Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering." *IEEE Signal Processing Magazine* 30.4 (2013): 51-61.

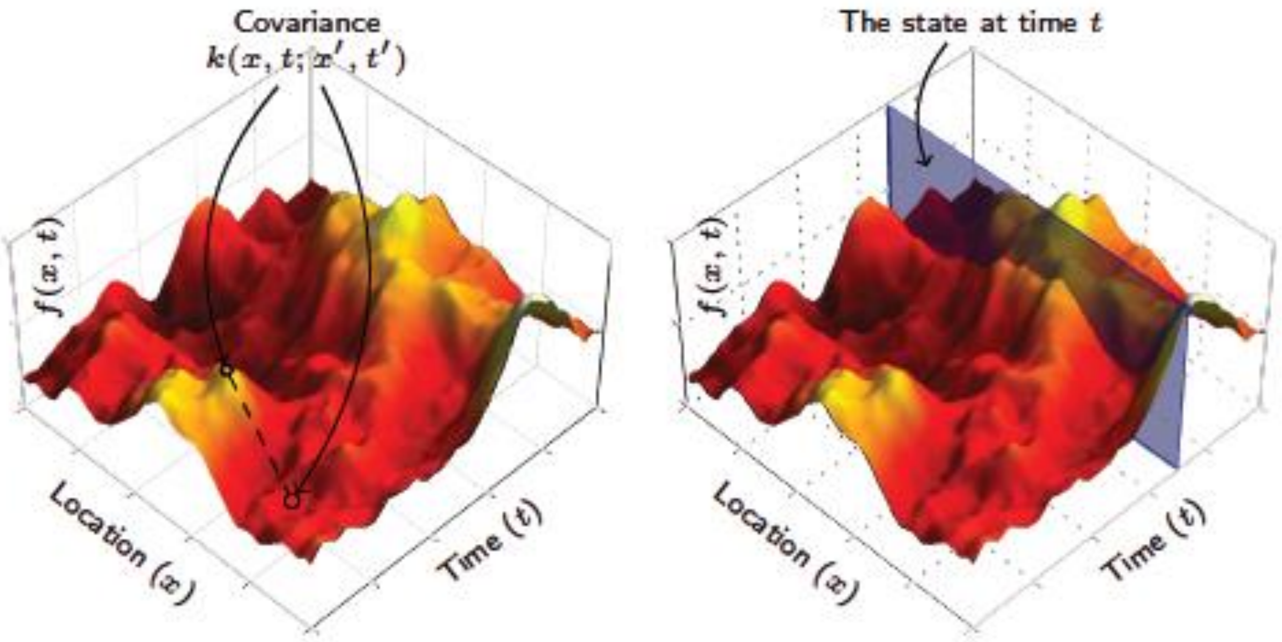
Problem statement

What is stochastic process?

Gaussian process $\{f(i) = X_i, i \in I\}$ is a stochastic process in which all the finite-dimensional distributions are multivariate Gaussian distributions for any finite choice of variables.

If we add Markov property for one dimensional, we call that Gaussian process to Gauss-Markov process or Gauss Markov random field

If the dimension is time, we refer to it as a Spatio-temporal Gaussian process.



