§ Linear SDEs and Gaussian processes

In this lecture, we study a fundamental class of SDEs of the form

The results in this lecture also applies to SDEs with input $+ u(t)dt$

$$dX(t) = A(t)X(t)\,dt + B(t)\,dW(t), \quad X(0) = X_0, \tag{1}$$

where $X \in \mathbb{R}^d$, $A : [0, \infty) \to \mathbb{R}^{d\times d}$, $B : [0, \infty) \to \mathbb{R}^{d \times w}$, $W \in \mathbb{R}^w$.

These SDEs are linear SDEs and they are important in our context because their solutions are Gaussian processes (GPs) which are widely used in many applications, for instance, machine learning, signal processing, and control.

Thanks to the linearity in the drift and dispersion functions, we can solve linear SDE's solutions in closed-~~form~~ form by using a similar routine to that of linear ODEs.

Theorem 22. For any starting time s, the solution to the linear SDE in ① is

$$X(t) = F(t)X(s) + F(t)\int_s^t F(\tau)^{-1}B(\tau)\,dW(\tau),$$

where $F \in \mathbb{R}^{d\times d}$ is a matrix-valued function that solves the matrix-ODE

$$\frac{dF(t)}{dt} = A(t)F(t), \qquad F(s) = I_d \in \mathbb{R}^{d\times d}.$$

Proof. The result is a direct consequence of Itô's formula, and the properties of transition matrix / semigroup F. To see this, Since $\frac{dF(t)}{dt} = A(t)F(t)$, we have

$$\frac{dF(t)^{-1}}{dt} = -F(t)^{-1}A(t).$$

why? $\frac{d(F(t)F(t)^{-1})}{dt} = 0 = \frac{dF(t)}{dt}F(t)^{-1} + F(t)\frac{dF(t)^{-1}}{dt} = 0$

Now apply Itô's formula on a mapping $(t,x) \mapsto F(t)^{-1}x(t)$ we obtain

$$F(t)^{-1}X(t) = F(s)^{-1}X(s) + \int_s^t \frac{dF(t)^{-1}}{dt}(\tau)\,X(\tau)\,d\tau + \int_s^t F(\tau)^{-1}A(\tau)X(\tau)\,d\tau + \int_s^t F(\tau)^{-1}B(\tau)\,dW(\tau)$$

$$= X(s) + \int_s^t F(\tau)^{-1}B(\tau)\,dW(\tau).$$

Multiply $F(t)$ on both side of the equation we conclude the result.

□

Recall the solution

$$X(t) = F(t)\,X(s) + F(t)\int_s^t F(\tau)^{-1} B(\tau)\, dW(\tau),$$

where F solves $\frac{dF(t)}{dt} = A(t)\,F(t)$, $F(s) = \mathrm{Id}$. Even if the function F solves the linear ODE, it is not possible to write F in closed-form in general, except for a few special cases. For instance

1) A is time-invariant. Then $F(t) = e^{(t-s)A}$, hence

$$X(t) = e^{(t-s)A} X(s) + \int_s^t e^{(t-\tau)A} B(\tau)\, dW(\tau)$$

2) A is self-~~commutative~~ commuting, viz., $A(t)A(s) = A(s)A(t)$ for all t, s. Then $F(t) = e^{\int_s^t A(\tau)\,d\tau}$.

3) A is one-dimensional, then A is self-commuting.

For more general systems, we may have to approximate F. The common methods for doing so include, for example, Peano-Baker series and Magnus expansion.

Example 23. Consider the Ornstein -- Uhlenbeck SDE

$$dX(t) = -\theta X(t)\,dt + \sigma\,dW(t), \quad X(0) = X_0.$$

By the theorem, the solution is

$$X(t) = e^{-\theta t} X_0 + \sigma e^{-\theta t} \int_0^t e^{\theta s}\,dW(s)$$

Think: how would you simulate this process?

By the explicit solution of linear SDE, we can right away tell that the solution is a Gaussian process.

Definition 24. A vector of random variables $X_{1:T} := \{X_1, X_2, \ldots X_T\} \in \mathbb{R}^T$ is jointly Normal distributed if for every non-trivial linear combinations of them are Normal distributed. That is, $\sum_{i=1}^{T} \lambda_i X_i$ is Normal for all non-trivial λ's. An equivalent saying is that $\mathbb{E}[e^{i\langle \lambda_{1:T}, X_{1:T}\rangle}] = e^{i\langle \lambda_{1:T}, \mu_{1:T}\rangle - \frac{1}{2}\langle \lambda_{1:T}, \Sigma \lambda_{1:T}\rangle}$ for some mean $\mu_{1:T}$ and covarice matrix Σ.

Does the mutivarite Normal density function define Normal vectors?

Definition 25. A stochastic process X is called a Gaussian process, if for every $T > 0$ and ~~times~~ reals $t_1 < t_2 < \cdots t_T$, the random variables $X_1(t_1), X(t_2), \ldots X(t_T)$ are jointly Normal distributed.

