Lecture 5: Further Topics; Series Expansions, Feynman–Kac, Girsanov Theorem, Filtering Theory

Simo Särkkä

Aalto University
Tampere University of Technology
Lappeenranta University of Technology
Finland

November 22, 2012

Contents

Karhunen–Loeve expansions [1/2]

- On a fixed time interval $[t_0, t_1]$ the standard Brownian motion has a Karhunen–Loeve series expansion

$$\beta(t) = \sum_{i=1}^{\infty} z_i \int_{t_0}^{t} \phi_i(\tau)\, \mathrm{d}\tau.$$

- $z_i \sim \mathrm{N}(0, 1)$, $i = 1, 2, 3, \ldots$ are independent random variables.
- $\{\phi_i(t)\}$ is an orthonormal basis of the Hilbert space with inner product

$$\langle f, g \rangle = \int_{t_0}^{t_1} f(\tau)\, g(\tau)\, d\tau.$$

- In fact, this is just a Fourier series and thus:

$$z_i = \int_{t_0}^{t} \phi_i(\tau)\, \mathrm{d}\beta(\tau).$$

Karhunen–Loeve expansions [2/2]

- We could now consider approximating the SDE

$$\mathrm{d}x = f(x,t)\,\mathrm{d}t + L(x,t)\,\mathrm{d}\beta.$$

by using a finite expansion

$$\mathrm{d}\beta(t) = \sum_{i=1}^{N} z_i\,\phi_i(t)\,\mathrm{d}t.$$

- However, this converges to the Stratonovich SDE

$$\mathrm{d}x = f(x,t)\,\mathrm{d}t + L(x,t)\circ\mathrm{d}\beta.$$

- That is, we can approximate the Stratonovich SDE with

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x,t) + L(x,t)\sum_{i=1}^{N} z_i\,\phi_i(t).$$

- A special case of Wong–Zakai approximations.

Wiener Chaos Expansions

- Let's consider the infinite expansion

$$\mathrm{d}x = f(x, t)\,\mathrm{d}t + L(x, t) \sum_{i=1}^{\infty} z_i\,\phi_i(t)\,\mathrm{d}t,$$

- The solution can be seen as a function (or functional) of the form

$$x(t) = U(t; z_1, z_2, \ldots).$$

- It is now possible to form a Fourier–Hermite series expansion of RHS in the variables $z_1, z_2, \ldots$.
- Leads to a polynomial series expansion which is called Wiener chaos expansion or polynomial chaos.

Feynman–Kac formulae and parabolic PDEs [1/5]

- Feynman–Kac formula gives a link between parabolic partial differential equations (PDEs) and SDEs.
- Consider the following PDE for function $u(x, t)$:

$$\frac{\partial u}{\partial t} + f(x)\,\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2} = 0$$
$$u(x, T) = \Psi(x).$$

- Let's define a process $x(t)$ on the interval $[t', T]$ as follows:

$$\mathrm{d}x = f(x)\,\mathrm{d}t + \sigma(x)\,\mathrm{d}\beta, \qquad x(t') = x'.$$

Feynman–Kac formulae and parabolic PDEs [2/5]

- Using Itô formula for $u(x,t)$ gives:

$$
\begin{aligned}
\mathrm{d}u &= \frac{\partial u}{\partial t}\,\mathrm{d}t + \frac{\partial u}{\partial x}\,\mathrm{d}x + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}\,\mathrm{d}x^2 \\
&= \frac{\partial u}{\partial t}\,\mathrm{d}t + \frac{\partial u}{\partial x}\,f(x)\,\mathrm{d}t + \frac{\partial u}{\partial x}\,\sigma(x)\,\mathrm{d}\beta + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}\,\sigma^2(x)\,\mathrm{d}t \\
&= \underbrace{\left[\frac{\partial u}{\partial t} + f(x)\,\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\,\frac{\partial^2 u}{\partial x^2}\right]}_{=0}\,\mathrm{d}t + \frac{\partial u}{\partial x}\,\sigma(x)\,\mathrm{d}\beta \\
&= \frac{\partial u}{\partial x}\,\sigma(x)\,\mathrm{d}\beta.
\end{aligned}
$$

