Manifold Learning and Artificial Intelligence
Lecture 7

Generative AI (5)

General SDE, Kolmogorov Differential Equation and the probability flow ODE

Momiao Xiong, University of Texas School of Public Health

- Time: 9:00 pm, US East Time, 01/14/2023
- 10:00 am, Beijing Time. 01/15/2023
- Zoom

https://uwmadison.zoom.us/w/93316139423?tk=wfbmsTfN2fgERto_HI1WKtBzh94d3HO02XV
Cexqd8.DQMAAAAVuhNJnxZ2dGVCbllIYlR3ZWJBNHl5LXlYMjBnAAAAAAAAAAAAAAAAAAAA
AAAAAA&pwd=Q0NVWFYvRFg5RmxCNkwxMmYrbW41dz09

Github Address: https://ai2healthcare.github.io/

- **General Stochastic Differential Equations**
- **Reverse – Time SDE**
- **Divergence of Matrix-valued Functions**
- **Itô’s Formula**
- **Kolmogorov Backward Differential Equation**
- **Kolmogorov Forward Differential Equation**
- **The Probability Flow Ordinary Differential Equation**

Discuss some unclear notation in the paper:

Song et al. 2021. SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

17. General Stochastic Differential Equations

17.1. Definition

- **Forward SDE**

$$d\boldsymbol{X}(t) = \boldsymbol{\mu}(\boldsymbol{X}(t), t)dt + \boldsymbol{\sigma}(\boldsymbol{X}(t), \boldsymbol{t})\mathrm{d}\,\boldsymbol{W}(t)$$

$$\boldsymbol{X}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_d(t) \end{bmatrix}, \boldsymbol{\mu}(\boldsymbol{X}(t), t) = \begin{bmatrix} \mu_1\left(\boldsymbol{X}(t)\right), t \\ \vdots \\ \mu_d(X(t), t) \end{bmatrix}$$

$$\boldsymbol{\sigma}(\boldsymbol{X}(t), \boldsymbol{t}) = \begin{bmatrix} \sigma_{\mathbf{11}}(\boldsymbol{X}(t), t) & \cdots & \sigma_{1d}(\boldsymbol{X}(t), t) \\ \vdots & \vdots & \vdots \\ \sigma_{d1}(\boldsymbol{X}(t), t) & \cdots & \sigma_{dd}(\boldsymbol{X}(t), t) \end{bmatrix} = \begin{bmatrix} \sigma_{1.}(\boldsymbol{X}(t), t) \\ \vdots \\ \sigma_{d.}(\boldsymbol{X}(t), t) \end{bmatrix}, \boldsymbol{W}(t) = \begin{bmatrix} W_1(t) \\ \vdots \\ W_d(t) \end{bmatrix}$$

- **Reverse – Time SDE**

- **Divergence of Matrix-valued Functions**

$$F(x) = \begin{bmatrix} f_{11}(x) & \cdots & f_{1d}(x) \\ \vdots & \vdots & \vdots \\ f_{d1}(x) & \cdots & f_{dd}(x) \end{bmatrix} = \begin{bmatrix} f_{1.}(x) \\ \vdots \\ f_{d.}(x) \end{bmatrix}, \text{ where } f_{i.}(x) = [f_{i1}(x) \quad \cdots \quad f_{id}(x)]$$

$$\nabla \cdot f_{i.}(x) = \frac{\partial f_{i1}(x)}{\partial x_1} + \cdots + \frac{\partial f_{id}(x)}{\partial x_d}$$

$$\nabla \cdot F(x) = \begin{bmatrix} \nabla \cdot f_{1.}(x) \\ \vdots \\ \nabla \cdot f_{d.}(x) \end{bmatrix}$$

• Reverse – Time SDE

$$\boldsymbol{dX} = \left\{\boldsymbol{\mu}(\boldsymbol{X},t) - \boldsymbol{\nabla}\cdot\left[\boldsymbol{\sigma}(\boldsymbol{X},t)\boldsymbol{\sigma}(\boldsymbol{X},t)^{\boldsymbol{T}}\right] - \boldsymbol{\sigma}(\boldsymbol{X},t)\boldsymbol{\sigma}(\boldsymbol{X},t)^{\boldsymbol{T}}\nabla_{\boldsymbol{x}}\log P(\boldsymbol{X},t)\right\}\boldsymbol{dt} + \boldsymbol{\sigma}(\boldsymbol{X},t)d\overline{\boldsymbol{W}}(\mathrm{t})$$

where

$$\boldsymbol{\sigma}(\boldsymbol{X},t)\boldsymbol{\sigma}(\boldsymbol{X},t)^{\boldsymbol{T}} = \begin{bmatrix} \sigma_{1.}(\boldsymbol{X}(t),t)\sigma_{1.}(\boldsymbol{X}(t),t)^T & \cdots & \sigma_{1.}(\boldsymbol{X}(t),t)\sigma_{d.}(\boldsymbol{X}(t),t)^T \\ \vdots & \vdots & \vdots \\ \sigma_{d.}(\boldsymbol{X}(t),t)\sigma_{1.}(\boldsymbol{X}(t),t)^T & \cdots & \sigma_{d.}(\boldsymbol{X}(t),t)\sigma_{d.}(\boldsymbol{X}(t),t)^{\boldsymbol{T}} \end{bmatrix}$$