It is not difficult to show that

$$X(t) = F(t)X(0) + F(t)\int_0^t F(\tau)^{-1} B(\tau)\, dW(\tau)$$

is a GP by definition. The key lies in the fact that the Wiener integral $\int_0^t F(\tau)^{-1} B(\tau)\, dW(\tau)$ is a Gaussian random variable for all t. (See, Kuo, 2006, pp. 11).

Another heuristic way to see that X is a GP is by the stationary solution to the Kolmogorov forward equation. Imagine an SDE of the form

$$\begin{aligned} dX(t) &= \frac{1}{2}\nabla_x \log\left(\frac{1}{Z} e^{-X(t)^T \Sigma X(t)}\right) dt + dW(t) \\ &= -\frac{1}{2}\Sigma X(t)\, dt + dW(t) \end{aligned}$$

which is a linear SDE, following the stationary PDF that is a Gaussian $N(0, \Sigma)$.

Any GPs are completly characterised by their mean and covariance functions, denoted by:

$$m(t) := \mathbb{E}[X(t)]$$

$$C(t,s) := \underline{\mathrm{Cov}[X(t), X(s)]} := \mathbb{E}[(X(t)-m(t))(X(s)-m(s))^T]$$

In the machine learning community, we conveniently use a shorthand notation

we sometimes call this (cross)-covariance

$$X(t) \sim GP(m(t), C(t,s))$$

to denote a GP.

Then, we derive explicitly the mean and covariance functions of linear SDEs. Recall

$X(0)$ and $W(t)$ are independent!

$$X(t) = F(t)X(0) + F(t)\int_0^t F(s)^{-1} B(s)\, dW(s),$$

Then,

$$m(t) = \mathbb{E}[X(t)] = F(t)\,\mathbb{E}[X(0)] + 0$$

$$C(t,s) = \mathrm{Cov}[X(t), X(s)] = \mathrm{Cov}[F(t)X(0), F(s)X(0)] + \underline{\mathrm{Cov}\left[F(t)\int_0^t F(\tau)^{-1}B(\tau)\,dW(\tau),\ F(s)\int_0^s F(\tau)^{-1}B(\tau)\,dW(\tau)\right]}$$

Itô isometry

$$= F(t)\,\mathrm{Cov}[X(0)]\,F(s)^T + F(t)\int_0^{\min(t,s)} F(\tau)^{-1}\underbrace{B(\tau)B(\tau)^T}_{\Gamma(\tau)}(F(\tau)^{-1})^T\,d\tau\ F(s)^T$$

$$v(t) = \mathrm{Cov}[X(t)] = F(t)\left(\mathrm{Cov}[X(0)] + \int_0^t F(\tau)^{-1}B(\tau)B(\tau)^T(F(\tau)^{-1})^T\,d\tau\right)F(t)^T$$

We can also express the (cross)-covariance in terms of the (marginal)-covariance as

$$\mathrm{Cov}[X(t), X(s)] = \begin{cases} \mathrm{Cov}[X(t)]\,(F(s)^{-1})^T F(t)^T, & t \le s. \\ F(t)\,F(s)^{-1}\,\mathrm{Cov}[X(s)], & t > s. \end{cases}$$

But how do we comput these covariances? By their formulae, it seems that we need to compute some complicated Wiener integrals. In fact, we can compute them by solving a system of ODEs. We can verify that $\mathbb{E}[X(t)] = m(t)$ and $V(t) := \mathrm{Cov}[X(t)]$ solve the linear ODEs

$$\frac{dm(t)}{dt} = A(t)\,m(t), \qquad m(0) = \mathbb{E}[X(0)]$$

$$\frac{dV(t)}{dt} = A(t)\,V(t) + V(t)\,A(t)^T + B(t)B(t)^T, \qquad V(0) = \mathrm{Cov}[X(0)]$$

which are easy to solve. Furthermore, if ~~the m~~ A and B are time-invariant, then we have the following simplified results:

$$\frac{dm(t)}{dt} = A\, m(t), \quad \Rightarrow m(t) = e^{At} m(0)$$

$$\frac{dV(t)}{dt} = A\, V(t) + V(t) A^T + BB^T,$$

how do you solve this ODE in closed-form? Hint: rearrange the terms so that they are linear in V, e.g., $[\cdots] \cancel{V(t)} + BB^T$ for some matrix $[\cdots]$.

$\underline{\mathrm{Vec}(EFG) = (G^T \otimes E)\,\mathrm{vec}(F)}$

$$F(t) = e^{At}$$

$$C(t,s) = \begin{cases} V(t)\, e^{(s-t)A}, & t \le s \\ e^{(t-s)A} V(t), & t > s. \end{cases}$$

It is worth remarking that to obtain the stationary solution $\frac{dV(t)}{dt} = 0$. We have the celebrated Lyapunov equation

$$AV + VA^T + BB^T = 0.$$

How to solve the Lyapunov equation? Is it a linear equation?