Feynman–Kac formulae and parabolic PDEs [3/5]

- We now have

$$\mathrm{d}u = \frac{\partial u}{\partial x}\,\sigma(x)\,\mathrm{d}\beta.$$

- Integrating from t' to T now gives

$$u(x(T),T) - u(x(t'),t') = \int_{t'}^{T} \frac{\partial u}{\partial x}\,\sigma(x)\,\mathrm{d}\beta,$$

- Substituting the initial and terminal terms we get:

$$\Psi(x(T)) - u(x',t') = \int_{t'}^{T} \frac{\partial u}{\partial x}\,\sigma(x)\,\mathrm{d}\beta.$$

Feynman–Kac formulae and parabolic PDEs [4/5]

- Take expectations from both sides

$$\mathrm{E}\left[\Psi(x(T)) - [u(x',t')]\right] = \underbrace{E\left[\int_{t'}^{T} \frac{\partial u}{\partial x}\,\sigma(x)\,\mathrm{d}\beta\right]}_{=0}$$

- leads to

$$u(x',t') = \mathrm{E}[\Psi(x(T))].$$

- Thus we can solve the value of $u(x',t')$ for arbitrary x' and t' as follows:
  1. Start the following process from x' and time t' and let it run until time T:

  $$\mathrm{d}x = f(x)\,\mathrm{d}t + \sigma(x)\,\mathrm{d}\beta$$

  2. The value of $u(x',t')$ is the expected value of $\Psi(x(T))$ over the process realizations.

Feynman–Kac formulae and parabolic PDEs [5/5]

- Can be generalized to equations of the form

$$\frac{\partial u}{\partial t} + f(x)\,\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2} - r\,u = 0$$
$$u(x, T) = \Psi(x),$$

- The SDE is the same and the corresponding solution is

$$u(x', t') = \exp(-r\,(T - t'))\;\mathrm{E}[\Psi(x(T))]$$

- Can be generalized in various ways: multiple dimensions, $r(x)$, constant terms, etc.

Solving Boundary Value Problems with Feynman–Kac [1/3]

- Consider the following boundary value problem for $u(x)$ on Ω:

$$f(x)\,\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2} = 0$$
$$u(x) = \Psi(x), x \in \partial\Omega.$$

- Again, define a process $x(t)$ as follows:

$$\mathrm{d}x = f(x)\,\mathrm{d}t + \sigma(x)\,\mathrm{d}\beta, \qquad x(t') = x'.$$

- Application of Itô formula to $u(x)$ gives

$$\mathrm{d}u = \underbrace{\left[f(x)\,\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\,\frac{\partial^2 u}{\partial x^2}\right]}_{=0}\mathrm{d}t + \frac{\partial u}{\partial x}\,\sigma(x)\,\mathrm{d}\beta$$
$$= \frac{\partial u}{\partial x}\,\sigma(x)\,\mathrm{d}\beta.$$

Solving Boundary Value Problems with Feynman–Kac [2/3]

- Let T_e be the first exit time of $x(t)$ from Ω.
- Integration from t' to T_e gives

$$u(x(T_e)) - u(x(t')) = \int_{t'}^{T_e} \frac{\partial u}{\partial x} \, \sigma(x) \, \mathrm{d}\beta.$$

- But the value of $u(x)$ on the boundary is $\Psi(x)$ and $x(t') = x'$, which leads to:

$$\Psi(x(T_e)) - u(x') = \int_{t'}^{T_e} \frac{\partial u}{\partial x} \, \sigma(x) \, \mathrm{d}\beta.$$

- Taking expectation and rearranging then gives

$$u(x') = \mathrm{E}[\Psi(x(T_e))].$$

Solving Boundary Value Problems with Feynman–Kac [3/3]

