Non-linear Kalman filtering and smoothing based inference in non-linear latent force models

Simo Särkkä

Aalto University, Espoo, Finland

June 13th, 2013

Contents

The Basic Idea of State-Space Representation

- Assume that our latent force model is of the form

$$\frac{dx_f(t)}{dt} = g(x_f(t)) + u(t),$$

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

- We measure the system at discrete instants of time:

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

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

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

- 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 $\mathbf{x} = (x_f, u)$, we get a two-dimensional SDE

$$\frac{d\mathbf{x}}{dt} = \underbrace{\begin{pmatrix} g(x_1(t)) + x_2(t) \\ -\lambda\, x_2(t) \end{pmatrix}}_{\mathbf{f}(\mathbf{x})} + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{\mathbf{L}} w(t)$$

- We can now rewrite the measurement model as

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

- Thus the result is a model of the generic form

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

- This model can now be efficiently tackled with non-linear Kalman filtering and smoothing.

What is a stochastic differential equation (SDE)?

- At first, we have an ordinary differential equation (ODE):

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t).$$

- Then we add white noise to the right hand side:

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

- Generalize a bit by adding a multiplier matrix on the right:

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

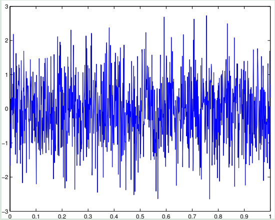
- Now we have a stochastic differential equation (SDE).
- $\mathbf{f}(\mathbf{x}, t)$ is the drift function and $\mathbf{L}(\mathbf{x}, t)$ is the dispersion matrix.

White noise

White noise

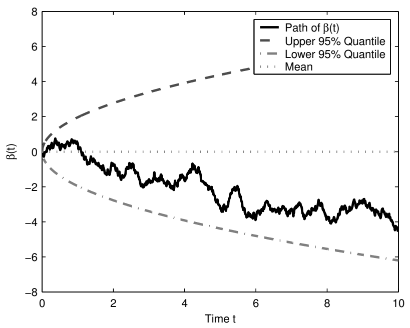
1. $\mathbf{w}(t_1)$ and $\mathbf{w}(t_2)$ are independent if $t_1 \neq t_2$.
2. $t \mapsto \mathbf{w}(t)$ is a Gaussian process with the mean and covariance:

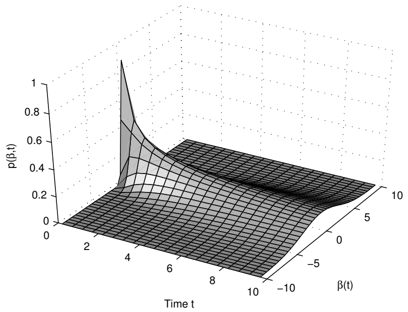
$$\mathrm{E}[\mathbf{w}(t)] = \mathbf{0}$$
$$\mathrm{E}[\mathbf{w}(t)\,\mathbf{w}^{\mathsf{T}}(s)] = \delta(t-s)\,\mathbf{Q}.$$



- $\mathbf{Q}$ is the spectral density of the process.
- The sample path $t \mapsto \mathbf{w}(t)$ is discontinuous almost everywhere.
- White noise is unbounded and it takes arbitrarily large positive and negative values at any finite interval.

What does a solution of SDE look like?



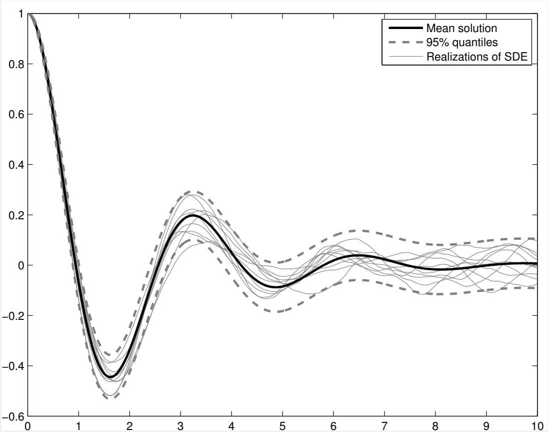


- *Left*: Path of a Brownian motion which is solution to stochastic differential equation

$$\frac{dx}{dt} = w(t)$$

- *Right*: Evolution of probability density of Brownian motion.

