AM216 Stochastic Differential Equations

Lecture 06

List of topics in this lecture

- Long road to concluding dW/dt is a white noise:
  - Energy spectrum density (ESD), power spectrum density (PSD);
  - Stationary stochastic process, auto-correlation function $R(t)$;
  - Wiener-Khinchin theorem (PSD is Fourier transform of $R(t)$);
  - dW/dt has a uniform PSD (definition of white noise).
- Constrained Wiener process, Bayes Theorem

Recap

The short story of white noise:

1) $Z(t) \equiv \dfrac{dW}{dt}$ is not a regular function.

2) $E\Big(Z(t)Z(s)\Big) = \delta(t-s)$

3) $\int \exp(-i2\pi\xi t) E\Big(Z(t)Z(0)\Big) dt = 1$

4) $Z(t)$ is a white noise (we will clarify what this means).

Plan for the long story

We will first address the definition of white noise in item 4). We discuss a general stationary stochastic process and in that context define white noise in steps listed below

- Energy spectrum density (ESD)
- Power spectrum density (PSD)
- A general stationary stochastic process and its PSD
- Relation between PSD and auto-correlation function
- Definition of white noise based on PSD

Before we proceed with the plan, we finish the properties of Fourier transform.

Properties of Fourier transform

1) $F\left[\rho_{N(0,\sigma^2)}(t)\right] = F\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-t^2}{2\sigma^2}\right)\right] = \exp\left(-2\pi^2\sigma^2\xi^2\right)$
2) $F[\delta(x)] = 1$
3) $F[1] = \delta(\xi)$
4) Parseval's theorem

$$\int\left|y(t)\right|^2 dt = \int\left|\hat{y}(\xi)\right|^2 d\xi$$

Proof:

$$\int\left|\hat{y}(\xi)\right|^2 d\xi = \int \hat{y}(\xi)\overline{\hat{y}(\xi)}d\xi$$

$$= \int\left(\int \exp(-i2\pi\xi t)y(t)dt \int \exp(i2\pi\xi s)\overline{y(s)}ds\right)d\xi$$

$$= \int\left(\int\int \exp\left(-i2\pi\xi(t-s)\right)y(t)\overline{y(s)}dt\,ds\right)d\xi$$

Change the order of integration

$$= \int\int y(t)\overline{y(s)}\left(\underbrace{\int \exp\left(-i2\pi\xi(t-s)\right)d\xi}_{F[1]=\delta(t-s)}\right)dt\,ds$$

$$= \int\int y(t)\overline{y(s)}\delta(t-s)dt\,ds$$

$$= \int y(s)\overline{y(s)}ds = \int\left|y(s)\right|^2 ds$$

The long story of white noise

Energy spectrum density (ESD)

In physics problems,

$$\text{Energy} \propto \int\left|y(t)\right|^2 dt = \int\left|\hat{y}(\xi)\right|^2 d\xi$$

Note: Here "energy" refers to the energy dissipated.

Examples:

$y(t)$ = electric current

$$\text{Energy} = \int R \cdot y(t)^2 \, dt \,, \qquad R = \text{electrical resistance}$$

$y(t)$ = velocity

$$\text{Energy} = \int b \cdot y(t)^2 \, dt \,, \qquad b = \text{viscous drag coefficient}$$

For mathematical convenience, we scale the energy to define

$$\text{Energy} = \int \left| y(t) \right|^2 dt = \int \left| \hat{y}(\xi) \right|^2 d\xi$$

We like to know how the energy is distributed in the frequency dimension.

Definition of energy spectrum density (ESD)

$$\text{ESD} \equiv \left| \hat{y}(\xi) \right|^2 = \left| \int \exp(-i2\pi\xi t) \, y(t) \, dt \right|^2$$

Caution: $\left| \hat{y}(\xi) \right|^2$ is an unnormalized density.

$$\int \left| \hat{y}(\xi) \right|^2 d\xi = \int \left| y(t) \right|^2 dt = \text{Energy} \neq 1$$

Other examples of unnormalized density:

Population density: X number of persons per square mile

Pollution density: X amount of chemicals per unit volume of air or water

Car density: X number of cars per thousand persons

X number of cars per square mile

Caution: (different definitions of energy spectrum density)

