LECTURE 1 INTRODUCTION TO SPDES

1. Kinds of questions we want to ask

What are SPDEs? PDEs + noise;

Why (When) SPDEs are useful?

- an evolutionary process itself is stochastic in nature
- detect "signal" against "random noise"
- data may be missing due to experimental reasons or intrinsically unobservable,
- we want to integrate multiple (heterogeneous) data and incorporate prior knowledge in a flexible and principled way

When solving SPDEs, many issues arise such as

- Do my SPDEs have solutions, in what senses?
- How do I solve these SPDEs efficiently? (Numerical PDEs + stochastic simulation) one important topic is the simulation of noise/Gaussian processes.
- Stability, convergence, computational cost, etc

What we will not address: basic probability as background covered in APMA263, APMA264.

Check APMA 2812B for more theoretical views. (Fridays 12-2:30).

2. Review of basic probability

2.1. **Gaussian random variables.** On a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ ($\Omega = \mathbb{R}$), if a cumulative distribution function of a random variable X is normal, i.e.,

$$\mathbb{P}(X < x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\, dy, \quad \sigma > 0.$$

then the random variable X is called a Gaussian (normal) random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here X is completely characterized by its mean μ and its standard deviation σ. We denote $X \sim \mathcal{N}(\mu, \sigma^2)$. The probability density function of X is

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

When $\mu = 0$ and $\sigma = 1$, we call X a standard Gaussian (normal) random variable. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$, i.e., Z is a standard Gaussian (normal) random variable.

Example 2.1. *If X_i are mutually independent Gaussian random variables, then $\sum_{j=1}^{N} a_i X_i$ is a Gaussian random variable for any $a_i \in \mathbb{R}$. In particular, if $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then*

$$\alpha X_1 + \beta X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \alpha^2\sigma_1^2 + \beta^2\sigma_2^2).$$

Date: October 3, 2019.

2.2. **Gaussian random vectors.**

Definition 2.2 (Gaussian random vector)**.** *A $\mathbb{R}^n$-valued random vector $X = (X_1, X_2, \dots, X_n)^\top$ has an n-variate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ if $X = \boldsymbol{\mu} + AZ$ where the matrix A is of size $n \times n$, $\Sigma = AA^\top$, and $Z = (Z_1, Z_2, \dots, Z_n)^\top$ is a vector with independent standard Gaussian (normal) components.*

When $n = 1$, X is a (univariate) Gaussian random variable. When Σ is non-singular, the probability density of X is

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \left|\Sigma\right|^{1/2}}} e^{-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}{2}}.$$

Example 2.3 (Equivalent definition)**.** *A set of random variables $\{X_i\}_{i=1}^n$ are called* jointly Gaussian *if $\sum_{i=1}^n a_i X_i$ is a Gaussian random variable for any $a_i \in \mathbb{R}$. Then $X = (X_1, \cdots, X_n)^\top$ is a Gaussian random vector if and only if $\{X_i\}_{i=1}^n$ are* jointly Gaussian*.*

The correlation of two random variables (vectors) is a normalized version of the covariance, with values ranging from -1 to 1:

$$\mathrm{Corr}(X, Y) = \frac{\mathbb{C}\mathrm{ov}[(X, Y)]}{\sqrt{\mathbb{V}\mathrm{ar}[X]\mathbb{V}\mathrm{ar}[Y]}}, \quad \mathbb{C}\mathrm{ov}[(X, Y)] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^\top].$$

When $\mathrm{Corr}(X, Y) = 0$, we say X and Y are uncorrelated.

Example 2.4. *For a Gaussian random variable $X \sim \mathcal{N}(0, \sigma^2)$ and a Bernoulli Z with $\mathbb{P}(Z = \pm 1) = \frac{1}{2}$, X and Z are independent. Show that*

a) *the product ZX is a Gaussian random variable;*
b) *and X and ZX are uncorrelated;*
c) *but X and ZX are not independent.*

Check whether $(X, ZX)^\top$ is a Gaussian random vector or not.

