Fundamentals of Stochastic Filtering

Lecture 3: The Zakai Equation and the Kushner-Stratonovich Equation

Zeju Sun

Geometry, Intelligent Control, and Bioinformatics Interdisciplinary Group
Beijing Institute of Mathematical Sciences and Applications (BIMSA)

sunzeju@bimsa.cn

October 10, 2025

Contents

Discrete-Time vs. Continuous-Time Filtering Problem

- Continuous-time Filtering Problem:

$$\begin{cases} dX_t = f(X_t)dt + \sigma(X_t)dV_t \\ dY_t = h(X_t)dt + dW_t \end{cases} \quad t \in [0, T]. \tag{1}$$

- Time-discretization: $0 = t_0 < t_1 < \cdots < t_n = T$

$$\begin{cases} \bar{X}_{t_k} = \bar{X}_{t_{k-1}} + f(\bar{X}_{t_{k-1}})\Delta t + \sigma(\bar{X}_{t_{k-1}})\Delta V_{t_k} \\ \Delta Y_{t_k} = h(\bar{X}_{t_k})\Delta t + \Delta W_{t_k} \end{cases}, \quad k = 1, \cdots, n. \tag{2}$$

- Theoretical solution of discrete-time filtering problem:
  - Prediction Step:

$$p_{k+1|k}(x|\delta y_{1:k}) = \int p_{k|k}(z|\delta y_{1:k}) q_{k+1|k}(x|z) dz \tag{3}$$

  - Update Step:

$$p_{k+1|k+1}(x|\delta y_{1:k+1}) \propto l(x, \delta y_{k+1}) p_{k+1|k}(x|\delta y_{1:k}). \tag{4}$$

Discrete-Time vs. Continuous-Time (Informal)

- For the simplicity of notations, let us consider one-dimensional problems:
- For the discrete-time case, the likelihood function

$$\begin{aligned} l(x, \delta y_k) := P\left(\Delta Y_{t_k} = \delta y_k | \bar{X}_{t_k} \in dx\right) &= \frac{1}{\sqrt{2\pi\Delta t}} \exp\left(-\frac{1}{2\Delta t}(\delta y_k - h(x)\Delta t)^2\right) \\ &= \frac{1}{\sqrt{2\pi\Delta t}} \exp\left(-\frac{1}{2\Delta t}(\delta y_k)^2\right) \times \exp\left(h(x)\delta y_k - \frac{1}{2}(h(x))^2\Delta t\right) \end{aligned} \tag{5}$$

- For a given test function φ,

$$E\left[\varphi(\bar{X}_{t_k})|\Delta Y_{t_{1:k}}\right] = \frac{1}{c}\tilde{E}\left[\varphi(\bar{X}_{t_k})\tilde{Z}_k|\Delta Y_{t_{1:k}} = \delta y_{1:k}\right] \tag{6}$$

where

$$\begin{aligned} \tilde{Z}_k &= \exp\left(h(\bar{X}_{t_k})\Delta Y_k - \frac{1}{2}h(\bar{X}_{t_k})^2\Delta t\right) \\ &\approx \exp\left(\int_{t_k}^{t_{k+1}} h(X_t)dY_t - \frac{1}{2}\int_{t_k}^{t_{k+1}} h(X_t)^2 dt\right) \end{aligned} \tag{7}$$

Discrete-Time v.s. Continuous-Time (Informal)

- $\tilde{E}$ means that the expectation is taken under the probability measure $\tilde{P}$ such that $X_{t_k}|\Delta Y_{t_{1:(k-1)}}$ and ΔY_{t_k} are mutually independent, and $\Delta Y_{t_k} \sim \mathcal{N}(0, \Delta t)$.
- $\tilde{Z}_k$ is the Radon-Nikodym derivative between the probability measures P and $\tilde{P}$.
- The denominator in the Bayesian formula

$$\int l(x, \delta y_k) p_{k|k-1}(x|\delta y_{1:(k-1)}) dx = \tilde{E}[\tilde{Z}_k | \Delta Y_{t_{1:k}} = \delta y_{1:k}] \tag{8}$$

Girsanov's Theorem

- Let M be a continuous martingale, and let Z be the associated exponential martingale:

$$Z_t = \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right). \tag{9}$$

- If Z is a uniformly integrable martingale, then a new measure Q, equivalent to P, may be defined by

$$\frac{dQ}{dP} = Z_\infty. \tag{10}$$

- Furthermore, if X is a continuous P local martingale, then $X_t - \langle X, M\rangle_t$ is a Q-local martingale.