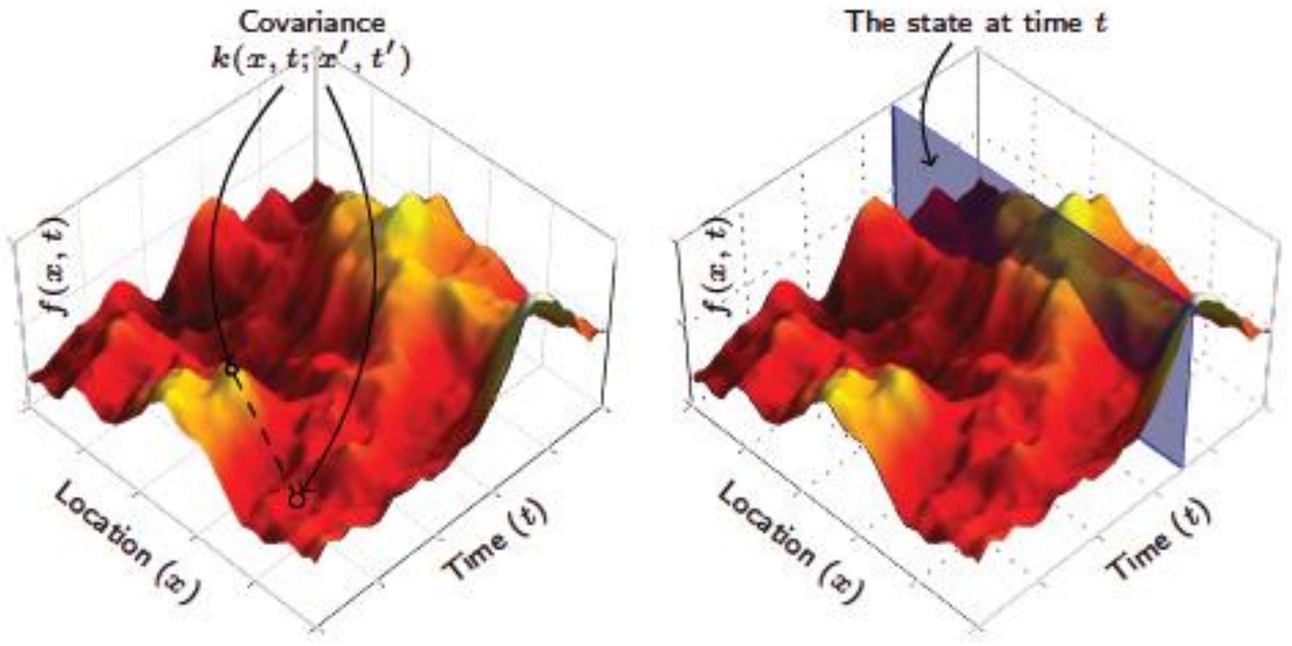
(a) Covariance function view (b) State space model view

Särkkä, Simo, Arno Solin, and Jouni Hartikainen. "Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering." *IEEE Signal Processing Magazine* 30.4 (2013): 51-61.

Problem statement

What is stochastic process?

We can train the Spatio-temporal Gaussian process **with linear scalability** over the number of data.



(a) Covariance function view (b) State space model view

Särkkä, Simo, Arno Solin, and Jouni Hartikainen. "Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering." *IEEE Signal Processing Magazine* 30.4 (2013): 51-61.

Spatial temporal Gaussian process

Markov property

We refer to $f(x)$ as f_x.

For Markov property, we derive the random process as

$$\begin{aligned} &\mathrm{f}_t|\mathrm{f}_t, \mathrm{f}_{t-1}, \dots, \mathrm{f}_1 \\ &= \mathrm{f}_t|\mathrm{f}_t, \mathrm{f}_{t-1}, \dots, \mathrm{f}_{t-p} \end{aligned}$$

In this presentation, for simplicity, we only deal with the linear form.

$\mathrm{f}_t + \alpha_1 \mathrm{f}_{t-1} + \cdots + \alpha_p \mathrm{f}_{t-p} = Z_t$ where Z_t is white Gaussian noise process and X_t has kernel function k.

For discrete index set I, we can give Markov property as shown above.

Spatial temporal Gaussian process

Markov property

For continuous index set, we can replace the time-lag by differentiation

We can show this by checking equivalent between sampling of continuous Spatio-temporal Gaussian process of and the discrete.

$\mathrm{a}_0 \mathrm{f}_t + \alpha_1 \mathrm{f}_t^1 + \cdots + \alpha_p \mathrm{f}_t^p = Z_t$ where $\mathrm{f}_{t'}^{(k)} = \frac{d^k f_t}{dt^k}|_{t=t'}$ is k-th differentiation of f_t which is Gaussian process with some kernel function k and Z_t is a white noise process.

How to solve

Connect GP and SDE

To analyze the X_t,

1) From fixed kernel k, Convert GP with k into the state space model with the autoregressive coefficients $a_i\ (i = 0, \dots, p)$.

GP with k with parameter θ

$\mathrm{a}_0(\theta)\mathrm{f}_t + \alpha_1(\theta)\mathrm{f}_t^{(1)} + \cdots + \alpha_p(\theta)\mathrm{f}_t^{(p)} = Z_t$ where $\mathrm{f}_t^{(d)}$ is d-th differentiation of f which is Gaussian process with some kernel function k and Z_t is a white noise process.

How to solve

Connect GP and SDE

$a_0 f_t + \alpha_1 f_t^{(1)} + \cdots + \alpha_p f_t^{(p)} = Z_t$ where $f_t^{(d)}$ is d-th differentiation of f which is Gaussian process with some kernel function k and Z_t is a white noise process.