- Thus we can solve the boundary value problem

$$f(x)\,\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2} = 0$$
$$u(x) = \Psi(x), x \in \partial\Omega.$$

at point x' as follows:

1. Start the following process from x':

$$\mathrm{d}x = f(x)\,\mathrm{d}t + \sigma(x)\,\mathrm{d}\beta.$$

2. Compute the following expectation at the positions $x(T_e)$ of the first exit times from the domain Ω:

$$u(x') = \mathrm{E}[\Psi(x(T_e))].$$

Girsanov theorem [1/6]

- Let's denote the whole path of the Itô process $\mathbf{x}(t)$ on a time interval $[0, t]$ as:

$$\mathscr{X}_t = \{\mathbf{x}(\tau) \;:\; 0 \le \tau \le t\}.$$

- Let $\mathbf{x}(t)$ be the solution to

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\;\mathrm{d}t + \mathrm{d}\boldsymbol{\beta}.$$

- Formally define the probability density of the whole path as

$$p(\mathscr{X}_t) = \lim_{N \to \infty} p(\mathbf{x}(t_1), \ldots, \mathbf{x}(t_N)).$$

- Not normalizable, but we can define the following for suitable $\mathbf{y}$:

$$\frac{p(\mathscr{X}_t)}{p(\mathscr{Y}_t)} = \lim_{N \to \infty} \frac{p(\mathbf{x}(t_1), \ldots, \mathbf{x}(t_N))}{p(\mathbf{y}(t_1), \ldots, \mathbf{y}(t_N))}.$$

Girsanov theorem [2/6]

- The Girsanov theorem is a way to make this kind of analysis rigorous.
- Connected to path integrals which can be considered as expectations of the form

$$\mathsf{E}[h(\mathscr{X}_t)] = \int h(\mathscr{X}_t)\, p(\mathscr{X}_t)\, \mathrm{d}\mathscr{X}_t.$$

- This notation is purely formal, because the density $p(\mathscr{X}_t)$ is actually an infinite quantity.
- But the expectation (path integral) is indeed finite.

Girsanov theorem [3/6]

Likelihood ratio of Itô processes

Consider the Itô processes

$$\begin{aligned} \mathrm{d}\mathbf{x} &= \mathbf{f}(\mathbf{x}, t)\, \mathrm{d}t + \mathrm{d}\boldsymbol{\beta}, \qquad & \mathbf{x}(0) &= \mathbf{x}_0, \\ \mathrm{d}\mathbf{y} &= \mathbf{g}(\mathbf{y}, t)\, \mathrm{d}t + \mathrm{d}\boldsymbol{\beta}, \qquad & \mathbf{y}(0) &= \mathbf{x}_0, \end{aligned}$$

where the Brownian motion $\boldsymbol{\beta}(t)$ has a non-singular diffusion matrix $\mathbf{Q}$.

- The ratio of the probability laws of $\mathscr{X}_t$ and $\mathscr{Y}_t$ is given as

$$\frac{p(\mathscr{X}_t)}{p(\mathscr{Y}_t)} = Z(t)$$

$$Z(t) = \exp\left(-\frac{1}{2} \int_0^t [\mathbf{f}(\mathbf{y}, \tau) - \mathbf{g}(\mathbf{y}, \tau)]^{\mathsf{T}}\, \mathbf{Q}^{-1}\, [\mathbf{f}(\mathbf{y}, \tau) - \mathbf{g}(\mathbf{y}, \tau)]\, \mathrm{d}\tau \right.$$
$$\left. + \int_0^t [\mathbf{f}(\mathbf{y}, \tau) - \mathbf{g}(\mathbf{y}, \tau)]^{\mathsf{T}}\, \mathbf{Q}^{-1}\, \mathrm{d}\boldsymbol{\beta}(\tau) \right)$$

Girsanov theorem [4/6]

Likelihood ratio of Itô processes (cont.)