$$\nabla\cdot\boldsymbol{\sigma}(\boldsymbol{X},t)\boldsymbol{\sigma}(\boldsymbol{X},t)^{\boldsymbol{T}} = \begin{bmatrix} \sum_{j=1}^{\boldsymbol{d}}\frac{\partial}{\partial x_j}\sigma_{1.}(\boldsymbol{X}(t),t)\sigma_{j.}(\boldsymbol{X}(t),t)^T \\ \vdots \\ \sum_{j=1d}\frac{\partial}{\partial x_j}\sigma_{d.}(\boldsymbol{X}(t),t)\sigma_{j.}(\boldsymbol{X}(t),t)^T \end{bmatrix}$$

$$\boldsymbol{\sigma}(\boldsymbol{X},\boldsymbol{t}) = \begin{bmatrix} \sigma_{1.}(\boldsymbol{X}(t),t) \\ \vdots \\ \sigma_{d.}(\boldsymbol{X}(t),t) \end{bmatrix}$$

$$\sigma(X,t)^T = \begin{bmatrix} \sigma_{1.}(\boldsymbol{X}(t),t)^T & \cdots & \sigma_{d.}(\boldsymbol{X}(t),t)^{\boldsymbol{T}} \end{bmatrix}$$

Brian D O Anderson. Reverse-time diffusion equation models. Stochastic Process. Appl., 12(3): 313–326, May 1982.

- **Example 1**

Forward SDE

$$dx = \mu(x,t)dt + \sigma(t)dw$$

Reverse – Time SDE

$$dx = \big(\mu(x,t) - \sigma(t)^2 \nabla_x \log P(x,t)\big)dt + \sigma(t)d\overline{w}$$

- **Example 2**

Forward SDE

$$dx = \mu(x,t)dt + \sigma(x,t)dw$$

Reverse – Time SDE

$$dx = \left(\mu(x,t) - \frac{\partial \sigma(x,t)^2}{\partial x} - \sigma(x,t)^2 \nabla_x \log P(x,t)\right)dt + \sigma(x,t)d\overline{w}$$

17.2. Itô's Formula for High Dimensional Diffusion Process

- **Purpose**

Ito's Lemma is a key component in the Ito Calculus, used to determine the derivative of a time-dependent function of a stochastic process. It performs the role of the chain rule and Taylor expansion in a stochastic setting, analogous to the chain rule in ordinary differential calculus. Ito's Lemma is a cornerstone of quantitative finance and it is intrinsic to the derivation of the Black-Scholes equation for contingent claims (options) pricing.

Calculate the moments of diffusion process.

Itô’s Formula for High Dimensional Diffusion Process

- **Let** $\boldsymbol{X}(\boldsymbol{t}) \in R^d, \boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}), \boldsymbol{t}) \in R^d, \boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}), \boldsymbol{t}) \in \boldsymbol{R}^{\boldsymbol{d}\times\boldsymbol{d}}, \boldsymbol{W}(\boldsymbol{t}) \in \boldsymbol{R}^{\boldsymbol{d}}$
  - **High Dimensional Diffusion Model**

$$\boldsymbol{dX}(\boldsymbol{t}) = \boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}), \boldsymbol{t})\,\mathrm{dt} + \boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}), \boldsymbol{t})\mathrm{d}\,\boldsymbol{W}(\boldsymbol{t}) \qquad \textbf{(1)}$$

- **Itô’s Formula**

Consider twice differentiable scalar function $f(t, \boldsymbol{X})$. Define Hessian Matrix:

$$H_{xx}f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

$$\boldsymbol{df(t,X_t) = \frac{\partial f}{\partial t}dt + (\nabla_x f)^T dX_t + \frac{1}{2}(dX_t)^T(H_{xx}f)dX_t}$$

$$\boldsymbol{= \left\{\frac{\partial f}{\partial t} + (\nabla_x f)^T \mu(X(t),t) + \frac{1}{2}Tr\left[\sigma(X(t),t)^T H_{xx} f \sigma(X(t),t)\right]dt\right\} + (\nabla_x f)^T\, \sigma(X(t),t)\mathrm{d}W_t}$$

- **Proof**

Recall that $\boldsymbol{dW} \approx \mathbf{0}$ and $\boldsymbol{dW(dW)^T} \approx \boldsymbol{I}\, dt, dt^2 \approx 0, \boldsymbol{dtdW} \approx \mathbf{0}$,

Using Taylor expansion, we obtain

$$df = f(t+dt, \boldsymbol{X}_t + d\boldsymbol{X}_t) - \mathrm{f}(\mathrm{t}, \boldsymbol{X}_t) = \frac{\partial f}{\partial t}dt + (\nabla_x f)^T d\boldsymbol{X}_t + \frac{1}{2}(d\boldsymbol{X}_t)^T(H_{xx}f)d\boldsymbol{X}_t \quad (2)$$