To analyze the X_t,

1) From fixed kernel k, Convert GP with k into the state space model with the autoregressive coefficients a_i $(i = 0, \dots, p)$.

For stationary kernel function k, define new function $C(\Delta t) = k(t, t + \Delta t)$

From the Wiener-Khinchin theorem: the function C is the Fourier analysis function of power spectrum $S(w)$. .

Let $\tilde{f}$ denote the fourier analysis function of f.

$$S(w) = \left|\tilde{f}(iw)\right|^2 \text{ (by the definition of power spectrum for } f_t\text{)}$$

$$= \int_{-\infty}^{\infty} C(t) e^{-2\pi i w t} dt \text{ (by the Wiener-Khinchin theorem) where}$$

How to solve

Connect GP and SDE

To analyze the f_t,

1) From fixed kernel k, Convert GP with k into the state space model with the autoregressive coefficients a_i $(i = 0, \ldots, p)$.

$\mathrm{a}_0\mathrm{f}_t + \alpha_1 \mathrm{f}_t^{(1)} + \cdots + \alpha_p \mathrm{f}_t^{(p)} = Z_t$ we cam derive the fourier analysis of f as follows:

$$\tilde{\mathrm{f}}(\mathrm{iw}) = \frac{1}{\mathrm{a}_0 + \alpha_1(\mathrm{iw}) + \cdots + \alpha_p(\mathrm{iw})^{\mathrm{p}}}\tilde{\mathrm{z}}(\mathrm{iw}) = \mathrm{G}(\mathrm{iw})\tilde{\mathrm{z}}(\mathrm{iw})$$

$\mathrm{S}(\mathrm{w}) = \left|\tilde{\mathrm{f}}(iw)\right|^2 = |\mathrm{G}(\mathrm{iw})\tilde{\mathrm{z}}(\mathrm{iw})\,|^2 = \mathrm{G}(\mathrm{iw})\mathrm{q}_\mathrm{c}\mathrm{G}(-\mathrm{iw})$ where $\mathrm{q}_\mathrm{c} = |\tilde{\mathrm{z}}(\mathrm{iw})|^2$ then $\mathrm{S}(\mathrm{w}) = \frac{\text{constant}}{\text{polynomial with } \mathrm{w}^2}$

From kernel we can calculate the rational function form $\mathrm{S}(\mathrm{w}) = \int_{-\infty}^{\infty} \mathrm{k}(\mathrm{t})\mathrm{e}^{-2\pi\mathrm{iwt}}\mathrm{dt}$ or approximate it such a form, for example, Taylor series expansions or Pad´e approximants.

How to solve

Connect GP and SDE

To analyze the f_t,

1) From fixed kernel k, Convert GP with k into the state space model with the autoregressive coefficients $a_i \ (i = 0, \dots, p)$.

$$\mathrm{S(w)} \ = \mathrm{G(iw)q_cG(-iw)}$$

Procedure for finding a stable transfer function $\mathrm{G(iw)}$ is called spectral factorization.

We required the transfer function should be stable which means that all of its poles (zeros of denominator) are in upper half of complex plane.

We also want the transfer function to be minimum phase, which happens when the zeros of the numerators are also in the half of complex plance.

How to solve

Connect GP and SDE

To analyze the f_t,

1) From fixed kernel k, Convert GP with k into the state space model with the autoregressive coefficients $a_i \ (i = 0, \dots, p)$.

More, generally, we can approximate $\mathrm{G(iw)} \ = \frac{\mathrm{a_0} + \alpha_1 \mathrm{(iw)} + \cdots + \alpha_p \mathrm{(iw)^q}}{\mathrm{a_0} + \alpha_1 \mathrm{(iw)} + \cdots + \alpha_p \mathrm{(iw)^p}}$ from the kernel function.

Then, we get the $\mathrm{S(w)} = \frac{\text{qth polynomial with } \mathrm{w}^2}{\text{pth polynomial with } \mathrm{w}^2} \ (\mathrm{p} > \mathrm{q})$

In control theory [3], there exists a number of methods to convert a transfer function of form into an equivalent state space model.

[3] Ljung, Lennart. *Control theory: multivariable and nonlinear methods*. Taylor & Francis, 2000.