- For an arbitrary functional $h(\bullet)$ of the path from 0 to t we have

$$\mathsf{E}[h(\mathscr{X}_t)] = \mathsf{E}[Z(t)\, h(\mathscr{Y}_t)],$$

- Under the probability measure defined through the transformed probability density

$$\tilde{p}(\mathscr{X}_t) = Z(t)\, p(\mathscr{X}_t)$$

the process

$$\tilde{\beta} = \beta - \int_0^t [\mathbf{f}(\mathbf{y}, \tau) - \mathbf{g}(\mathbf{y}, \tau)] \, \mathrm{d}\tau,$$

is a Brownian motion with diffusion matrix **Q**.

- Derivation on the blackboard follows.

Girsanov theorem [5/6]

Girsanov I

Let $\boldsymbol{\theta}(t)$ be a process driven by a standard Brownian motion $\boldsymbol{\beta}(t)$ such that $\mathrm{E}\left[\int_0^t \boldsymbol{\theta}^\mathsf{T}(\tau)\,\boldsymbol{\theta}(\tau)\,\mathrm{d}\tau\right] < \infty$, then under the measure defined by the formal probability density

$$\tilde{p}(\Theta_t) = Z(t)\,p(\Theta_t),$$

where $\Theta_t = \{\boldsymbol{\theta}(\tau) \;:\; 0 \le \tau \le t\}$, and

$$Z(t) = \exp\left(\int_0^t \boldsymbol{\theta}^\mathsf{T}(\tau)\,\mathrm{d}\boldsymbol{\beta} - \frac{1}{2}\int_0^t \boldsymbol{\theta}^\mathsf{T}(\tau)\,\boldsymbol{\theta}(\tau)\,\mathrm{d}\tau\right),$$

the following process is a standard Brownian motion:

$$\tilde{\boldsymbol{\beta}}(t) = \boldsymbol{\beta}(t) - \int_0^t \boldsymbol{\theta}(\tau)\,\mathrm{d}\tau.$$

Girsanov theorem [6/6]

Girsanov II

Let $\boldsymbol{\beta}(\omega, t)$ be a standard n-dimensional Brownian motion under the probability measure $\mathbb{P}$. Let $\boldsymbol{\theta} : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}^n$ be an adapted process such that the process Z defined as

$$Z(\omega, t) = \exp\left\{ \int_0^t \boldsymbol{\theta}^\mathsf{T}(\omega, t)\, \mathrm{d}\boldsymbol{\beta}(\omega, t) - \frac{1}{2} \int_0^t \boldsymbol{\theta}^\mathsf{T}(\omega, t)\, \boldsymbol{\theta}(\omega, t)\, \mathrm{d}t \right\},$$

satisfies $\mathrm{E}[Z(\omega, t)] = 1$. Then the process

$$\tilde{\boldsymbol{\beta}}(\omega, t) = \boldsymbol{\beta}(\omega, t) - \int_0^t \boldsymbol{\theta}(\omega, \tau)\; \mathrm{d}\tau$$

is a standard Brownian motion under the probability measure $\tilde{\mathbb{P}}$ defined via the relation $\mathrm{E}\left[\mathrm{d}\tilde{\mathbb{P}} \big/ \mathrm{d}\mathbb{P}\,(\omega) \,\middle|\, \mathscr{F}_t \right] = Z(\omega, t)$, where $\mathscr{F}_t$ is the natural filtration of the Brownian motion $\boldsymbol{\beta}(\omega, t)$.

Applications of Girsanov theorem

- Removal of drift: define $\theta(t)$ in terms of the drift function suitably such that in the transformed SDE the drift cancels out.
- Weak solutions of SDEs: Select $\theta(t)$ such that an easy process $\tilde{\mathbf{x}}(t)$ solves the SDE with the constructed $\tilde{\beta}(t)$.
- Kallianpur–Striebel formula (Bayes' rule in continuous time) and stochastic filtering theory.
- Importance sampling and exact sampling of SDEs.