Exercise 2.5. *Consider a Gaussian vector (X, Y) with X, Y satisfying $\mathbb{C}ov[(X, Y)] = 0$ (uncorrelated). Then X and Y are independent.*

Exercise 2.6. *Assume that X and Y are Gaussian random variables, then $X+Y$ independent of $X - Y$ implies X independent of Y if and only if (X, Y) is a Gaussian random vector.*

Exercise 2.7 (Marginally Gaussian but not Jointly Gaussian)**.** *Let X and Y be two Gaussian random variables. Consider the following joint PDF*

$$f_{X,Y}(x, y) = \frac{1}{\pi \sigma_X \sigma_Y} \exp(-\frac{1}{2}(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2})) \mathbf{1}_{\{xy>0\}}.$$

This joint PDF is not jointly Gaussian.

2.3. **Gaussian processes. Gaussian processes** on $[0, T]$ can be written in the following form

$$\mu(t) + \sum_{k=1}^{\infty} \sqrt{\lambda_k} m_k(t) \xi_k, \tag{2.1}$$

where $\lambda_k \geq 0$, ξ_k's are i.i.d. standard normal, and m_k's form a complete orthonormal system (CONS) in $L^2[0, T]$.

Example 2.8.

- *Gaussian white noise*

$$\sum_{k=1}^{\infty} m_k(t) \xi_k, \quad \lambda_k \equiv 1. \tag{2.2}$$

- *Brownian motion*

$$\sum_{k=1}^{\infty} \int_0^t m_k(s)\, ds \xi_k, \tag{2.3}$$

Taking the cosine basis, we can write the Brownian motion as

$$\sqrt{\frac{1}{T}} t + \sum_{k=2}^{\infty} \frac{\sqrt{2T}}{(k-1)\pi} \sin(\frac{(k-1)\pi}{T} t) \xi_k. \tag{2.4}$$

Remark 2.9. *All definitions can be extended to a bounded domain in $\mathbb{R}^d$. On an unbounded domain, we may not have the definition via series. Instead, we will need integration with respect to Brownian motions.*

Gaussian processes can also be defined through Gaussian random vectors.

Definition 2.10 (Gaussian process)**.** *A collection of random variables $X = \{X(t)\}_{t\in\mathcal{T}}$ is called a Gaussian process, if the joint distribution of any finite number of its members is Gaussian. In other words, a Gaussian process is a $\mathbb{R}^d$-valued stochastic process with continuous time (or with index) t such that $(X(t_0), X(t_1), \cdots, X(t_n))^\top$ is a $n+1$-dimensional Gaussian random vector for any $t_0 < t_1 < \cdots < t_n$ in $\mathcal{T}$.*

The Gaussian process is denoted as $X = \{X(t)\}_{t\in\mathcal{T}}$ where $\mathcal{T}$ is a set of indexes. Here $\mathcal{T}$ can be $\mathbb{R}$ or $[0, \infty)$ while in the first definition it might not be working when $T = \infty$.

Exercise 2.11. *Use Defnition 2.10 to show that the following process is a Gaussian process. $X(t) = \sum_{k=1}^{\infty} \frac{1}{k} m_k(t)\xi_k$, where $\{m_k(t)\}$ is a CONS in $L^2([0,T])$ and ξ_k's are i.i.d. standard normal.*

2.4. **Characterization of Gaussian processes.** The consistency theorem of Kolmogorov [Karatzas and Shreve, 1991, Theorem 2.2] implies that the finite dimensional distribution of a Gaussian stochastic process $X(t)$ is uniquely characterized by two functions: the mean function $\mu_t = \mathbb{E}[X(t)]$ and the covariance function $C(t,s) = \mathbb{C}\text{ov}[X(t), X(s)]$.

Here all Gaussian processes are defined in $\mathbb{R}^d$, with Euclidean norms.

Definition 2.12 (Stationary processes)**.** *A random process with $\mathbb{E}[X^2(t)] < \infty$ is called stationary in the wide sense if its mean $\mu(t)$ and its covariance function $K(t,s)$ do not change under a shift of the parameters.*

Stationary kernels depends on the difference of two points:

$$K(x,y) = K_s(x-y), \tag{2.5}$$

The kernel in (2.5) is sometimes called anisotropic stationary kernel (depending on the distance and also the direction).