How to solve

Filtering and smoothing

To analyze the f_t,

2) Solve the state-space model using filtering and smoothing with observation x_i, y_i

$$y = f(x) + \epsilon, \epsilon \sim N(0, \sigma^2_{noise})$$

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

$$y_k = \mathbf{H}\,\mathbf{f}(t_k) + \varepsilon_k,$$

$\mathrm{a}_0(\theta)\mathrm{f}_t + \alpha_1(\theta)\mathrm{f}_t^{(1)} + \cdots + \alpha_p(\theta)\mathrm{f}_t^{(p)} = Z_t$ where $\mathrm{f}_t^{(d)}$ is d-th differentiation of f which is Gaussian process with some kernel function k and Z_t is a white noise process.

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

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

STOCHASTIC SYSTEMS LAB.

How to solve

Filtering and smoothing

To analyze the f_t,

2) Solve the state-space model using filtering and smoothing with observation x_i, y_i

$$y = f(x) + \epsilon, \epsilon \sim N(0, \sigma_{noise}^2)$$

$$\frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} = \mathbf{F}\,\mathbf{f}(t) + \mathbf{L}\,\mathbf{w}(t),$$

$$y_k = \mathbf{H}\,\mathbf{f}(t_k) + \varepsilon_k,$$

$$\frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & -a_1 & \dots & -a_{m-1} \end{pmatrix} \mathbf{f}(t) + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} w(t).$$

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

- Kalman filter can be used for computing all the Gaussian filtering distributions:

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k).$$

- Rauch–Tung–Striebel smoother then computes the corresponding smoothing distributions

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:T}) = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^s, \mathbf{P}_k^s).$$

How to solve

Filtering and smoothing

To analyze the f_t,

2) Solve the state-space model

We can represent the defined linear differential equation $\mathrm{a}_0\mathrm{f}_t + \alpha_1\mathrm{f}_t^{(1)} + \cdots + \alpha_p\mathrm{f}_t^{(p)} = Z_t$ as SDE form follows:

To analyze the f_t,
2) Solve the state-space model
We can represent the defined linear differential equation $\mathrm{a}_0\mathrm{f}_t + \alpha_1\mathrm{f}_t^{(1)} + \cdots + \alpha_p\mathrm{f}_t^{(p)} = Z_t$ as SDE form follows:
Our state-space model can be represented linear stochastic differential equation.
Solve the state−space model with Kalman filtering and Rauch–Tung–Striebel smoother

$$\frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{m-1} \end{pmatrix} \mathbf{f}(t) + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} w(t).$$

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

Our state−space model can be represented linear stochastic differential equation.

Solve the state−space model with Kalman filtering and Rauch–Tung–Striebel smoother

To analyze the f_t,

2) Solve the state-space model using filtering and smoothing with observation x_i, y_i

$$y = f(x) + \epsilon, \epsilon \sim N(0, \sigma_{noise}^2)$$

- Kalman filter can be used for computing all the Gaussian filtering distributions:

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k).$$

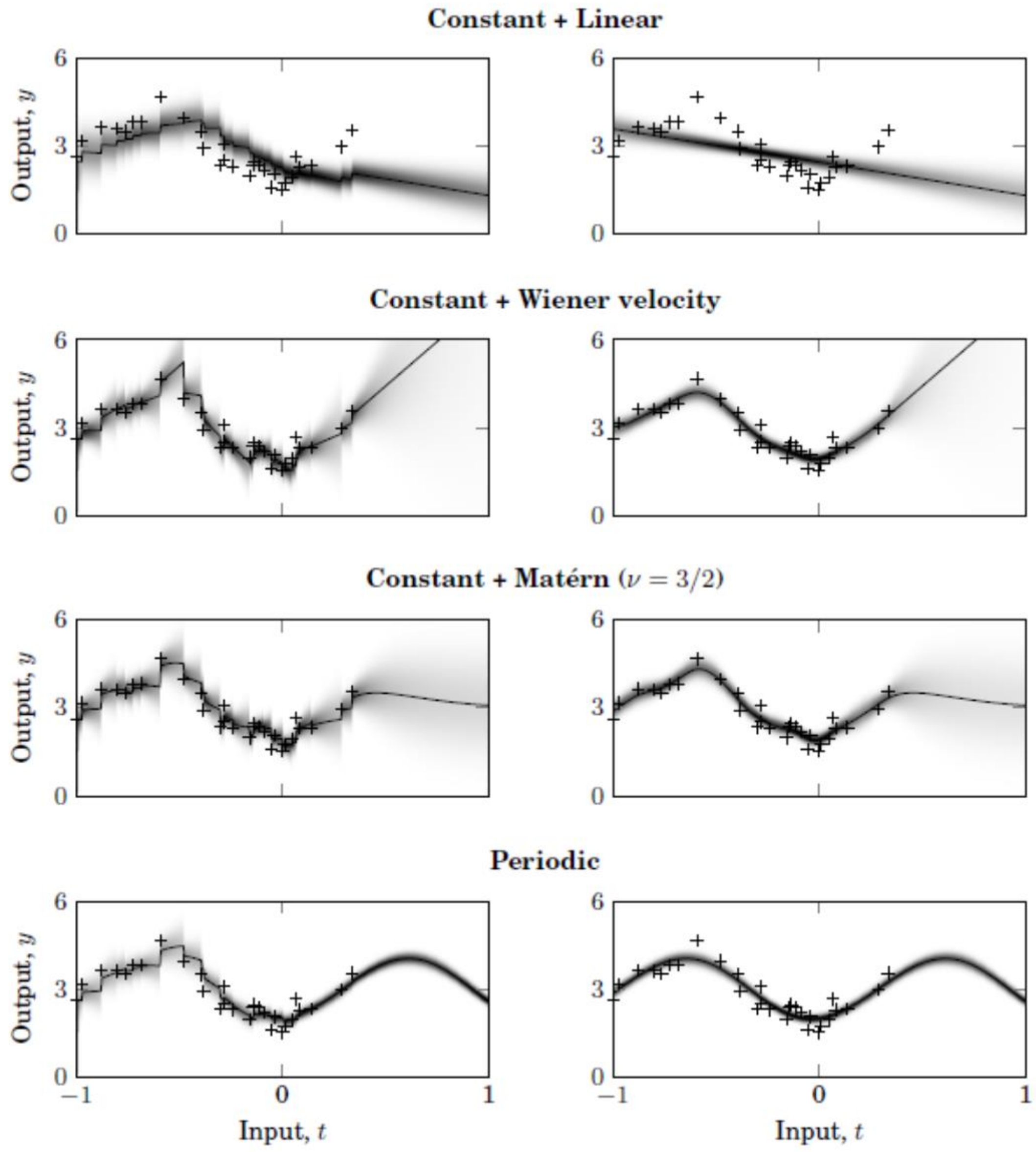
- Rauch–Tung–Striebel smoother then computes the corresponding smoothing distributions

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:T}) = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^s, \mathbf{P}_k^s).$$

How to solve

Filtering and smoothing

Filtering (left) and smoothing (right) solutions to Gaussian process regression problems with different prior models.



How to solve

Filtering and smoothing

To analyze the f_t,

3) Calculate the likelihood and estimate the kernel parameter.

From smoothing, we get the likelihood to estimate the kernel parameter.