Girsanov's Theorem

- Let W_t be a P-Brownian motion and define Q as in Girsanov's Theorem, the n

$$\tilde{W}_t = W_t - \langle W, M \rangle_t \tag{11}$$

  is a Q-Brownian motion.

- Let $\{B^i\}_{t \geq 0}$ be continuous local martingales starting from zero for $i = 1, \cdots, n$. Then $B_t = (B_t^1, \cdots, B_t^n)$ is a Brownian motion with respect to $(\Omega, \mathcal{F}, P)$ adapted to the filtration $\mathcal{F}_t$, if and only if

$$\langle B^i, B^j \rangle_t = \delta_{ij} t, \ \forall \ i, j \in \{1, 2, \cdots, n\}. \tag{12}$$

Change-of-Probability-Measure Method

- Consider the filtering problem

$$\begin{cases} dX_t = f(X_t)dt + \sigma(X_t)dV_t \\ dY_t = h(X_t)dt + dW_t \end{cases} \quad t \in [0, T]. \tag{13}$$

- Let $Z = \{Z_t, t > 0\}$ be the process defined by

$$Z_t = \exp\left(-\sum_{i=1}^{m} \int_0^t h^i(X_s)dW_s^i - \frac{1}{2}\sum_{i=1}^{m} \int_0^t h^i(X_s)^2 ds \right) \tag{14}$$

- Additional properties are required such that Z is a martingale.

Change-of-Probability-Measure Method

- For fixed $t \geq 0$, Z_t introduce a probability measure $\tilde{P}^t$ on $\mathcal{F}_t$ by specifying its Radon-Nikodym derivative

$$\left.\frac{d\tilde{P}^t}{dP}\right|_{\mathcal{F}_t} = Z_t \tag{15}$$

- Assume that we can generate $\{\tilde{P}^t : t \geq 0\}$ to a measure $\tilde{P} = \tilde{P}^\infty$ on $\mathcal{F}_\infty = \sigma\left(\cup_{0\leq t<\infty}\mathcal{F}_t\right)$.

Theorem (Proposition 3.13)

If Z is a martingale, then under $\tilde{P}$, the observation process Y is a Brownian motion independent of X; additionally, the law of X under $\tilde{P}$ is the same as its law under P.

Change-of-Probability-Measure Method

- Let $\tilde{Z} = \{\tilde{Z}_t, t \geq 0\}$ be the process defined as $\tilde{Z}_t = Z_t^{-1}$ for $t \geq 0$.

$$\tilde{Z}_t = \exp\left(\sum_{i=1}^{m} \int_0^t h^i(X_s)dY_s^i - \frac{1}{2}\sum_{i=1}^{m} \int_0^t h^i(X_s)^2 ds\right) \quad (16)$$

- According to Itô's formula,

$$d\tilde{Z}_t = \sum_{i=1}^{m} \tilde{Z}_t h^i(X_t)dY_t^i \quad (17)$$

which is a stochastic differential equation under $\tilde{P}$.

Unnormalized Conditional Distribution

Theorem (Kallianpur-Striebel, Proposition 3.16)

For every bounded measurable test function $\varphi \in \mathcal{B}(\mathbb{R}^d)$*, for fixed* $t \geq 0$*,*

$$E[\varphi(X_t)|\mathcal{Y}_t] := \pi_t(\varphi) = \frac{\tilde{E}[\tilde{Z}_t\varphi(X_t)|\mathcal{Y}_t]}{\tilde{E}[\tilde{Z}_t|\mathcal{Y}_t]}, \quad a.s. \tag{18}$$

- With the Kallianpur-Striebel formula, let us define the unnormalized conditional distribution of X to be the measure-valued process $\rho = \{\rho_t : t \geq 0\}$, which is determined by the values $\rho_t(\varphi)$ for $\varphi \in \mathcal{B}(\mathbb{R}^d)$, which are given for $t \geq 0$ by

$$\rho_t(\varphi) = \tilde{E}\left[\tilde{Z}_t\varphi(X_t)|\mathcal{Y}_t\right] \tag{19}$$

The Zakai Equation

Theorem (Theorem 3.24)

If mild conditions are satisfied, then the process ρ_t satisfies the following evolution equation, called the Zakai equation,

$$\rho_t(\varphi) = \pi_0(\varphi) + \int_0^t \rho_s(A\varphi)ds + \int_0^t \rho_s(\varphi h^\top)dY_s, \ a.s., \ \forall\ t \geq 0; \tag{20}$$

for any $\varphi \in \mathcal{D}(A)$.