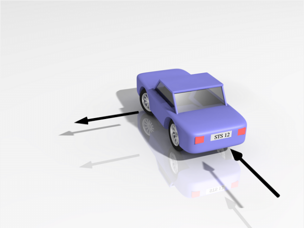
What does a solution of SDE look like? (cont.)



Paths of stochastic spring model

$$\frac{d^2x(t)}{dt^2} + \gamma \frac{dx(t)}{dt} + \nu^2 x(t) = w(t).$$

Example: State Space Model for a Car [1/2]



- The dynamics of the car in 2d (x_1, x_2) are given by the Newton's law:

$$\mathbf{f}(t) = m\,\mathbf{a}(t),$$

where $\mathbf{a}(t)$ is the acceleration, m is the mass of the car, and $\mathbf{f}(t)$ is a vector of (unknown) forces acting the car.

- We shall now model $\mathbf{f}(t)/m$ as a 2-dimensional white noise process:

$$\begin{aligned} d^2 x_1/dt^2 &= w_1(t) \\ d^2 x_2/dt^2 &= w_2(t). \end{aligned}$$

Example: State Space Model for a Car [2/2]

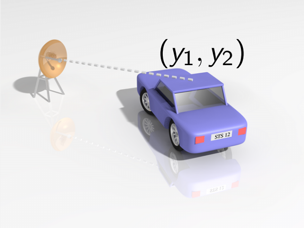
- If we define $x_3(t) = dx_1/dt$, $x_4(t) = dx_2/dt$, then the model can be written as a first order system of differential equations:

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{\mathbf{A}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}}_{\mathbf{L}} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}.$$

- In shorter matrix form:

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

Measurement Model for a Car



- Assume that the position of the car (x_1, x_2) is measured and the measurements are corrupted by Gaussian measurement noise $e_{1,k}, e_{2,k}$:

$$
\begin{aligned}
y_{1,k} &= x_1(t_k) + e_{1,k} \\
y_{2,k} &= x_2(t_k) + e_{2,k}.
\end{aligned}
$$

- The measurement model can be now written as

$$
\mathbf{y}_k = \mathbf{H}\,\mathbf{x}(t_k) + \mathbf{e}_k, \qquad \mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}
$$

Model for Car Tracking

- The dynamic and measurement models of the car now form a linear Gaussian state-space model:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\,\mathbf{x} + \mathbf{L}\,\mathbf{w}$$
$$\mathbf{y}_k = \mathbf{H}\,\mathbf{x}(t_k) + \mathbf{r}_k,$$

- In this case it is possible to solve the transition density explicitly:

$$p(\mathbf{x}(t_k) \,|\, \mathbf{x}(t_{k-1})) = \mathrm{N}(\mathbf{x}(t_k) \,|\, \mathbf{\Psi}_k\,\mathbf{x}(t_{k-1}), \mathbf{W}_k)$$

  where $\mathbf{\Psi}_k$ and $\mathbf{W}_k$ can be expressed in terms of the matrix exponential function.

- Thus we can actually write the model as

$$\mathbf{x}_k = \mathbf{\Psi}_k\,\mathbf{x}_{k-1} + \mathbf{q}_k$$
$$\mathbf{y}_k = \mathbf{H}\,\mathbf{x}_k + \mathbf{r}_k,$$

  where $\mathbf{q}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{W}_k)$.

Latent Force Model for a Car

- We could also start from

$$\begin{aligned} d^2 x_1 / dt^2 &= u(t) \\ d^2 x_2 / dt^2 &= v(t). \end{aligned}$$

  where u and v are, say, Matern 3/2 processes.
- Thus we have, e.g.:

$$\frac{d\mathbf{u}(t)}{dt} = \begin{pmatrix} 0 & 1 \\ -\lambda^2 & -2\lambda \end{pmatrix} \mathbf{u}(t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w_u(t), \quad j = 1, 2$$

  where $\mathbf{u}(t) = (u(t), du(t)/\,dt)$.

Latent Force Model for a Car (cont.)