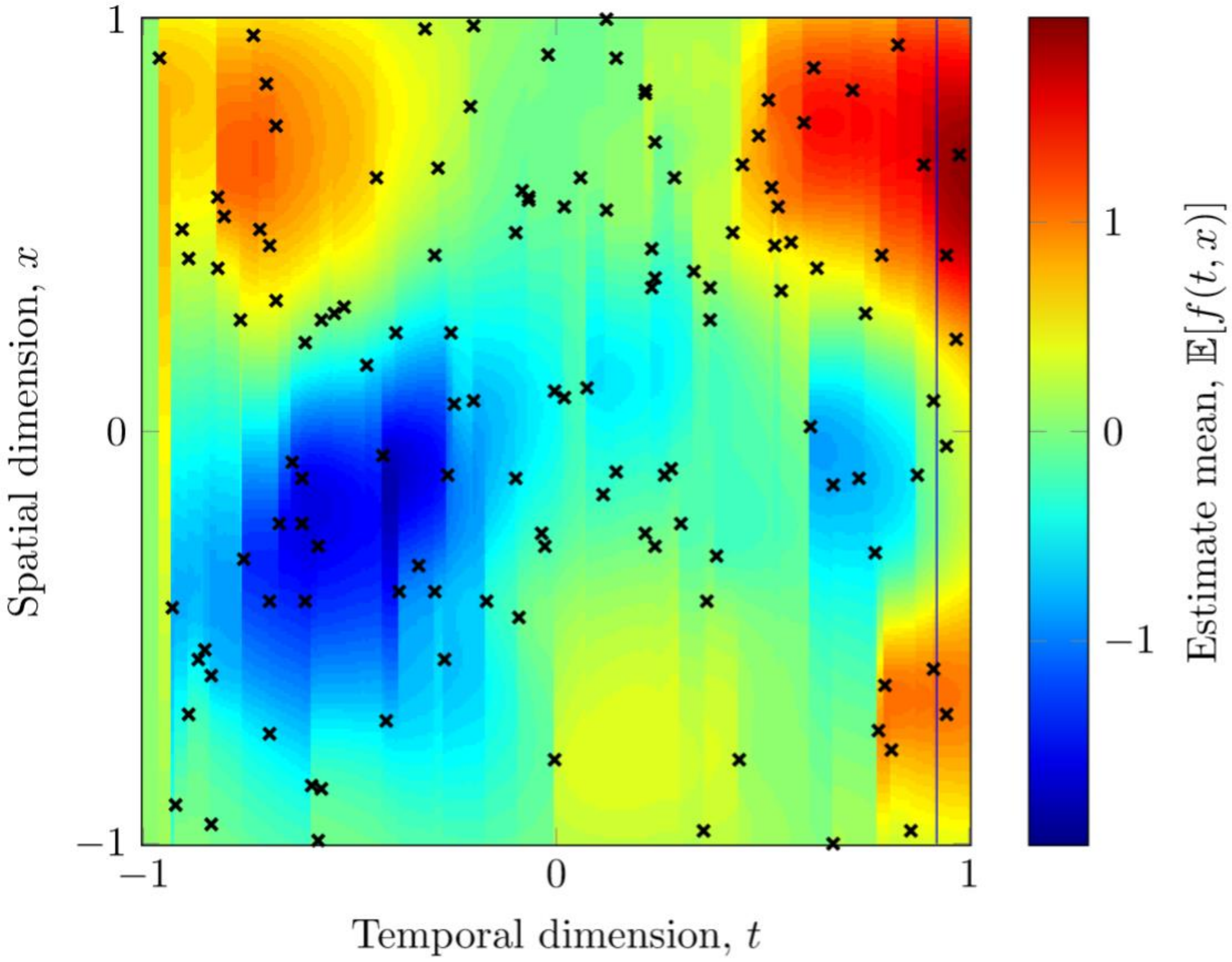
- Kalman filter can be used for computing all the Gaussian filtering distributions:

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k).$$

- Rauch–Tung–Striebel smoother then computes the corresponding smoothing distributions

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:T}) = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k^s, \mathbf{P}_k^s).$$

Spatio-temporal GP regression



Appendix

From discrete-time to continuous-time

We reformulate the

$$\frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} = \mathbf{F}\,\mathbf{f}(t) + \mathbf{L}\,\mathbf{w}(t),$$
$$y_k = \mathbf{H}\,\mathbf{f}(t_k) + \varepsilon_k,$$

$$\frac{df(t)}{dt} = Af(t) + Lz(t)$$
$$z(t) \text{ has zero mean and } E[z(t)z(t+a)] = Q_c\delta(a)$$
$$y_k = Hf(t_k) + \epsilon_k \qquad \epsilon_k \sim N(0, \sigma^2_{noise})$$

A_k is the discrete-time transition matrix and Q_K is the process noise covariance matrix.

$$\begin{aligned} \mathbf{f}_k &= \mathbf{A}_{k-1}\,\mathbf{f}_{k-1} + \mathbf{q}_{k-1}, && \text{where} \quad \mathbf{q}_{k-1} \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}_{k-1}), \\ y_k &= \mathbf{H}\,\mathbf{f}_k + \varepsilon_k, && \text{where} \quad \varepsilon_k \sim \mathrm{N}(0, \sigma^2_{\text{noise}}). \end{aligned} \tag{3.7}$$

$$\mathbf{A}_k = \mathbf{\Phi}(\Delta t_k), \tag{3.8}$$

$$\mathbf{Q}_k = \int_0^{\Delta t_k} \mathbf{\Phi}(\Delta t_k - \tau)\,\mathbf{L}\,\mathbf{Q}_\mathrm{c}\,\mathbf{L}^\mathsf{T}\,\mathbf{\Phi}(\Delta t_k - \tau)^\mathsf{T}\,\mathrm{d}\tau, \tag{3.9}$$

$$\Delta t_k = t_{k+1} - t_k$$

$$\Phi(\tau) = \exp(A\tau)$$

Appendix

Proof)