Example 25.

$$dX(t) = -\theta X(t)\,dt + \sigma\, dW(t), \quad X(0) \sim N(m_0, V_0).$$

The solution and the mean and covariance functions are:

$$X(t) = e^{-\theta t} X_0 + \sigma e^{-\theta t} \int_0^t e^{\theta s} \, dW(s).$$

$$\begin{cases} \dfrac{dm(t)}{dt} = -\theta m(t), & m(0) = m_0 \\ \dfrac{dv(t)}{dt} = -2\theta v(t) + \sigma^2 & v(0) = v_0 \end{cases}$$

$$\to \quad m(t) = e^{-\theta t} m_0$$

$$v(t) = v_0 e^{-2\theta t} + \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t})$$

The solution to the associated Lyapunov equation $-2\theta V + \sigma^2 = 0$ is $V = \frac{\sigma^2}{2\theta}$.

If we set the initial variance $V_0 = \frac{\sigma^2}{2\theta}$ be the stationary one, the $v(t) = \frac{\sigma^2}{2\theta}$ is time-invariant, and it follows that $C(t,s) = \frac{\sigma^2}{2\theta} \exp(-\theta|t-s|)$ which is the celebrated exponential kernel for Gaussian processes.

Since now we know the mean and covariance functions of linear SDEs, we can exactly simulate them, say for example, at $t_1 < t_2 < \cdots t_T$. The canonical way is to comput the means $m_{1:T} := \{m(t_1), m(t_2) \ldots m(t_T)\}$ and the covariance matrix

$$C_{1:T} = \begin{bmatrix} c(t_1,t_1) & c(t_1,t_2) & \cdots & c(t_1,t_T) \\ c(t_2,t_1) & c(t_2,t_2) & & \\ \vdots & & \ddots & \\ c(t_T,t_1) & & & c(t_T,t_T) \end{bmatrix} \in \mathbb{R}^{T\times T},$$

How expensive? $O(T^3)$

then draw $\sim \underline{N(m_{1:T}, C_{1:T})}$. But this is a fuss and is <u>expensive.</u> Since we know that X is a Markov process, we can make the simulation efficient by drawing the samples sequentially in time. Recall that X is a aP-SDE, so its transition distribution is Normal too. Recall Algorithm 14, the conditional mean and covariance is all we need. We can discretise the SDE solution at $t_1 < t_2 < \cdots t_T$ as

$$X(t_k) = \overbrace{F(t_k)X(t_{k-1})}^{\mathbb{E}[X(t_k) \mid X(t_{k-1})]} + Q(t_k),$$

$$Q(t_k) \sim N(0, \underline{\mathrm{Cov}[X(t_k) \mid X(t_{k-1})]}),$$

the conditional covariance does Not depend on $X(t_{k-1})$ here!

and recall that the covariance $\mathrm{Cov}[X(t_k) \mid X(t_{k-1})] = F(t)\int_{t_{k-1}}^{t_k} F(\tau)^{-1} B(\tau)B(\tau)^T(F(\tau)^{-1})^T d\tau \, F(t)^T$ is the solution of the ODE $\frac{dv(t)}{dt} = \cdots$

starting from $\underline{v(0)=0}$.

Algorithm 26. Simulate Linear SDEs.

Input: Times $t_1 < t_2 < \cdots t_T$, and initial mean m_0 and covariance V_0.

output: $X(t_1), X(t_2), \ldots X(t_T)$

1. Draw $X(0) \sim N(m_0, V_0)$
2. For $k = 1, 2, \ldots T$ do
3. solve

$$① \begin{cases} \frac{dm(t)}{dt} = A(t)m(t), & m(t_{k-1}) = X_{k-1} \\ \frac{dV(t)}{dt} = A(t)V(t) + V(t)A(t)^T + B(t)B(t)^T, & V(t_{k-1}) = 0. \end{cases}$$

at t_k to obtain $m(t_k)$ and $V(t_k)$

4. Draw $\xi_k \sim N(0, I_d)$
5. $X(t_k) = m(t_k) + \sqrt{V(t_k)}\,\xi_k$

cholesky decomposition

Return $X(t_1), X(t_2), \ldots X(t_T)$

Remark: How do we solve ①? If A and B do not depend on time, then they are just linear time-invariant ODEs, and the solution is, e.g., $m(t_k) = e^{(t_k - t_{k-1})A} X(t_{k-1})$.

The method in Axelsson and Gustafsson, 2015, see also, Särkkä and Solin, 2019, PP. 83 provides a convenient

please read it!

way to compute the covariance ODE $V(t)$. please see the function 'discretise_lti_sde' in 'lec5_linear_sde_mean_cov.ipynb'.

$$\frac{dX_1(t)}{dt} = X_2(t)$$

Example. 27. Motion model.

$$d\begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} dW(t).$$

This is a model that obeys Newton's motion Law. The variable X_1 and X_2 represent the position and the velocity of an object, respectively.

Compute its conditional mean and covariance, the sample a trajectory based on these. Does this mode has a stationary variance? why?

See 'lec5_linear_sde_mean_cov.ipynb'.