In electrical engineering (EE), energy spectrum density is defined as

$$\text{ESD} \equiv \Phi(\omega) = \left| \frac{1}{\sqrt{2\pi}} \int \exp(-i\omega t) \, y(t) \, dt \right|^2$$

$\Phi(\omega)$ and $\left| \hat{y}(\xi) \right|^2$ are related by

$$\Phi(\omega) = \frac{1}{2\pi}\left|\hat{y}(\xi)\right|^2, \qquad \xi = \frac{\omega}{2\pi}$$

Power spectrum density

Energy spectrum density is meaningful only when

$$\text{Energy} = \int \left|y(t)\right|^2 dt = \text{finite}$$

Example:

$y(t)$ = electric current = y_0, constant in time

$$\text{Energy dissipated} = \int R \cdot y_0^2\, dt = \infty$$

When the total energy is not finite, we look at the energy per unit time.

$$\text{Power} = \frac{\text{Energy}}{\text{time}} = \text{finite}$$

Definition of power spectrum density (PSD)

$$\text{PSD} \equiv \lim_{T\to\infty} \frac{\left|\int_{-T}^{T} \exp(-i2\pi\xi t)\, y(t)\, dt\right|^2}{2T}$$

Expression of power spectrum density (PSD)

We write PSD into a more workable expression.

$$\text{PSD} = \lim_{T\to\infty} \frac{\int_{-T}^{T} \exp(-i2\pi\xi t)\, y(t)\, dt \int_{-T}^{T} \exp(i2\pi\xi s)\, \overline{y(s)}\, ds}{2T}$$

$$= \lim_{T\to\infty} \frac{\int_{-T}^{T}\int_{-T}^{T} \exp(-i2\pi\xi(t-s))\, y(t)\, \overline{y(s)}\, dt\, ds}{2T}$$

change of variable $\tau = t - s$

$$= \lim_{T\to\infty} \frac{\int_{-T}^{T}\int_{-T-s}^{T-s} \exp(-i2\pi\xi\tau)\, y(\tau+s)\, \overline{y(s)}\, d\tau\, ds}{2T}$$

Draw the integration region in s-τ plane.

For each s, the range for τ is $[-T-s, T-s]$.

For each τ, the range for s is $[a(\tau), b(\tau)]$

where

$$a(\tau)=\begin{cases} -T-\tau, & \tau\in[-2T,\ 0] \\ -T, & \tau\in[0,\ 2T] \end{cases}$$

$$b(\tau)=\begin{cases} T, & \tau\in[-2T,\ 0] \\ T-\tau, & \tau\in[0,\ 2T] \end{cases}$$

Change the order of integration

$$\text{PSD}=\lim_{T\to\infty}\frac{\int_{-2T}^{2T}\exp(-i2\pi\xi\tau)\int_{a(\tau)}^{b(\tau)}y(\tau+s)\overline{y(s)}ds\,d\tau}{2T} \qquad \text{(PSD01)}$$

So far, we worked with deterministic process $y(t)$.

Next we introduce stochastic process and stationary stochastic process.

Stationary stochastic process

Definition of stochastic process

A stochastic process is a function of time that varies with the random outcome of an experiment.

$$\underbrace{y(t)}_{\text{Short notation}} = \underbrace{y(t,\omega)}_{\text{Full notation}} \qquad \omega = \text{random outcome of an experiment}$$

Definition of stationary stochastic process

Let $y(t)$ be a stochastic process. We say $y(t)$ is stationary if for any set of time instances $(t_1, t_2, \ldots, t_k)$, the joint distribution of $(y(t+t_1), y(t+t_2), \ldots, y(t+t_k))$ is independent of t.

Example:

- $W(t)$ is a stochastic processes.
- $Z(t)=\dfrac{dW(t)}{dt}$ is a stochastic process.

  Note: For finite dt, dW/dt is well defined, with no complication at all.

Example:

- $W(t)$ is not stationary. $E\left(W(t)^2\right)=t$ varies with t.
- $Z(t)=\dfrac{dW(t)}{dt}$ is stationary

  Note: The joint distribution is invariant under a shift.

Properties of stationary stochastic process

For a stationary stochastic process, we have

- $E(y(t)) = E(y(0))$
- $E\left(y(s+\tau)\overline{y(s)}\right)=E\left(y(\tau)\overline{y(0)}\right)$

Caution:

These are necessary conditions for a stationary stochastic process.