- Now we get

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ u_1 \\ v_1 \\ u_2 \\ v_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -\lambda^2 & 0 & -2\lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & -\lambda^2 & 0 & -2\lambda \end{pmatrix}}_{\mathbf{A}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ u_1 \\ v_1 \\ u_2 \\ v_2 \end{pmatrix} + \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{L}} \begin{pmatrix} w_1 \\ w_2 \\ w_u \\ w_v \end{pmatrix}$$

$$\mathbf{y}_k = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ u_1 \\ u_2 \\ v_1 \\ v_2 \end{pmatrix} + \mathbf{e}_k,$$

- But this is just a linear Gaussian state-space model:

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

$$\mathbf{y}_k = \mathbf{H}\,\mathbf{x}(t_k) + \mathbf{r}_k,$$

Mathematical Meaning and Notation of SDE [1/2]

- What is Itô stochastic calculus then?
- Let's take a look at the scalar equation

$$\frac{dx(t)}{dt} = f(x(t)) + L(x(t))\,w(t).$$

- Integrating from s to t gives

$$x(t) - x(s) = \int_s^t f(x(t))\,dt + \int_s^t L(x(t))\,w(t)\,dt.$$

- White noise is unbounded and discontinuous almost everywhere – the second integral cannot be defined as Riemann, Stieltjes, or Lebesgue integral!

Aalto University

Mathematical Meaning and Notation of SDE [2/2]

- Itô's idea: define $d\beta(t) = w(t)\,dt$, where $\beta(t)$ is the Wiener/Brownian process:

$$x(t) - x(s) = \int_s^t f(x(t))\,dt + \int_s^t L(x(t))\,d\beta(t).$$

- Commonly used shorthand notation for the above:

$$dx(t) = f(x(t))\,dt + L(x(t))\,d\beta(t).$$

- In stochastics literature you see this in form:

$$dX_t(\omega) = f(X_t(\omega))\,dt + L(X_t(\omega))\,d\beta_t(\omega).$$

Itô Integral

- The Itô integral is defined as the limit of the expression

$$\begin{aligned}\int_s^t L(x(t))\, d\beta(t) = {} & L(x(t_1))\,[\beta(t_2) - \beta(t_1)] \\ & + L(x(t_2))\,[\beta(t_3) - \beta(t_2)] \\ & + \ldots \\ & + L(x(t_{n-1}))\,[\beta(t_n) - \beta(t_{n-1})]\end{aligned}$$

- The key issue is that b is evaluated at the beginning of interval, that is, we have $L(x(t_1))\,[\beta(t_2) - \beta(t_1)]$ instead of, say, $L(x(t_2))\,[\beta(t_2) - \beta(t_1)]$.
- In Riemann, Stieltjes, or Lebesgue integral the result should be independent of the evaluation point.
- The resulting calculus is called Itô calculus or the stochastic calculus.

Stratonovich Integral and SDEs

- If L is evaluated in the middle of the interval $t_i^* = (t_i + t_{i+1})/2$, we get Stratonovich integral and Stratonovich calculus:

$$\begin{aligned}\int_s^t L(x(t)) \circ d\beta(t) = {} & L(x(t_1^*)) \, [\beta(t_2) - \beta(t_1)] \\ & + \dots \\ & + L(x(t_{n-1}^*)) \, [\beta(t_n) - \beta(t_{n-1})]\end{aligned}$$

- The corresponding Stratonovich SDE is then often denoted as

$$dx(t) = f(x(t)) \, dt + L(x(t)) \circ d\beta(t).$$

- If L is constant (or depends only on time), then $L \circ d\beta(t) = L \, d\beta(t)$, i.e., the Itô and Stratonovich SDEs are equivalent.
- When deriving the properties, means and covariance equations, distributions and such, we need to know if we have Itô or Stratonovich SDEs.

What kind of solutions do SDEs have?

- Path of solution: Draw random path $\mathbf{w}(t)$ (or $\beta(t)$) and solve the equation using it as the input.
  - Monte Carlo simulation of SDE solutions.
  - Used in particle filtering and smoothing methods.
- Distribution of solution: Given many random $\mathbf{w}(t)$'s, what is the distribution of the state $p(\mathbf{x}(t))$?
  - Solution is given by the Fokker-Planck-Kolmogorov PDE.
  - Used in grid based and basis function methods (FEM, BEM).
- Moments: What are the mean and covariance of $\mathbf{x}(t)$?
  - Ordinary differential equations for the mean and covariance.
  - Used in non-linear Kalman (Gaussian) filters and smoothers.