- In this lecture, we will first provide an informal proof of the Zakai equation in order to illustrate the basic ideas, and then we will make the proof rigorous.

The Zakai Equation—An Informal Derivation

- Notice that $\rho_t(\varphi) = \tilde{E}[\tilde{Z}_t \varphi(X_t) | \mathcal{Y}_t]$. In this informal derivation of the Zakai equation, the key insight is to apply the Itô's formula to the stochastic process $\tilde{Z}_t \varphi(X_t)$.

$$
\begin{aligned}
d\left(\tilde{Z}_t \varphi(X_t)\right) &= \tilde{Z}_t d\varphi(X_t) + \varphi(X_t) d\tilde{Z}_t + d\langle \tilde{Z}, \varphi(X) \rangle_t \\
&= \tilde{Z}_t \left(A\varphi(X_t)dt + \sum_{i=1}^{d} \frac{\partial \varphi}{\partial x_i}(X_t) \sum_{j=1}^{p} \sigma^{ij}(X_t) dV_t^j \right) \\
&\quad + \varphi(X_t) \tilde{Z}_t \sum_{i=1}^{m} h^i(X_t) dY_t^i.
\end{aligned}
\tag{21}
$$

The Zakai Equation–An Informal Derivation

- Here we use the fact that V_t and Y_t are independent, and thus $d\langle \tilde{Z}, \varphi(X)\rangle_t = 0$.
- Take conditional expectation on $\mathcal{Y}_t$ with respect to $\tilde{P}$, we obtain the desired result:

$$d\rho_t(\varphi) = \rho_t(A\varphi)dt + \rho_t(\varphi h^\top)dY_t. \tag{22}$$

The Zakai Equation–A Formal Derivation

- c.f. *Fundamentals of Stochastic Filtering*, Section 3.5, Page 62-65.

Theorem (Theorem 3.24)

If the following conditions are satisfied:

$$E\left[\int_0^t \|h(X_s)\|^2 ds\right] < \infty,\ E\left[\int_0^t Z_s\|h(X_s)\|^2 ds\right] < \infty,\ \forall\ t > 0, \tag{23}$$

$$\tilde{P}\left[\int_0^t [\rho_s(\|h\|)]^2 ds < \infty\right] = 1. \tag{24}$$

then the process ρ_t satisfies the following evolution equation, called the Zakai equation,

$$\rho_t(\varphi) = \pi_0(\varphi) + \int_0^t \rho_s(A\varphi)ds + \int_0^t \rho_s(\varphi h^\top)dY_s,\ a.s.,\ \forall\ t \geq 0; \tag{25}$$

for any $\varphi \in \mathcal{D}(A)$.

The Zakai Equation – A Formal Derivation

- The approximation of $\tilde{Z}_t$ by

$$\tilde{Z}_t^\epsilon = \frac{\tilde{Z}_t}{1 + \epsilon \tilde{Z}_t} \tag{26}$$

- Apply the Itô's rule, because $\tilde{Z}_t^\epsilon$ is bounded.
- Dominated Convergence Theorem: convergence of $\tilde{Z}_t^\epsilon$ to $\tilde{Z}_t$, as $\epsilon \to 0$.

The Kushner-Stratonovich Equation

Theorem

If the above conditions are satisfied then the conditional distribution of the signal π_t, satisfies the following evolution equation called the Kushner-Stratonovich equation:

$$\begin{aligned}\pi_t(\varphi) = \pi_0(\varphi) &+ \int_0^t \pi_s(A\varphi)ds \\ &+ \int_0^t (\pi_s(\varphi h^\top) - \pi_s(h^\top)\pi_s(\varphi))(dY_s - \pi_s(h)ds)\end{aligned} \tag{27}$$

for any $\varphi \in \mathcal{D}(A)$.