Substituting equation (1) into equation (2) yields

$$df = \frac{\partial f}{\partial t}dt + (\nabla_x f)^T(\boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\,\mathrm{dt} + \boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\mathrm{d}\,\boldsymbol{W}(\boldsymbol{t}))$$

$$\frac{1}{2}(\boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\,\mathrm{dt} + \boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\mathrm{d}\,\boldsymbol{W}(\boldsymbol{t}))^T\,(H_{xx}f)(\boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\,\mathrm{dt} + \boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\mathrm{d}\,\boldsymbol{W}(\boldsymbol{t}))$$

$$= \frac{\partial f}{\partial t} dt + (\nabla_x f)^T (\boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\ \mathrm{dt} + \boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\mathrm{d}\,\boldsymbol{W}(\boldsymbol{t})) + \frac{1}{2}\boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})^T\ (H_{xx}f)\ \boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})(dt)^{\mathbf{2}}$$

$$+ \frac{1}{2} dt\boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})^T\ (H_{xx}f)\ (\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\mathrm{d}\,\boldsymbol{W}(\boldsymbol{t})) + \frac{1}{2}(\mathrm{d}\,\boldsymbol{W}(\boldsymbol{t})^T\left(\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\right)^T (H_{xx}f)\boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})dt$$

$$+ \frac{1}{2} Tr(\mathrm{d}\,\boldsymbol{W}(\boldsymbol{t}))^T\left(\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\right)^T (H_{xx}f)(\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\mathrm{d}\,\boldsymbol{W}(\boldsymbol{t}))$$

$$= \frac{\partial f}{\partial t} dt + (\nabla_x f)(^T \boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\ \mathrm{dt} + \boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\mathrm{d}\,\boldsymbol{W}(\mathrm{t})) + 0 + 0 + 0 + 0$$

$$+ \frac{1}{2} Tr(\left(\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\right)^T\ (H_{xx}f)\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\boldsymbol{dW}(\boldsymbol{t})\left(\boldsymbol{dW}(\boldsymbol{t})\right)^{\boldsymbol{T}})$$

$$\boldsymbol{dW}(\boldsymbol{t})\left(\boldsymbol{dW}(\boldsymbol{t})\right)^{\boldsymbol{T}} \approx Idt$$

$$= \frac{\partial f}{\partial t} dt + (\nabla_x f)^T (\boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\ \mathrm{dt} + \boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\mathrm{d}\,\boldsymbol{W}(\mathrm{t})) + 0 + 0 + 0 + 0$$

$$+ \frac{1}{2} Tr(\left(\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\right)^T\ (H_{xx}f)\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\mathrm{dt}$$

$$= \left[\frac{\partial f}{\partial t} + (\nabla_x f)^T\ \boldsymbol{\mu}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t}) + \frac{1}{2} Tr(\left(\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\right)^T\ (H_{xx}f)\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t}))\right]\mathrm{dt} + (\boldsymbol{\nabla_x f})^{\boldsymbol{T}}\boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})\mathrm{d}\,\boldsymbol{W}($$

A -functional Itô’s formula and its applications in mathematical finance,

stochastic Processes and their Applications Volume 148, June 2022, Pages 299-323

Wright-Fisher models, mutation frequency is modeled as a diffusion process.

Karlin and Taylor (1981) A second Course in Stochastic Process

Mutation effect in identifying concerned variant is a function of mutation frequency and can be calculated using Ito’s formula.

Solution to SDE is a Markov Process

- **Solution to SDE (1) is given by**

$$X(t) - X(s) = \int_s^t \mu(x(\tau), \tau) d\tau + \int_s^t \sigma(x(\tau), \tau) dW(\tau)$$

which can be calculated using only $X(u), s \leq u \leq t$, and do not require knowing $X(u), u \leq s$.

17.3. Kolmogorov Backward Differential Equation

- **Definition of Function $u(t, X)$ and Derivation**

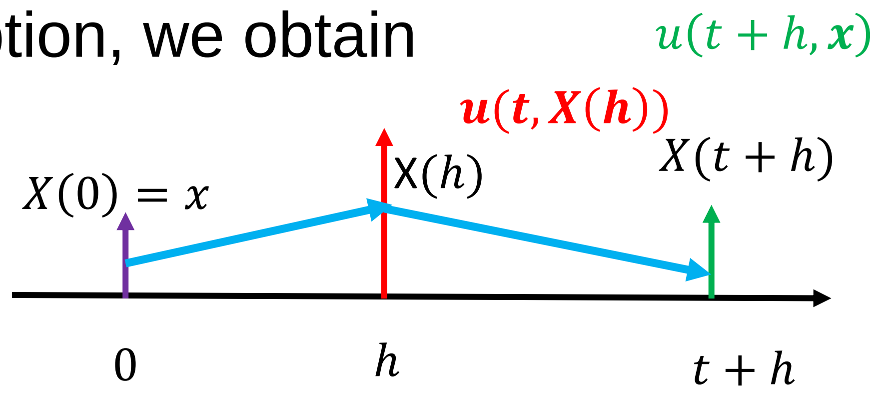
Define

$$u(t\ X), = E[g(\boldsymbol{X}(t))|X(0) = X] \quad (3)$$

By Markov property and stationary assumption, we obtain

(4)

$$E[g(\boldsymbol{X}(h+t)|\boldsymbol{X}(h)] = u(t, \boldsymbol{X}(h)) \quad (5)$$