Markov Property and Transition Density

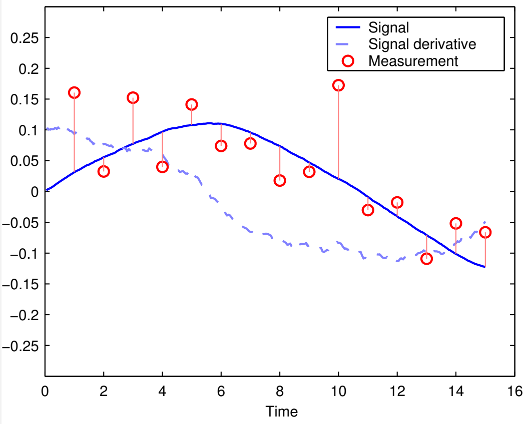
- Itô's stochastic differential equations are Markovian in the sense that

$$p(\mathbf{x}(t_k) \mid \{\mathbf{x}(\tau) : t_0 \le \tau \le t_{k-1}\}) = p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))$$

- This follows from the fact that Itô integrals over Brownian motion are Martingales.
- Due to Markovianity the transition densities $p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))$ characterize the probability law of the process completely.
- The transition density is the Green's function of the Fokker–Planck PDE – and thus intractable in most cases.

Continuous-Discrete State Estimation Problem

- Estimate the unobserved continuous-time signal (= state) from noisy discrete-time measurements



Mathematical Problem Formulation

- Mathematical model is (the special case considered here):

$$d\mathbf{x} = \mathbf{f}(\mathbf{x})\,dt + \mathbf{L}\,d\boldsymbol{\beta}(t)$$
$$\mathbf{y}_k = \mathbf{H}\,\mathbf{x}(t_k) + \mathbf{r}_k.$$

- The dynamics of state $\mathbf{x}(t) \in \mathbb{R}^n$ are modeled as Itô-type stochastic differential equations (SDE, Itô diffusion).
- $\boldsymbol{\beta}(t) \in \mathbb{R}^s$ is a vector of Brownian motions (Wiener processes) with diffusion matrix $\mathbf{Q}$ and dimension $s \leq n$.
- $\mathbf{r}_k \in \mathbb{R}^d$ is a Gaussian random variable $\mathbf{r}_k \sim \mathrm{N}(0, \mathbf{R})$.
- We can think SDE as white noise driven differential equation

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

where the white noise is defined as $\mathbf{w}(t) = d\boldsymbol{\beta}(t)/dt$.

Bayesian Filtering and Smoothing Solution

- We don't aim to compute the full (infinite-dimensional) posterior of the state, but instead only its time-marginals.
- Filtering/prediction solutions: Compute the posterior distribution(s)

$$p(\mathbf{x}(t) \mid \mathbf{y}_1, \ldots, \mathbf{y}_k), \qquad t \in [t_k, t_{k+1}).$$

- Smoothing solution: Compute the posterior distribution(s)

$$p(\mathbf{x}(t) \mid \mathbf{y}_1, \ldots, \mathbf{y}_T), \qquad t \in [t_0, t_T].$$

- If we could solve the transition density $p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1}))$, the model would reduce to a discrete-time model:

$$\begin{aligned} \mathbf{x}(t_k) &\sim p(\mathbf{x}(t_k) \mid \mathbf{x}(t_{k-1})) \\ \mathbf{y}_k &\sim p(\mathbf{y}_k \mid \mathbf{x}(t_k)). \end{aligned}$$

- Linear Gaussian systems can be treated with Kalman filter and RTS smoother – e.g. the car model and linear LFMs.

Formal Bayesian Continuous-Discrete Filter

- General continuous-discrete filtering model:

$$
\begin{aligned}
d\mathbf{x} &= \mathbf{f}(\mathbf{x})\,dt + \mathbf{L}(\mathbf{x})\,d\beta(t) \\
\mathbf{y}_k &\sim p(\mathbf{y}_k \,|\, \mathbf{x}(t_k)).
\end{aligned}
$$

Continuous-Discrete Bayesian Optimal filter

1. Prediction step: Solve the Fokker-Planck-Kolmogorov PDE

$$
\frac{\partial p}{\partial t} = -\sum_i \frac{\partial}{\partial x_i} (f_i(\mathbf{x})\,p) + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \Big([\mathbf{L}(\mathbf{x})\,\mathbf{Q}\,\mathbf{L}^T_{ij}(\mathbf{x})\,p \Big)
$$