They are not sufficient conditions.

Auto-correlation function

Definition of auto-correlation function

For a stationary stochastic process,

$$R(\tau)\equiv E\left(y(s+\tau)\overline{y(s)}\right)=E\left(y(\tau)\overline{y(0)}\right)$$

is called the auto-correlation function.

Note: $R(\tau)$ is a function of τ only, independent of s.

Caution: be careful with the term "auto-correlation"

Auto-correlation coefficient is defined as

$$\rho(\tau)\equiv\frac{E\left(\left[y(\tau)-E(y(0))\right]\left[\overline{y(0)-E(y(0))}\right]\right)}{\text{var}(y(0))}$$

Auto-correlation function is defined as

$$R(\tau)\equiv E\left(y(\tau)\overline{y(0)}\right)$$

Wiener-Khinchin theorem (relation between PSD and auto-correlation function)

For a stationary stochastic process, the power spectrum density (PSD) is

$$\text{PSD}\equiv\underbrace{s(\xi)}_{\substack{\text{New notation}\\ \text{for PSD}}}=\lim_{T\to\infty}\frac{E\left(\left|\int_{-T}^{T}\exp(-i2\pi\xi t)y(t)dt\right|^2\right)}{2T}$$

We use (PSD01) obtained above to rewrite $s(\xi)$ as

$$s(\xi) = \lim_{T\to\infty} \frac{E\left(\int_{-2T}^{2T} \exp(-i2\pi\xi\tau) \int_{a(\tau)}^{b(\tau)} y(\tau+s)\overline{y(s)}\,ds\,d\tau \right)}{2T}$$

Change the order of integration and taking average

$$= \lim_{T\to\infty} \frac{\int_{-2T}^{2T} \exp(-i2\pi\xi\tau) \int_{a(\tau)}^{b(\tau)} E\left(y(\tau+s)\overline{y(s)}\right) ds\,d\tau}{2T}$$

$$= \lim_{T\to\infty} \frac{\int_{-2T}^{2T} \exp(-i2\pi\xi\tau) R(\tau)(b(\tau)-a(\tau))\,d\tau}{2T}$$

The term $(b(\tau) - a(\tau))$ has the expression:

$$b(\tau)-a(\tau) = \begin{cases} 2T+\tau, & \tau\in[-2T,\ 0] \\ 2T-\tau, & \tau\in[0,\ 2T] \end{cases}$$
$$= 2T - |\tau|$$

Substituting it into the expression of $s(\xi)$ yields

$$s(\xi) = \lim_{T\to\infty} \int_{-2T}^{2T} \exp(-i2\pi\xi\tau) R(\tau)\left(1 - \frac{\tau}{2T} \right) d\tau$$

Taking the limit as $T \to \infty$, we arrive at

$$s(\xi) = \int_{-\infty}^{+\infty} \exp(-i2\pi\xi\tau) R(\tau)\,d\tau$$

We just derived the Wiener-Khinchin theorem.

Wiener-Khinchin theorem:

For a stationary stochastic process $y(t)$, the power spectrum density, $s(\xi)$, and the auto-correlation function, $R(t)$, are related by

$$s(\xi) = \int_{-\infty}^{+\infty} \exp(-i2\pi\xi t) R(t)\,dt$$

In other words, the PSD is Fourier transform of the auto-correlation function.

Definition of white noise

Let $y(t)$ be a stationary stochastic process. We say $y(t)$ is a white noise if

$$s(\xi) = const$$

In other words, the power of a white noise is uniformly distributed in the frequency dimension.

The Wiener-Khinchin theorem established above tells us that a white noise has two equivalent defining characters.

$$s(\xi) = const \qquad \Longleftrightarrow \qquad R(t) = E\left(y(t)\overline{y(0)}\right) \propto \delta(t)$$

Working out items in the short story

We re-write the short story in terms of the auto-correlation funvtion $R(\tau)$ and power spectrum density $s(\xi)$.

1) $Z(t) \equiv \dfrac{dW}{dt}$ is not a regular function.

2) $R(\tau) = E\left(Z(s+\tau)Z(s)\right) = \delta(\tau)$

3) $s(\xi) = \int \exp(-i2\pi\xi t)R(t)dt = 1$

4) $Z(t)$ is a white noise.