Using conditioning, we obtain

$$u(t+h, X) = E[g(\boldsymbol{X}(h+t))\big|\boldsymbol{X}(0) = X] = E[E[\mathbf{g}(\mathbf{X}\,(\mathrm{t}+\mathrm{h}))|\mathbf{X}\,(\mathrm{h})]\ \mathbf{X}(0) = \mathrm{X}] \quad (6)$$

Substituting equation (5) into equation (6) yields

$$u(t+h, X) = E[u(t, \boldsymbol{X}(h))|\boldsymbol{X}(0) = X] \quad (7)$$

Let $\Delta\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{h}) - X$

Since the probability of $\Delta \boldsymbol{X} > \boldsymbol{\varepsilon}$ is small, we can assume that $\Delta \boldsymbol{X}$ is small.

Now by definition of derivative and using equations (3) and (7), we Have

$$u(t+h, X) - u(t, X) = E[u(t, X + \Delta \boldsymbol{X}) - u(t, X) | \boldsymbol{X}(0) = X] \quad \textbf{(8)}$$

Using Taylor expansion, we have

$$\boldsymbol{x} + \Delta \boldsymbol{X} = \boldsymbol{X}(\boldsymbol{h})$$

$$u(t, X + \Delta \boldsymbol{X}) - u(t, X) = (\nabla_x u)^T \Delta \boldsymbol{X} + \frac{1}{2} (\Delta \boldsymbol{X})^T \frac{\partial^2 u}{\partial x \partial x^T} \Delta \boldsymbol{X} \quad \textbf{(9)}$$

$$= (\nabla_x u)^T (\boldsymbol{\mu}(\boldsymbol{X}, \boldsymbol{t}) \boldsymbol{dt} + \boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t}), \boldsymbol{t}) \boldsymbol{dW}) + \frac{1}{2} (\mu(X, t) dt + \sigma(X(t), t) dW)^{\boldsymbol{T}}$$

$$\frac{\partial^2 u}{\partial x \partial x^T} (\mu(X, t) dt + \sigma(X(t), t) dW)$$

Substituting equation (1) into equation (9)

$$= (\nabla_x u)^T \mu(X, t) dt + (\nabla_x u)^T \sigma(X(t), t) dW + \frac{1}{2} (\mu(X, t) dt)^T \frac{\partial^2 \boldsymbol{u}}{\partial x \partial x^T} \mu(X, t) dt$$

$$+ 2 * \frac{1}{2} \mu(X, t) dt \frac{\partial^2 u}{\partial x \partial x^T} \sigma(X(t), t) dW + \frac{1}{2} (\sigma(X(t), t) dW)^T \frac{\partial^2 u}{\partial x \partial x^T} \sigma(X(t), t) dW \quad \textbf{(10)}$$

Using trace property of matrix, we obtain

$$(\sigma(X(t),t)dW)^T \frac{\partial^2 u}{\partial x \partial x^T} \sigma(X(t),t)dW = Tr\left((\sigma(X(t),t)dW)^T \frac{\partial^2 u}{\partial x \partial x^T} \sigma(X(t),t)dW\right)$$

$$= Tr\left(\frac{\partial^2 u}{\partial x \partial x^T} \sigma(X(t),t)dW(\sigma(X(t),t)dW)^T\right) \qquad \textbf{(11)}$$

Substituting equation (11) into equation (10) yields

$$u(t, X + \Delta\boldsymbol{X}) - u(t,X) = (\nabla_x u)^T \mu(X,t)dt + (\nabla_x u)^T \sigma(X(t),t)d\boldsymbol{W} + \frac{1}{2}(\mu(X,t)dt)^T \frac{\partial^2 \boldsymbol{u}}{\partial x \partial x^T} \mu(X,t)dt$$

$$+2 * \frac{1}{2} \mu(X,t)dt \frac{\partial^2 u}{\partial x \partial x^T} \sigma(X(t),t)d\boldsymbol{W} + \frac{1}{2} Tr\left(\frac{\partial^2 \boldsymbol{u}}{\partial x \partial x^T} \sigma(X(t),t)\boldsymbol{dW}(\boldsymbol{dW})^T \sigma(X(t),t)^T\right)$$

Recall that $\boldsymbol{dW} \approx \boldsymbol{0}$ and $\boldsymbol{dW}(\boldsymbol{dW})^{\boldsymbol{T}} \approx \boldsymbol{I}\, dt, dt^2 \approx 0, \boldsymbol{dtdW} \approx \boldsymbol{0}$,

Taking expectations on both sides of above equation, we obtain

$$E[u(t, X + \Delta\boldsymbol{X}) - u(t,X)|\mathbf{X}(0) = X] = (\nabla_x u)^T \mu(X,t)dt + \frac{1}{2} Tr(\sigma(X(t),t)\sigma(X(t),t)^T \frac{\partial^2 u}{\partial x \partial x^T}\, \mathrm{dt}$$