2. Update step: Apply the Bayes' rule.

$$
p(\mathbf{x}(t_k) \,|\, \mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k \,|\, \mathbf{x}(t_k))\,p(\mathbf{x}(t_k) \,|\, \mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k \,|\, \mathbf{x}(t_k))\,p(\mathbf{x}(t_k) \,|\, \mathbf{y}_{1:k-1})\,d\mathbf{x}(t_k)}
$$

Continuous-Discrete Non-Linear Kalman Filtering [1/2]

- The current special case of the model is:

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

- We can now apply Gaussian (process) approximation to the posterior of the process $\mathbf{x}(t)$ – when combined with approximate Bayesian filter, leads to non-linear Kalman filters.
- Note that we can easily generalize to non-linear measurement model $\mathbf{H}\, \mathbf{x}(t_k) \to \mathbf{h}(\mathbf{x}(t_k))$.
- The resulting approximation is of the form

$$p(\mathbf{x}(t) \mid \mathbf{y}_{1:k}) \approx \mathrm{N}(\mathbf{x}(t) \mid \mathbf{m}(t), \mathbf{P}(t)), \quad t \in [t_k, t_{k+1}),$$

where $\mathbf{m}(t)$ and $\mathbf{P}(t)$ are computed by the non-linear Kalman filter.
- Different brands: EKF, UKF, CKF, GHKF, etc.

Continuous-Discrete Non-Linear Kalman Filtering [2/2]

Continuous-Discrete Non-Linear Kalman Filter

1. Prediction step: Integrate the following time t_{k-1} to t_k^-:

$$
\begin{aligned}
\frac{d\mathbf{m}}{dt} &= \mathrm{E}[\mathbf{f}(\mathbf{x})] \\
\frac{d\mathbf{P}}{dt} &= \mathrm{E}[(\mathbf{x} - \mathbf{m}_k)\,\mathbf{f}^T(\mathbf{x})] + \mathrm{E}[\mathbf{f}(\mathbf{x})\,(\mathbf{x} - \mathbf{m})^T] + \mathrm{E}[\mathbf{L}(\mathbf{x})\,\mathbf{Q}\,\mathbf{L}^T(\mathbf{x})].
\end{aligned}
$$

2. Update step: Update step is the linear Kalman filter update:

$$
\begin{aligned}
\mathbf{S}_k &= \mathbf{H}\,\mathbf{P}(t_k^-)\,\mathbf{H}^T + \mathbf{R}_k \\
\mathbf{K}_k &= \mathbf{P}(t_k^-)\,\mathbf{H}_k^T\,\mathbf{S}_k^{-1} \\
\mathbf{m}(t_k) &= \mathbf{m}(t_k^-) + \mathbf{K}_k\,[\mathbf{y}_k - \mathbf{H}\,\mathbf{m}(t_k^-)] \\
\mathbf{P}(t_k) &= \mathbf{P}(t_k^-) - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T.
\end{aligned}
$$

Formal Bayesian Continuous-Discrete Smoother

- Continuous-discrete (-time) smoothing refers to recursive computation of the distributions

$$p(\mathbf{x}(t) \mid \mathbf{y}_{1:T}), \qquad t \in [t_0, t_T].$$

- Discrete-time smoothing refers to computation of $p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:T})$ for $k = 1, \ldots, T$.
- The (discrete-time) Bayesian smoothing equation is

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:n}) = p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \int \frac{p(\mathbf{x}_{k+1} \mid \mathbf{y}_{1:n})\, p(\mathbf{x}_{k+1} \mid \mathbf{x}_k)}{p(\mathbf{x}_{k+1} \mid \mathbf{y}_{1:k})}\, d\mathbf{x}_{k+1}.$$

- The continuous-time version of the above is a quite complicated partial differential equation (PDE).
- Continuous-discrete non-linear Gaussian smoother can be derived by computing the continuous-time limit of the Gaussian discrete-time smoother.