Continuous-Time Stochastic Filtering Theory [1/3]

- Consider the following filtering model:

$$\begin{aligned} \mathrm{d}\mathbf{x}(t) &= \mathbf{f}(\mathbf{x}(t), t)\, \mathrm{d}t + \mathbf{L}(\mathbf{x}, t)\, \mathrm{d}\beta(t) \\ \mathrm{d}\mathbf{y}(t) &= \mathbf{h}(\mathbf{x}(t), t)\, \mathrm{d}t + \mathrm{d}\eta(t). \end{aligned}$$

- Given that we have observed $\mathbf{y}(t)$, what can we say (in statistical sense) about the hidden process $\mathbf{x}(t)$?
- The first equation defines dynamics of the system state and the second relates measurements to the state.
- Physical interpretation of the measurement model:

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t), t) + \epsilon(t),$$

  where $\mathbf{z}(t) = \mathrm{d}\mathbf{y}(t)/\,\mathrm{d}t$ and $\epsilon(t) = \mathrm{d}\eta(t)/\,\mathrm{d}t$.
- For example, $\mathbf{x}(t)$ may contain position and velocity of a car and $\mathbf{y}(t)$ might be a radar measurement.

Continuous-Time Stochastic Filtering Theory [2/3]

- The solution can be solved using Bayesian inference.
- This Bayesian solution is surpricingly old, as it dates back to work of Stratonovich around 1950.
- The aim is to compute the filtering (posterior) distribution

$$p(\mathbf{x}(t) \mid \mathscr{Y}_t).$$

  where $\mathscr{Y}_t = \{\mathbf{y}(\tau) \ : \ 0 \leq \tau \leq t\}$.
- The solutions are called the Kushner–Stratonovich equation and the Zakai equation.
- The solution to the linear Gaussian problem is called Kalman–Bucy filter.

Continuous-Time Stochastic Filtering Theory [3/3]

Remark on notation

If we define a linear operator $\mathscr{A}^*$ as

$$\mathscr{A}^*(\bullet) = -\sum_i \frac{\partial}{\partial x_i}[f_i(x,t)\,(\bullet)] + \frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial x_i\,\partial x_j}\{[\mathbf{L}(\mathbf{x},t)\,\mathbf{Q}\,\mathbf{L}^\mathsf{T}(\mathbf{x},t)]_{ij}\,(\bullet)\}.$$

Then the Fokker–Planck–Kolmogorov equation can be written compactly as

$$\frac{\partial p}{\partial t} = \mathscr{A}^* p.$$

Kushner–Stratonovich equation

Kushner–Stratonovich equation

The stochastic partial differential equation for the filtering density $p(\mathbf{x}, t \mid \mathscr{Y}_t) \triangleq p(\mathbf{x}(t) \mid \mathscr{Y}_t)$ is

$$\begin{aligned}
&\mathrm{d}p(\mathbf{x}, t \mid \mathscr{Y}_t) = \mathscr{A}^* \, p(\mathbf{x}, t \mid \mathscr{Y}_t) \, \mathrm{d}t \\
&+ (\mathbf{h}(\mathbf{x}, t) - \mathrm{E}[\mathbf{h}(\mathbf{x}, t) \mid \mathscr{Y}_t])^\mathsf{T} \, \mathbf{R}^{-1} (\mathrm{d}\mathbf{y} - \mathrm{E}[\mathbf{h}(\mathbf{x}, t) \mid \mathscr{Y}_t] \, \mathrm{d}t) \, p(\mathbf{x}, t \mid \mathscr{Y}_t),
\end{aligned}$$

where $\mathrm{d}p(\mathbf{x}, t \mid \mathscr{Y}_t) = p(\mathbf{x}, t + \mathrm{d}t \mid \mathscr{Y}_{t+\mathrm{d}t}) - p(\mathbf{x}, t \mid \mathscr{Y}_t)$ and

$$\mathrm{E}[\mathbf{h}(\mathbf{x}, t) \mid \mathscr{Y}_t] = \int \mathbf{h}(\mathbf{x}, t) \, p(\mathbf{x}, t \mid \mathscr{Y}_t) \, \mathrm{d}\mathbf{x}.$$