When $K(x,y)$ depends only on $\|x-y\|$ (the distance between x and y), it is called the isotropic kernel

$$K(x,y) = K_I(\|x-y\|).$$

One important example of practical interest is the Martern kernel:

$$M_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{d}{\rho}\right)^\nu K_\nu\left(\sqrt{2\nu}\frac{d}{\rho}\right), \quad d = \|x-y\|, \tag{2.6}$$

where $\rho, \nu > 0$ and $K_\nu(\cdot)$ is the modified Bessel function of the second kind.

When $\nu = 0.5$, it becomes the exponential kernel $\frac{2\sigma^2}{\pi} \exp(-\frac{d}{\rho})$.

When $\nu = \infty$, it becomes a Gaussian kernel.

In Matlab, the Matern kernel with $\nu = 1.5, 2.5$ have been implemented `https://www.mathworks.com/help/stats/kernel-covariance-function-options.html`.

Remark 2.13. *All the isotropic kernels are defined on $\mathbb{R}^d$ but sometime it is convenient to define compactly supported kernels: $K_I(\|x-y\|) = 0$ if $\|x-y\| \geq L > 0$. In computation, the covariance matrix $K_I(x_i, x_j)$ with be sparse since when $|i-j|$ is large, the entry $K_I(x_i, x_j) = 0$. For example, a compactly supported Martern kernel is $M_\nu(d)\max(1-\frac{d}{L})^{\nu_1}, 0)$ where $\nu_1 \geq (d+1)/2$ such that the "truncated" kernel is positive definite.*

Remark 2.14. *The SPDE $(\kappa^2 - \Delta)^{\alpha/2} X(s) = \dot{W}(s)$, $s \in \mathbb{R}^d$ where $\kappa = \sqrt{2\nu}/\rho$ and $\alpha = \nu + d/2$. Here $\dot{W}(s)$ is the white noise sheet in $\mathbb{R}^d$. Then there is only one stationary solution $X(s)$, which has Matern kernel and with variance $\frac{\Gamma(\nu)}{\Gamma(\nu+d/2)(4\pi)^{d/2}\kappa^{2\nu}}$.*

2.5. **Brownian sheets, white noise sheets, cylindrical processes.** Let $\{e_k(x)\}$ be CONS in $L^2(D)$ and W_k's be independent Brownian motions.

$$\dot{W}^Q(t,x) = \sum_{k=1}^{\infty} \sqrt{q_k} e_k(x) \dot{W}_k(t), \quad q_k \geq 0. \tag{2.7}$$

If $\sum_k q_k e_k(x) e_k(y) = Q(x-y)$, then this above process has a covariance $Q(x-y)\delta(t-s)$. When $q_k \equiv 1$, the process $\dot{W}^Q(t,x)$ is called space-time white noise.

In higher dimensions, we can define Brownian motions, Brownian sheet, see Chapter 1 of [Lototsky and Rozovsky, 2017]. The Brownian sheet $W_{s,t}$ on $[0,\infty) \otimes [0.\infty)$ has the kernel $(s \wedge s')(t \wedge t')$.

3. The color of noises

For the Gaussian process $\sum_{k=1}^{\infty} \sqrt{\lambda_k} m_k(t) \xi_k$, the decay of λ_k with k characterizes of smoothness of the process. When λ_k is proportional to $k^{-\alpha}$, where $0 \leq \alpha \leq 2$, the noise is called $1/f^\alpha$ noise.