Continuous-Discrete Gaussian Smoothing [1/2]

The following discrete-time Gaussian smoother is due to Särkkä and Hartikainen (2010):

Discrete-time Gaussian smoother

$$
\begin{aligned}
\mathbf{m}_{k+1}^{-} &= \int \mathbf{f}(\mathbf{x}_k)\,\mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k)\, d\mathbf{x}_k \\
\mathbf{P}_{k+1}^{-} &= \int [\mathbf{f}(\mathbf{x}_k) - \mathbf{m}_{k+1}^{-}]\,[\mathbf{f}(\mathbf{x}_k) - \mathbf{m}_{k+1}^{-}]^T\,\mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k)\, d\mathbf{x}_k + \mathbf{Q}_k \\
\mathbf{D}_{k+1} &= \int [\mathbf{x}_k - \mathbf{m}_k]\,[\mathbf{f}(\mathbf{x}_k) - \mathbf{m}_{k+1}^{-}]^T\,\mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k)\, d\mathbf{x}_k \\
\mathbf{G}_k &= \mathbf{D}_{k+1}\,[\mathbf{P}_{k+1}^{-}]^{-1} \\
\mathbf{m}_k^s &= \mathbf{m}_k + \mathbf{G}_k\,(\mathbf{m}_{k+1}^s - \mathbf{m}_{k+1}^{-}) \\
\mathbf{P}_k^s &= \mathbf{P}_k + \mathbf{G}_k\,(\mathbf{P}_{k+1}^s - \mathbf{P}_{k+1}^{-})\,\mathbf{G}_k^T.
\end{aligned}
$$

Continuous-Time Gaussian Smoothing [2/2]

The continuous-discrete Gaussian approximation based non-linear smoother (Särkkä and Sarmavuori, 2013) forms the approximations

$$p(\mathbf{x}(t) \,|\, \mathbf{y}_{1:T}) \approx \mathrm{N}(\mathbf{x}(t) \,|\, \mathbf{m}^s(t), \mathbf{P}^s(t)), \qquad t \in [t_0, t_T].$$

Continuous-Discrete Gaussian smoother

$$\begin{aligned}
\frac{d\mathbf{m}^s}{dt} &= \mathrm{E}[\mathbf{f}(\mathbf{x})] + \left\{ \mathrm{E}[\mathbf{f}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^T] + \mathrm{E}[\mathbf{L}(\mathbf{x})\,\mathbf{Q}\,\mathbf{L}^T(\mathbf{x})] \right\} \mathbf{P}^{-1}[\mathbf{m}^s - \mathbf{m}] \\
\frac{d\mathbf{P}^s}{dt} &= \left\{ \mathrm{E}[\mathbf{f}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^T] + \mathrm{E}[\mathbf{L}(\mathbf{x})\,\mathbf{Q}\,\mathbf{L}^T(\mathbf{x})] \right\} \mathbf{P}^{-1}\,\mathbf{P}^s \\
&\quad + \mathbf{P}^s\,\mathbf{P}^{-1} \left\{ \mathrm{E}[\mathbf{f}(\mathbf{x})\,(\mathbf{x}-\mathbf{m})^T] + \mathrm{E}[\mathbf{L}(\mathbf{x})\,\mathbf{Q}\,\mathbf{L}^T(\mathbf{x})] \right\}^T \\
&\quad - \mathrm{E}[\mathbf{L}(\mathbf{x})\,\mathbf{Q}\,\mathbf{L}^T(\mathbf{x})].
\end{aligned}$$

where the expectations are with respect to the filtering distribution.

Continuous-Discrete Particle Filtering and Smoothing

- In particle filtering and smoothing we use sequential Monte Carlo approximations to the distributions.
- The most common discrete-time framework is based on sequential importance sampling / resampling.
- Continuous-discrete particle methods can be formed by (a) discretizing the SDE (b) by computing continuous limits of sequential Monte Carlo methods (c) using specialized versions of SMC.
- There are also algorithms for direct sampling from the smoothing distribution.

SDE View of Latent Force Models [1/4]

- Let's now take a look at the non-linear state-space LFM methodology presented in Hartikainen and Särkkä (2012).
- Consider the latent force model (Lawrence et al., 2006)

$$\frac{dx_j(t)}{dt} = B_j + \sum_{r=1}^{R} S_{j,r} g_j(u_r(t)) - D_j x_j(t), \;\; j = 1, \ldots, N$$

- We can now use independent Gaussian process (GP) priors

$$u_r(t) \sim \mathrm{GP}(m(t), k_{u_r}(t, t')), \; r = 1, \ldots, R$$

  where $m(t)$ and $k_{u_r}(t, t')$ were suitably chosen mean and covariance functions.
- Recall that GPs with certain stationary covariance functions (e.g. Matérn) can be represented as state space models. Assume that we use such covariance function and set $m(t) = 0$.