Using equation (8) and above equation yields

$$u(t+h, X) - u(t,X) = \left[(\nabla_x u)^T \mu(X,t) + \frac{1}{2} Tr(\sigma(X(t),t)\sigma(X(t),t)^T \frac{\partial^2 u}{\partial x \partial x^T}\right] dt \qquad \textbf{(12)}$$

which implies that

$$\lim_{dt\to 0}\frac{u(t+dt,\boldsymbol{x})-u(t,\boldsymbol{x})}{dt}=(\nabla_x u)^T\mu(X,t)+\frac{1}{2}Tr(\sigma(X(t),t)\sigma(X(t),t)^T\frac{\partial^2 u}{\partial x\partial x^T},$$

or

$$\boldsymbol{\frac{\partial u}{\partial t}=\left(\mu(X,t)\right)^T\nabla_x u+\frac{1}{2}Tr(\sigma(X(t),t)\sigma(X(t),t)^T\frac{\partial^2 u}{\partial x\partial x^T}} \quad \textbf{(13)}$$

For one dimensional diffusion process, we have

$$\boldsymbol{\frac{\partial u}{\partial t}=\mu(X,t)\frac{\partial u}{\partial x}+\frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2}} \quad \textbf{(14)}$$

Define

$$g(Z(t))=\begin{cases}1 & Z\le Y\\ 0 & Otherwise\end{cases}$$

$$\boldsymbol{u(t\ X)=E[g\left(Z(t)\right)|X(0)=X]}$$

Then, $u(t,X)=P(t,X,Y)$

Applying equation (13), we obtain **Kolmogorov backward equation for transition probability**

- **High Dimensional Transitional Probability**

$$\frac{\partial P(t,X,Y)}{\partial t} = \left(\mu(X,t)\right)^T \nabla_x P(t,X,Y) + \frac{1}{2} Tr(\sigma(X(t),t)\sigma(X(t),t)^T \frac{\partial^2 P(t,X,Y)}{\partial x \partial x^T} \quad \textbf{(15)}$$

- **One Dimensional Transition Probability**

$$\frac{\partial P(t,x,y)}{\partial t} = \mu(x,t)\frac{\partial P(t,x,y)}{\partial x} + \frac{1}{2}\sigma(X,t)^2 \frac{\partial^2 P(t,x,y)}{\partial x^2} \quad \textbf{(16)}$$

- **High Dimensional Marginal Probability**

$$\frac{\partial P(t,X)}{\partial t} = \left(\mu(X,t)\right)^T \nabla_x P(t,X) + \frac{1}{2} Tr(\sigma(X(t),t)\sigma(X(t),t)^T \frac{\partial^2 P(t,X)}{\partial x \partial x^T}$$

Franco Flandoli

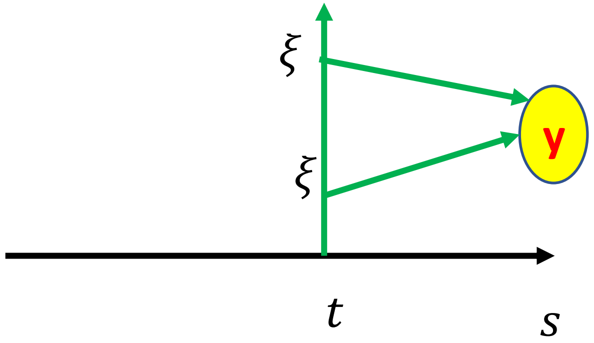
Numerical computation of probabilities for nonlinear SDEs in high dimension using Kolmogorov equation

Applied Mathematics and Computation Volume 436, 1 January 2023, 127520

Kolmogorov Forward Differential Equation

- **Pertinent variables are t and y, the state variable y a time t rather initial value x.**

Let $\varphi(t, y)$ be arbitrary smooth function satisfying



$$\varphi(t+s, y) = \int \varphi(t, \xi) p(s, \xi, y) d\xi \qquad (17)$$

Differentiating both dides of equation (17) yields

$$\frac{\partial \varphi(t+s, y)}{\partial t} = \frac{\partial \varphi(t+s, y)}{\partial s} = \int \varphi(t, \xi) \frac{\partial p(s, \xi, y)}{\partial s} d\xi \qquad (18)$$

Substituting equation (15) into equation (18) yields

$$\frac{\partial \varphi(t+s, y)}{\partial t} = \int \varphi(t, \xi) \left[\left(\mu(\xi, s)\right)^T \nabla_\xi P(s, \boldsymbol{\xi}, \boldsymbol{y}) + \frac{1}{2} Tr(\sigma(\xi(s), s) \sigma(\xi(s), s)^T \frac{\partial^2 P(s, \xi, y)}{\partial \xi \partial \xi^T} \right] d\xi$$

(19)