Zakai equation

Let $q(\mathbf{x}, t \mid \mathscr{Y}_t) \triangleq q(\mathbf{x}(t) \mid \mathscr{Y}_t)$ be the solution to Zakai's stochastic partial differential equation

$$\mathrm{d}q(\mathbf{x}, t \mid \mathscr{Y}_t) = \mathscr{A}^* \, q(\mathbf{x}, t \mid \mathscr{Y}_t) \, \mathrm{d}t + \mathbf{h}^\mathsf{T}(\mathbf{x}, t) \, \mathbf{R}^{-1} \, \mathrm{d}\mathbf{y} \, q(\mathbf{x}, t \mid \mathscr{Y}_t),$$

where $\mathrm{d}q(\mathbf{x}, t \mid \mathscr{Y}_t) = q(\mathbf{x}, t + \mathrm{d}t \mid \mathscr{Y}_{t+\mathrm{d}t}) - q(\mathbf{x}, t \mid \mathscr{Y}_t)$. Then we have

$$p(\mathbf{x}(t) \mid \mathscr{Y}_t) = \frac{q(\mathbf{x}(t) \mid \mathscr{Y}_t)}{\int q(\mathbf{x}(t) \mid \mathscr{Y}_t) \, \mathrm{d}\mathbf{x}(t)}.$$

Kalman–Bucy filter

The Kalman–Bucy filter is the exact solution to the linear Gaussian filtering problem

$$
\begin{aligned}
\mathrm{d}\mathbf{x} &= \mathbf{F}(t)\,\mathbf{x}\;\mathrm{d}t + \mathbf{L}(t)\;\mathrm{d}\boldsymbol{\beta} \\
\mathrm{d}\mathbf{y} &= \mathbf{H}(t)\,\mathbf{x}\;\mathrm{d}t + \mathrm{d}\boldsymbol{\eta}.
\end{aligned}
$$

Kalman–Bucy filter

The Bayesian filter, which computes the posterior distribution $p(\mathbf{x}(t) \mid \mathscr{Y}_t) = \mathrm{N}(\mathbf{x}(t) \mid \mathbf{m}(t), \mathbf{P}(t))$ for the above system is

$$
\begin{aligned}
\mathbf{K}(t) &= \mathbf{P}(t)\,\mathbf{H}^\mathsf{T}(t)\,\mathbf{R}^{-1} \\
\mathrm{d}\mathbf{m}(t) &= \mathbf{F}(t)\,\mathbf{m}(t)\;\mathrm{d}t + \mathbf{K}(t)\;[\mathrm{d}\mathbf{y}(t) - \mathbf{H}(t)\,\mathbf{m}(t)\;\mathrm{d}t] \\
\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} &= \mathbf{F}(t)\,\mathbf{P}(t) + \mathbf{P}(t)\,\mathbf{F}^\mathsf{T}(t) + \mathbf{L}(t)\,\mathbf{Q}\,\mathbf{L}^\mathsf{T}(t) - \mathbf{K}(t)\,\mathbf{R}\,\mathbf{K}^\mathsf{T}(t).
\end{aligned}
$$

Approximations of nonlinear filters

- Monte Carlo approximations (particle filters).
- Series expansions of processes.
- Series expansions of probability densities.
- Gaussian process approximations (non-linear Kalman filters).

Summary

- Brownian motion can be expanded into Karhunen–Loeve series.
- The series can be substituted into SDE leading to a class of Wong–Zakai approximations or to Wiener Chaos Expansions.
- Feynman–Kac formulae can be used to solve PDEs by simulating solutions of SDEs.
- Girsanov theorem is related to computation of likelihood ratios of processes.
- Applications of Girsanov theorem include removal drifts, solving SDEs and deriving results and methods for stochastic filtering theory.
- Filtering theory is related to Bayesian reconstruction of a hidden stochastic process $\mathbf{x}(t)$ from an observed stochastic process $\mathbf{y}(t)$.