SDE View of Latent Force Models [2/4]

- That is, we can formulate the GP priors on the components of $\mathbf{u}(t) = (u_1(t) \ldots u_R(t))^T$ as multivariate space space models (SDEs) of form
$$d\mathbf{z}_r(t) = \mathbf{F}_{z,r}\, \mathbf{z}_r(t)\, dt + \mathbf{L}_{z,r}\, d\beta_{z,r}(t)$$
where
$$\mathbf{z}_r(t) = \begin{pmatrix} u_r(t) & \frac{du_r(t)}{dt} & \ldots & \frac{d^{d_r-1} u_r(t)}{dt^{d_r-1}} \end{pmatrix}^T$$
and
$$\mathbf{F}_{z,r} = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_r^0 & \ldots & -a_r^{p_r-2} & -a_r^{p_r-1} \end{pmatrix}, \quad \mathbf{L}_{z,r} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ q_r \end{pmatrix}.$$

SDE View of Latent Force Models [3/4]

- Thus a general non-linear latent force model (LFM) with GP prior can be formulated as a continuous-discrete system of form

$$
\begin{aligned}
d\mathbf{x}(t) &= \mathbf{f}_f(\mathbf{x}(t), \mathbf{u}(t), t)\, dt, \\
d\mathbf{z}_r(t) &= \mathbf{F}_{z,r}\, \mathbf{z}_r(t)\, dt + \mathbf{L}_{z,r}\, d\boldsymbol{\beta}_{z,r}(t),\ r = 1, \ldots, R
\end{aligned}
$$

  where $\mathbf{x}(t) \in \mathbb{R}^M$ is the state.
- M is the number of state components needed in representing the output processes $\{x_j(t)\}_{j=1}^N$ in a vector form,
- $\mathbf{u}(t) \in \mathbb{R}^R$ are the latent force processes and
- $\mathbf{f}_f(\cdot)$ the dynamic model function corresponding to the LFM.

Aalto University

SDE View of Latent Force Models [4/4]

- We can further simplify the notation by constructing an augmented system with state $\mathbf{x}_a(t)$ comprising of the output process and latent forces as $\mathbf{x}_a(t) = (\mathbf{x}(t), \mathbf{z}_1(t), \ldots, \mathbf{z}_R(t))^T$ with dynamics

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

- To complete the model specification we assume that observations at discrete time instants $t_1, \ldots, t_T$ can be modeled as

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_a(t_k)) + \mathbf{r}_k, \quad k = 1, \ldots, T$$

  where $\mathbf{h}(\cdot)$ is the measurement model function, $\mathbf{y}_k \in \mathbb{R}^D$ is the measurement at time t_k and $\mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_k)$ is the measurement noise.
- This is now just a non-linear state space model and completely tractable with non-linear Kalman filtering and smoothing methods.

Summary

- Non-linear LFMs can be converted into state-space form by:
  1. Converting the latent GPs into state-space form.
  2. Forming an augmented state space model.
- Bayesian filtering and smoothing, in principle, provide the full solution to the problem.
- In practice, formal solution is intractable – involves, e.g., solutions to particle differential equations.
- Approximate inference in non-linear LFMs can be implemented with non-linear Kalman filters and smoothers.

References

- J. Hartikainen and S. Särkkä (2010). *Kalman Filtering and Smoothing Solutions to Temporal Gaussian Process Regression Models*. Proceedings of MLSP.
- S. Särkkä and J. Hartikainen (2010). *On Gaussian Optimal Smoothing of Non-Linear State Space Models*. IEEE Transactions on Automatic Control, Volume 55, Issue 8, Pages 1938–1941.
- J. Hartikainen, M. Seppänen and S. Särkkä (2012). *State-Space Inference for Non-Linear Latent Force Models with Application to Satellite Orbit Prediction*. Proceedings of ICML.
- S. Särkkä and J. Sarmavuori (2013). *Gaussian Filtering and Smoothing for Continuous-Discrete Dynamic Systems*. Signal Processing, Volume 93. Issue 2, Pages 500–510.
- S. Särkkä (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press. (Forthcoming).