- **Note that**

$$\left(\mu(\xi,s)\right)^T \nabla_\xi P(s,\boldsymbol{\xi},\boldsymbol{y}) = \sum_{i=1}^{d} \mu_i(\xi,s) \frac{\partial}{\partial \xi_i} P(s,\boldsymbol{\xi},\boldsymbol{y}) \tag{20}$$

$$\sigma(\xi(s),s)\sigma(\xi(s),s)^T = \begin{bmatrix} \sigma_{11}(\xi,s) & \cdots & \sigma_{1d}(\xi,s) \\ \vdots & \cdots & \vdots \\ \sigma_{d1}(\xi,s) & \cdots & \sigma_{dd}(\xi,s) \end{bmatrix} \begin{bmatrix} \sigma_{11}(\xi,s) & \cdots & \sigma_{d1}(\xi,s) \\ \vdots & \vdots & \vdots \\ \sigma_{1d}(\xi,s) & \cdots & \sigma_{dd}(\xi,s) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=1}^{d} \sigma_{11}^2(\xi,s) & \cdots & \sum_{k=1}^{d} \sigma_{1k}(\xi,s)\sigma_{dk}(\xi,s) \\ \vdots & \vdots & \vdots \\ \sum_{k=1}^{d} \sigma_{dk}(\xi,s)\sigma_{1k}(\xi,s) & \cdots & \sum_{k=1}^{d} \sigma_{dk}^2(\xi,s) \end{bmatrix} \tag{21}$$

$$Tr(\sigma(\xi(s),s)\sigma(\xi(s),s)^T \frac{\partial^2 P(\boldsymbol{s},\boldsymbol{\xi},\boldsymbol{y})}{\partial \xi \partial \xi^T}) = \sum_{j=1}^{d} \sum_{i=1}^{d} \sum_{k=1}^{d} \sigma_{jk}(\,(\xi(s),s)\sigma_{ik}((\xi(s),s) \frac{\partial P(s,\boldsymbol{\xi},\boldsymbol{y})}{\partial \xi_i \partial \xi_j}$$

(22)

- **Substituting equations (20)-(22) into equation (19) yields**

$$\frac{\partial\varphi(t+s,y)}{\partial t}=\sum_{i=1}^{d}\int[\varphi(t,\xi)\mu_i(\xi,s)]\frac{\partial}{\partial\xi_i}P(s,\boldsymbol{\xi},\boldsymbol{y})\mathrm{d}\xi_i$$

$$+\sum_{j=1}^{d}\sum_{i=1}^{d}\sum_{k=1}^{d}\int\left[\frac{1}{2}\varphi(t,\xi)\sigma_{jk}(\xi,s)\sigma_{ik}(\xi,s)\right]\frac{\partial}{\partial\xi_i\partial\xi_j}P(s,\boldsymbol{\xi},\boldsymbol{y})d\,\xi_i\mathrm{d}\xi_j \qquad \textbf{(23)}$$

Note that $P(s,+\infty,y)=P(s,-\infty,y)=0$

$$\int\varphi(t,\xi)\,\mu_i(\xi,s)\frac{\partial}{\partial\xi_i}P(s,\boldsymbol{\xi},\boldsymbol{y})\mathrm{d}\xi_i=\varphi(t,\xi)\mu_i(\xi,s)\boldsymbol{P(s,\xi,y)}\Big|-\int P(s,\boldsymbol{\xi},\boldsymbol{y})\frac{\partial}{\partial\xi_i}[\varphi(t,\xi)\mu_i(\xi,s)]d\xi_i$$

$$=-\int P(s,\boldsymbol{\xi},\boldsymbol{y})\frac{\partial}{\partial\xi_i}[\varphi(t,\xi)\mu_i(\xi,s)]\,d\xi_i \qquad \textbf{(24)}$$

$$\int\left[\frac{1}{2}\varphi(t,\xi)\sigma_{jk}(\xi,s)\sigma_{ik}(\xi,s)\right]\frac{\partial}{\partial\xi_i\partial\xi_j}P(s,\boldsymbol{\xi},\boldsymbol{y})d\xi_i\mathrm{d}\xi_j=\int P(s,\boldsymbol{\xi},\boldsymbol{y})\frac{\partial}{\partial\xi_i\partial\xi_j}\left[\frac{1}{2}\varphi(t,\xi)\sigma_{jk}(\xi,s)\sigma_{ik}(\xi,s)\right]d\xi_i\mathrm{d}\xi_j$$

$$\textbf{(25)}$$

Integration by parts

Substituting equations (24) and (25) into equation (23) yields

$$\frac{\partial \varphi(t+s,y)}{\partial t} = -\sum_{i=1}^{d} \int P(s,\boldsymbol{\xi},\boldsymbol{y}) \frac{\partial}{\partial \xi_i} [\varphi(t,\xi)\mu_i(\xi,s)]\, d\xi_i$$

$$+\sum_{j=1}^{d}\sum_{i=1}^{d}\sum_{k=1}^{d} \int P(s,\boldsymbol{\xi},\boldsymbol{y}) \frac{\partial}{\partial \xi_i \partial \xi_j} \left[\frac{1}{2}\varphi(t,\xi)\sigma_{jk}(\xi,s)\sigma_{ik}(\xi,s)\right] d\xi_i \mathrm{d}\xi_j \qquad \textbf{(26)}$$