For stationary Gaussian processes, we can characterize its smoothness through its covariance kernel and power spectrum. Recall the Fourier Transform in 1D

$$\hat{f}(\omega) = \int_{\mathbb{R}} f(t) e^{\mathrm{i}\omega t}\, dt \tag{3.1}$$

The inverse Fourier transform is defined as

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\omega) e^{-\mathrm{i}\omega t}\, dt \tag{3.2}$$

Table 1. A short table of Fourier transform

$f(t)$	$\hat{f}$	
1	$2\pi\delta(\omega)$	
$\delta(t)$	1	
$e^{-\frac{t^2}{2\sigma^2}}$	$\sigma\sqrt{2\pi} e^{-\sigma^2\omega^2/2}$	
$e^{-a\lvert t\rvert}$	$\frac{2a}{a^2+\omega^2}$	$a > 0$
$\frac{(\alpha t)^\nu}{\Gamma(\nu)2^{\nu-1}} K_\nu(\alpha t)$	$\frac{\Gamma(\nu+1/2)\alpha^{2\nu}}{\pi^{1/2}\Gamma(\nu)(\alpha^2+\omega^2)^{\nu+1/2}}$	$\alpha, \nu > 0$, $t \geq 0$

The auto-covariance function of a stationary process $X(t)$ (in the wide sense) is defined by $K(s) = \mathbb{E}[X(t-s)X^\top(t)]$ if $\mathbb{E}[X(t)] = 0$ (otherwise use $X(\cdot) - \mathbb{E}[X(\cdot)]$ in place of $X(\cdot)$.)

Theorem 3.1 (Wiener–Khinchin theorem)**.** *If the auto-covariance function of a stationary process $X(t)$ exists and is absolutely integrable, then there exists a monotone function $F(\omega)$ such that*

$$K(t) = \int_{\mathbb{R}} e^{\mathrm{i}\omega t}\, dF(\omega). \tag{3.3}$$

Here F is called power spectral distribution. If F is absolutely continuous with respect to the Lebsgue measure, then we denote the power spectral density (power spectrum) by f, i.e., $f(\omega)\, d\omega = F(\omega)\, d\omega$. Then

$$K(t) = \int_{\mathbb{R}} e^{\mathrm{i}\omega t} f(\omega)\, d\omega, \quad \text{and } f(\omega) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}\omega s} K(s)\, ds. \tag{3.4}$$

The power spectrum is actually a good characterization of smoothness of the stationary process $X(t)$. In fact, if $f(\omega) \sim \omega^{-\alpha}$, $0 \le \alpha \le 2$, the process $X(t)$ is called $\frac{1}{f^\alpha}$ noise. When $\alpha = 2$, it is called Brownian noise; When $\alpha = 2$, it is called pink noise; When $\alpha = 0$ ($f(\omega) \equiv 1$), it is called white noise.

Exercise 3.2. *A covariance function is positive definite if the corresponding covariance matrix from any finite time instants is positive definite.*

Exercise 3.3. *Show that when $K(s)$ is positive definite, then the power spectral density f is unique.*

In the next lecture, we will discuss a precise definition of smoothness of stochastic processes.

4. Why random and Brownian motion (white noise)?

Consider the following simple population growth model

$$dy(t) = k(t)y(t)\, dt, \quad y(0) = y_0. \tag{4.1}$$

Here $y(t)$ is the size of population and $k(t)$ is the relative growth rate. In practice, $k(t)$ is not completely known and is disturbed around some known quantity $\bar{k}(t)$:

$$k(t) = \bar{k}(t) + \text{`perturbation (noise)'}.$$

Here $\bar{k}(t)$ is deterministic and is usually known while no exact behavior exists of the perturbation (noise) term. The uncertainty (lack of information) about $k(t)$ (the perturbation term) is naturally represented as a stochastic quantity, denoted as $v(t, \omega)$. Here ω represents the randomness.

To address the dependence on ω, we then write the ordinary differential equation as

$$dy(t,\omega) = k(t,\omega)y(t,\omega)\, dt^1, \quad y(0) = y_0. \tag{4.2}$$

Here y_0 can be a random variable but we take $y_0 = 1$ (deterministic) for simplicity.