The Kushner-Stratonovich Equation

- Kallianpur-Striebel formula:

$$\pi_t(\varphi) = \rho_t(\varphi) \cdot \frac{1}{\rho_t(1)}. \tag{28}$$

- Itô's formula:

$$d\left(\frac{1}{\rho_t(1)}\right) = \frac{1}{\rho_t(1)}[-\pi_t(h^\top)dY_t + \pi_t(h^\top)\pi_t(h)dt]. \tag{29}$$

- Combining the Zakai equation.

The Innovation Process Approach

- Let $I = \{I_t : t \geq 0\}$ be the following process, called the innovation process:

$$I_t = Y_t - \int_0^t \pi_s(h)ds \tag{30}$$

Lemma (Proposition 2.30)

Under mild conditions, I_t is a $\mathcal{Y}_t$-adapted Brownian motion under the measure P.

The Innovation Process Approach

Lemma (Corollary 3.3, Martingale Representation Theorem)

Under mild conditions, every right continuous square integrable martingale $\eta = \{\eta_t : t \geq 0\}$, which is $\mathcal{Y}_t$-adapted has a representation

$$\eta_t = \eta_0 + \int_0^t \nu_s^\top dI_s. \tag{31}$$

Lemma

Define

$$N_t = \pi_t(\varphi) - \int_0^t \pi_s(A\varphi)ds, \tag{32}$$

then N is a $\mathcal{Y}_t$-adapted square integrable martingale under the probability measure P.

The Innovation Process Approach

- According to the Martingale Representation Theorem:

$$\pi_t(\varphi) = \pi_0(\varphi) + \int_0^t \pi_s(A\varphi)ds + \int_0^t \nu_s^\top dI_s \tag{33}$$

- Prove that

$$\nu_s = \pi_t(\varphi h) - \pi_s(\varphi)\pi_s(h). \tag{34}$$

- Total set ϵ_t:

$$d\epsilon_t = i\epsilon_t r_t^\top dY_t. \tag{35}$$

The Innovation Process Approach

- In fact,

$$\pi_t(\varphi) = \pi_0(\varphi) + \int_0^t \pi_s(A\varphi)ds + \int_0^t \nu_s^\top (dY_s - \pi_s(h)ds) \quad (36)$$

$$\varphi(X_t) = \varphi(X_0) + \int_0^t A\varphi(X_s)ds + M_t^\varphi \quad (37)$$

- Compare the difference between $\pi_t(\varphi)\epsilon_t$ and $\varphi(X_t)\epsilon_t$.

Comparison between Two Methods

- Conditions for the change-of-probability-measure method:

$$E\left[\int_0^t \|h(X_s)\|^2 ds\right] < \infty,\ E\left[\int_0^t Z_s\|h(X_s)\|^2 ds\right] < \infty,\ \forall\ t > 0, \quad (38)$$

$$\tilde{P}\left[\int_0^t [\rho_s(\|h\|)]^2 ds < \infty\right] = 1. \quad (39)$$

- Conditions for the innovation process approach:

$$E\left[\int_0^t \|h(X_s)\|^2 ds\right] < \infty,\ P\left(\int_0^t \|\pi_s(h)\|^2 ds < \infty\right) = 1. \quad (40)$$

The Correlated Noise Framework

- If the state process is governed by

$$dX_t = b(X_t)dt + \sigma(X_t)dV_t + \bar{\sigma}(X_t)dW_t, \tag{41}$$

then,

$$\begin{aligned} d\pi_t(\varphi) =& \pi_t(A\varphi)dt + (\pi_t((h^\top + B^\top)\varphi) - \pi_t(h^\top)\pi_t(\varphi)) \\ & \times (dY_t - \pi_t(h)dt) \end{aligned} \tag{42}$$

$$\rho_t(\varphi) = \rho_0(\varphi) + \int_0^t \rho_s(A\varphi)ds + \int_0^t \rho_s((h^\top + B^\top)\varphi)dY_s \tag{43}$$

where

$$B_k = \sum_{i=1}^{d} \bar{\sigma}_{ik} \frac{\partial}{\partial x_i}, \ \ k = 1, \cdots, m. \tag{44}$$