When $s \to 0$, we obtain

$$P(s,\boldsymbol{\xi},\boldsymbol{y}) \to P(\boldsymbol{\xi}=\boldsymbol{y}) = 1, \qquad \boldsymbol{\mu_i(\xi,0) = \mu_i(y)}$$

$$\int P(s,\boldsymbol{\xi},\boldsymbol{y}) \frac{\partial}{\partial \xi_i} [\varphi(t,\xi)\mu_i(\xi,s)]\, d\xi_i \to \frac{\partial}{\partial y_i}[\varphi(t,\boldsymbol{y})\mu_i(\boldsymbol{y})] \qquad \textbf{(27})$$

$$\int P(s,\boldsymbol{\xi},\boldsymbol{y}) \frac{\partial}{\partial \xi_i \partial \xi_j} \left[\frac{1}{2}\varphi(t,\xi)\sigma_{jk}(\xi,s)\sigma_{ik}(\xi,s)\right] d\xi_i \mathrm{d}\xi_j \to \frac{\partial}{\partial y_i \partial y_j}\left[\frac{1}{2}\varphi(t,\boldsymbol{y})\sigma_{jk}(\boldsymbol{y})\sigma_{ik}(\boldsymbol{y})\right]$$

(28)

Substituting equations (27) and (28) into equation (26) yields

$$\frac{\partial \varphi(t,y)}{\partial t} = -\sum_{i=1}^{d} \frac{\partial}{\partial y_i}[\varphi(t,\boldsymbol{y})\mu_i(\boldsymbol{y})] + \sum_{j=1}^{d} \sum_{i=1}^{d} \sum_{k=1}^{d} \frac{\partial}{\partial y_i \partial y_j}\left[\frac{1}{2}\varphi(t,\boldsymbol{y})\sigma_{jk}(\boldsymbol{y})\sigma_{ik}(\boldsymbol{y})\right]$$

(29)

Substituting $\varphi(t,\boldsymbol{y}) = p(t,\boldsymbol{x},\boldsymbol{y})$ into equation (29), we obtain

$$\frac{\partial p(t,\boldsymbol{x},\boldsymbol{y})}{\partial t} = -\sum_{i=1}^{d} \frac{\partial}{\partial y_i}[p(t,\boldsymbol{x},\boldsymbol{y})\mu_i(\boldsymbol{y})] + \sum_{j=1}^{d} \sum_{i=1}^{d} \sum_{k=1}^{d} \frac{\partial}{\partial y_i \partial y_j}\left[\frac{1}{2}p(t,\boldsymbol{x},\boldsymbol{y})\sigma_{jk}(\boldsymbol{y})\sigma_{ik}(\boldsymbol{y})\right]$$

(30)

When $d = 1$, equation (30) is reduced to

$$\frac{\partial p(t,x,y)}{\partial t} = -\frac{\partial}{\partial y}[p(t,x,y)\mu(y)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left[p(t,x,y)\sigma^2(y)\right]$$

(31)

Marginal Kolmogorov Forward Equations

$$\frac{\partial p(t,\boldsymbol{x})}{\partial t} = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}[p(t,\boldsymbol{x})\mu_i(x)] + \sum_{j=1}^{d}\sum_{i=1}^{d}\sum_{k=1}^{d} \frac{\partial}{\partial x_i \partial x_j}\left[\frac{1}{2}p(t,\boldsymbol{x})\sigma_{jk}(x)\sigma_{ik}(x)\right]$$

(A1)

Let

$$dX = \mu(X,t)dt + \sigma(X,t)dW, \qquad \sigma(X,t) = 0$$

Then,

$$\boldsymbol{\frac{\partial p(t,x)}{\partial t} = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}[p(t,x)\mu_i(x)]} \qquad \textbf{(A2)}$$

(A3)

$$\boldsymbol{dX = \mu(X,t)dt}$$

18. The Probability Flow Ordinary Differential Equation (ODE)

18.1. Definition

The probability flow ODE is defined as

$$\frac{dX}{dt} = \mu(X,t) - \frac{1}{2}\nabla\cdot\left[\sigma(X,t)\sigma(X,t)^T\right] - \frac{1}{2}\sigma(X,t)\sigma(X,t)^T\nabla_x \log P(X,t) \quad \textbf{(32)}$$

18.2. Derivation

Recall that Kolmogorov's forward equation (Fokker-Planck equation) for marginal density function of the process $X(t)$ defined in equation (1) is given by

$$\frac{\partial p(t,X)}{\partial t} = -\sum_{i=1}^{d}\frac{\partial}{\partial x_i}\left[p(t,X)\mu_i(t,X)\right] + \sum_{j=1}^{d}\sum_{i=1}^{d}\sum_{k=1}^{d}\frac{\partial}{\partial x_i \partial x_j}\left[\frac{1}{2}p(t,X)\sigma_{jk}(X,t)\sigma_{ik}(X,t)\right],$$

$$\textbf{(33)}$$

which can be rewritten as

$$\frac{\partial p(X,t)}{\partial t} = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}[p(X,t)\mu_i(X,t)] + \frac{1}{2}\sum_{i=1}^{d} \frac{\partial}{\partial x_i}[\sum_{j=1}^{d} \frac{\partial}{\partial x_j}[\sum_{k=1}^{d} p(X,t)\sigma_{jk}(X,t)\sigma_{ik}(X,t)]]$$