4.0.1. *Solutions to* (4.2). When $\bar{k}(t) = 0$, $k(t,\omega) = v(t,\omega)$ may take the following form:

- $v(t,\omega) = \xi \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable. The covariance (in this case variance) of $v(t,\omega)$ is 1.
- $v(t,\omega)$ where the covariance function of $V(t)$ is $\exp(-\frac{|t-s|}{A})$, with A the correlation time.
- $v(t,\omega) = W(t,\omega)$ is a standard Brownian motion where the covariance function is $\min(t,s)$.
- $v(t,\omega) = \dot{W}(t,\omega)$ is the white noise and the covariance function is $\delta(t-s)$.

[1]We assume that $k(t,\omega)$ is sufficiently smooth in t. It will be clear why some smoothness is needed when we introduce the stochastic product and caculus later on.

When $k(t,\omega) =: \dot{W}(t,\omega)$, Equation (4.2) is understood in the Stratonovich sense

$$dy = y \circ dW(t), \quad y(0) = y_0 \tag{4.3}$$

so that one can apply the classical chain rule. Here the circle means Stratonovich product.

The exact solution to Equation (4.2) is $y = y_0 \exp(K(t))$, where $K(t) = \int_0^t k(s)\, ds$ is again Gaussian with mean zero. It can be readily checked by the definition of moments that

- $K(t) \sim \mathcal{N}(0, t^2)$ when $k(t,\omega) =: \xi \sim \mathcal{N}(0,1)$;
- $K(t) \sim \mathcal{N}\big(0, 2At + 2A^2(\exp(-\frac{t}{A}) - 1)\big)$ when $k(t,\omega)$ has the two-point correlation function of $\exp(-\frac{|t_1 - t_2|}{A})$.
- $K(t) \sim \mathcal{N}(0, \frac{t^3}{3})$ when $k(t,\omega) = W(t,\omega)$ is a standard Brownian motion.
- $K(t) \sim \mathcal{N}(0, t)$ is the Brownian motion when $k(t,\omega) = \dot{W}(t,\omega)$ is the white noise.

Then the moments of the solution y are, for $m = 1, 2, \cdots$,

$$\mathbb{E}[y^m(t)] = y_0^m \exp(\frac{m^2}{2}\sigma^2), \tag{4.4}$$

where $\sigma^2 = t^2,\ 2At + 2A^2(\exp(-\frac{t}{A}) - 1),\ \frac{t^3}{3},\ t$ for the listed processes, respectively. These results are useful and can be used in checking the accuracy of different numerical schemes applied to Equation (4.2).

Appendix A. Non-stationary covariance kernels

Locally stationary kernels are defined by

$$K(x,y) = K_1(\frac{x+y}{2}) K_2(x-y). \tag{A.1}$$

Here $K_1 \geq 0$ and K_2 is a stationary kernel. When K_1 is a constant, the kernel is stationary and when K_2 is a constant, the kernel is called **exponentially convex kernel**. For example, $\exp(-\frac{x^2+y^2}{2}) = \exp(-2\frac{(x+y)^2}{2}) \exp(-2\frac{(x-y)^2}{2})$ is a locally stationary kernel. Also, $K_1(x+y)\delta(x-y)$ is a locally stationary kernel (the noise is called a locally stationary white noise).

There are non-stationary kernels $K(x,y)$ not functions of the difference $x - y$. Examples of non-stationary kernel

- polynomial kernels $K(x,y) = (x^\top y + a)^p$, $p > 0$ and $a \in \mathbb{R}$.
- separable non-stationary kernels: $K(x,y) = K_1(x)K_2(y)$, where K_1 and K_2 are stationary kernels. In this case, a huge computational cost can be saved as the covariance matrix will be rank-1.

Definition A.1 (Stationary reducible kernels)**.** *A nonstationary kernel is stationary reducible if the existsa bijective deformation Φ such that $K(x,y) = K_s^*(\Phi(x) - \Phi(y))$, where K_s^* is a stationary kernel.*

Similarly, we can define locally stationary reducible kernels.

References

[Karatzas and Shreve, 1991] Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition.

[Lototsky and Rozovsky, 2017] Lototsky, S. V. and Rozovsky, B. L. (2017). *Stochastic partial differential equations*. Universitext. Springer, Cham.