$$\frac{df(t)}{dt} - Af(t) = Lz(t)$$

$$\exp(-At)(\frac{df(t)}{dt} - Af(t)) = \exp(-At)Lz(t)$$

$$\frac{d}{dt}\left(\exp(-At)\, f(t)\right) = \exp(-At)\, Lz(t)$$

$$\exp(-A(t+\Delta t_k))\, f(t+\Delta t_k) - \exp(-At)\, f(t) = \int_t^{\mathrm{t}+\Delta t_k} \exp(-A\tau)\, Lz(\tau)d\tau$$

$$f(t+\Delta t_k) = \exp(A\Delta t_k)\, f(t) + \int_0^{\Delta t_k} \exp(A(\Delta t_k - \tau))\, Lz(\tau)d\tau$$

Appendix

Let's now consider the spatio-temporal model, $\mathrm{f}(\mathrm{x},\mathrm{t})$ with kernel $\mathrm{k}(\mathrm{x},\mathrm{t},\mathrm{x}',\mathrm{t}')$

We can get the (power) spectral density $\mathrm{S}(\mathrm{w_x},\mathrm{w_t}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{C}(\mathrm{x},\mathrm{t})e^{-2\pi \mathrm{i} \mathrm{w_t} \mathrm{t}}e^{-2\pi \mathrm{i} \mathrm{w_x} \mathrm{x}}\mathrm{dtdx}$

For fixed $\mathrm{w_x}$, we can get the stable transfer function for $\mathrm{w_t}$ and

$$S(w_x, w_t) = G(\mathrm{w_x}, \mathrm{iw_t})$$

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

Appendix

From one-dimensional to spatio-temporal model

$$a_0(iw_x)\tilde{f}_t(iw_x) + \alpha_1(iw_x)\tilde{f}_t^{(1)}(iw_x) + \cdots + \alpha_p(iw_x)\tilde{f}_t^{(p)}(iw_x) = \tilde{z}_t(iw_x)$$

Where z_t is a noise-process with covariance Q_c

Let's reformulate the differential equation into SDE form as left follows.

Next, we compute its inverse Fourier transform.

Where $A_{\mathrm{i}} = (a_{\mathrm{i}}\widetilde{(\mathrm{iw_x})})$ is the pseudo-differential operators.

$$\frac{\partial \tilde{\mathbf{f}}(i\,\boldsymbol{\omega}_x, t)}{\partial t} = \mathbf{A}(i\,\omega_x)\,\tilde{\mathbf{f}}(i\,\boldsymbol{\omega}_x, t) + \mathbf{L}\,\tilde{w}(i\,\boldsymbol{\omega}_x, t), \qquad (28)$$

where

$$\mathbf{A}(i\,\omega_x) = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0(i\,\boldsymbol{\omega}_x) & -a_1(i\,\boldsymbol{\omega}_x) & \dots & -a_{n-1}(i\,\boldsymbol{\omega}_x) \end{pmatrix}, \qquad (29)$$

$$\frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial t} = \mathcal{A}\,\mathbf{f}(\mathbf{x}, t) + \mathbf{L}\,w(\mathbf{x}, t), \qquad (30)$$

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

$$\mathcal{A} = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -\mathcal{A}_0 & -\mathcal{A}_1 & \dots & -\mathcal{A}_{n-1} \end{pmatrix}. \qquad (31)$$

Appendix

Recent methods are summarized in this paper [5]

[5] Wilkinson, William J., Simo Särkkä, and Arno Solin. "Bayes-Newton Methods for Approximate Bayesian Inference with PSD Guarantees." *arXiv preprint arXiv:2111.01721* (2021).

References

We refer to two papers:

Solin, Arno. "Stochastic differential equation methods for spatio-temporal Gaussian process regression." (2016).

Särkkä, Simo, Arno Solin, and Jouni Hartikainen. "Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering." *IEEE Signal Processing Magazine* 30.4 (2013): 51-61.

Thank you

Problem statement

- **Notation**

$y \in R$: Output data

$x \in \Omega$: Input data

$D = \{(x_i, y_i) | i = 1, \ldots, n\}$: Data set of n observation,

$f : \Omega \to R$ function. We refer to $f(x)$ as f_x.

Vectors are column vectors by default

Capital letters refer to matrix

- $X = (x_1, x_2, \ldots, x_n)^T$ $where\ x_i \in \Omega$