Using formula for derivative of product, we obtain **(34)**

$$\sum_{j=1}^{d} \frac{\partial}{\partial x_j}[\sum_{k=1}^{d} p(X,t)\sigma_{jk}(X,t)\sigma_{ik}(X,t)]$$

$$= \sum_{j=1}^{d} \frac{\partial}{\partial x_j}[\sum_{k=1}^{d} \sigma_{ik}(X,t)\sigma_{jk}(X,t)]P(X,t) + \sum_{j=1}^{d} \left[\sum_{k=1}^{d} \sigma_{ik}(X,t)\sigma_{jk}(X,t)\right] \frac{\partial}{\partial x_j} P(X,t)$$

Using formula for divergence of matrix valued function, we obtain **(35)**

$$\nabla \cdot \ \sigma(X,t)\sigma(X,t)^T = \begin{bmatrix} \sum_{j=1}^{d} \frac{\partial}{\partial x_j}[\sum_{k=1}^{d} \sigma_{1k}(X,t)\sigma_{jk}(X,t)] \\ \vdots \\ \sum_{j=1}^{d} \frac{\partial}{\partial x_j}[\sum_{k=1}^{d} \sigma_{dk}(X,t)\sigma_{jk}(X,t)] \end{bmatrix} \quad \textbf{(36)}$$

$$\sum_{j=1}^{d}\left[\sum_{k=1}^{d}\sigma_{ik}(X,t)\sigma_{jk}(X,t)\right]\frac{\partial}{\partial x_j}P(X,t)=\left[\sum_{k=1}^{d}\sigma_{ik}\sigma_{1k}\quad\cdots\quad\sum_{k=1}^{d}\sigma_{ik}\sigma_{dk}\right]\begin{bmatrix}\frac{\partial P(X,t)}{\partial x_1}\\ \vdots\\ \frac{\partial P(X.t)}{\partial x_d}\end{bmatrix}$$

$$=P(X,t)\left[\sigma(X,t)\sigma(X,t)^T\right]_i\nabla_x\log P(X,t)\qquad\textbf{(37)}$$

Substituting equations (36) and (37) yields

$$\boldsymbol{P(X,t)\nabla_x\log P(X,t)=P(X,t)\frac{1}{P(X,t)}\nabla_x P(X,t)=\nabla_x P(X,t)}$$

$$\sum_{j=1}^{d}\frac{\partial}{\partial x_j}[\sum_{k=1}^{d}p(X,t)\sigma_{jk}(X,t)\sigma_{ik}(X,t)]=$$

$$P(X,t)\left[\nabla\cdot\sigma(X,t)\sigma(X,t)^T\right]_i+P(X,t)\left[\sigma(X,t)\sigma(X,t)^T\right]_i\nabla_x\log P(X,t)\qquad\textbf{(38)}$$

Substituting equation (38) into equation (34), we obtain

$$\frac{\partial p(X,t)}{\partial t}=-\sum_{i=1}^{d}\frac{\partial}{\partial x_i}\{P(X,t)[\mu_i(X,t)-\frac{1}{2}\left[\nabla\cdot\sigma(X,t)\sigma(X,t)^T\right]_i-\frac{1}{2}\left[\sigma(X,t)\sigma(X,t)^T\right]_i\nabla_x\log P(X,t)]\}\qquad\textbf{(39)}$$

Let

$$\boldsymbol{\tilde{\mu}_i(X,t) = \mu_i(X,t) - \frac{1}{2}\left[\nabla \cdot \sigma(X,t)\sigma(X,t)^T\right]_i - \frac{1}{2}\left[\sigma(X,t)\sigma(X,t)^T\right]_i \nabla_x \log P(X,t)}$$

$$\tilde{\mu}(X,t) = \begin{bmatrix} \tilde{\mu}_1(X,t) \\ \vdots \\ \tilde{\mu}_d(X,t) \end{bmatrix} = \mu(X,t) - \frac{1}{2}\nabla \cdot \left[\sigma(X,t)\sigma(X,t)^T\right] - \frac{1}{2}\sigma(X,t)\sigma(X,t)^T \nabla_x \log P(X,t),$$

$$\mathrm{G}(\mathrm{X},\mathrm{t}) = 0$$

Define SDE as

$$dX = \tilde{\mu}(X,t)dt + G(X,t)dW \tag{40}$$

Then the marginal probability corresponding to SDE (40) satisfies equation (39).

Equation (40) is essentially ODE:

$$dX = \tilde{\mu}(X,t)dt,$$

which is the same se equation (32).

Graph-based Multi-ODE Neural Networks for SpatioTemporal Traffic Forecasting

Probability flow solution of the Fokker-Planck equation

Han Huang et al. 2023,

Conditional Diffusion Based on Discrete Graph Structures for Molecular Graph Generation

Lu, C. et al. 2022,
DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps

Every SDE has an associated probability flow ODE, which yields deterministic processes that sample from the same distribution as the SDE at each timestep. This establishes an equivalence to neural ODEs, allowing sampling via ODE solvers and exact computation of log-likelihoods.