- To show $Z(t)$ is a white noise (item 4), we only need to show $s(\xi)$ = const (item 3).
- To show $s(\xi)$ = 1(item 3), we only need to show $R(t) = \delta(t)$ (item 2)

Thus, the remaining task is to show item 2, which we do now.

Derivation of $R(t) = \delta(t)$ (a "formal" derivation")

We first calculate $E(W(t)W(s))$ for $t \geq s$.

$$E\left(W(t)W(s)\right) = E\left((W(t)-W(s)+W(s))W(s)\right)$$

$$= E\left((W(t)-W(s))W(s)\right) + E\left(W(s)^2\right) = 0 + s = s$$

Since $E(W(t)W(s)) = E(W(s)W(t))$, we obtain

$$E\left(W(t)W(s)\right) = \min(t,s)$$

Next, in the calculation of $E(Z(t)Z(s))$, we "formally" exchange the order of taking derivatives and taking average.

$$E\left(Z(t)Z(s)\right) = E\left(\frac{\partial}{\partial s}\frac{\partial}{\partial t}\left(W(t)W(s)\right)\right)$$

$$= \frac{\partial}{\partial s}\frac{\partial}{\partial t}E\left(W(t)W(s)\right) = \frac{\partial}{\partial s}\frac{\partial}{\partial t}\min(t,s)$$

As a function of t, we have

$$\min(t,s)=\begin{cases} t, & t<s \\ s, & t>s \end{cases}$$

Differentiating with respect to t, and then writing it as a function of s, we get

$$\frac{\partial}{\partial t}\min(t,s)=\begin{cases} 1, & t<s \\ 0, & t>s \end{cases} \qquad \text{(as a function of } t\text{)}$$

$$=\begin{cases} 0, & s<t \\ 1, & s>t \end{cases} \qquad \text{(as a function of } s\text{)}$$

Differentiating with respect to s, we arrive at

$$\frac{\partial}{\partial s}\frac{\partial}{\partial t}\min(t,s)=\delta(s-t)$$

Therefore, we conclude

$$E\Big(Z(t)Z(s)\Big)=\delta(t-s)$$

$$\Longrightarrow \qquad R(\tau)=E\Big(Z(s+\tau)Z(s)\Big)=\delta(\tau)$$

<u>Expressions of $R(t)$ and $s(\xi)$ for **finite** dt</u>

(Justification of the "formal" derivations above)

To emphasize the finite time step, let $\Delta t \equiv dt$. We have

$$Z(t)=\frac{W(t+\Delta t)-W(t)}{\Delta t} \qquad \text{a well defined stationary stochastic process}$$

$$E\Big(Z(t)Z(s)\Big)=E\left(\frac{W(t+\Delta t)-W(t)}{\Delta t}\cdot\frac{W(s+\Delta t)-W(s)}{\Delta t}\right)$$

$$=\begin{cases} 0, & |t-s|>\Delta t \\ \dfrac{\Delta t-|t-s|}{(\Delta t)^2}, & |t-s|\le \Delta t \end{cases} \qquad \text{(derivation not included)}$$

$$R(\tau)=E\Big(Z(s+\tau)Z(s)\Big)=\begin{cases} 0, & |\tau|>\Delta t \\ \dfrac{\Delta t-|\tau|}{(\Delta t)^2}, & |\tau|\le \Delta t \end{cases}$$

Taking the Fourier transform of $R(t)$, we obtain

$$s(\xi) = \int \exp(-i2\pi\xi t) R(t)\, dt = \int_{-\Delta t}^{\Delta t} \exp(-i2\pi\xi t) \frac{\Delta t - |t|}{(\Delta t)^2} dt$$

$$= 2\frac{\cosh(i2\pi\xi\Delta t) - 1}{(i2\pi\xi\Delta t)^2} \qquad \text{(derivation not included)}$$

Now, <u>after</u> we obtain $s(\xi)$ for finite Δt, we take the limit as $\Delta t \to 0$.

At any fixed ξ, as $\Delta t \to 0$, we have

$$\lim_{\Delta t \to 0} s(\xi) = \lim_{\Delta t \to 0} 2\frac{\cosh(i2\pi\xi\Delta t) - 1}{(i2\pi\xi\Delta t)^2} = 1$$

<u>Observation:</u>

- Mathematically, working with finite *dt* until taking the limit at the end is a rigorous approach in which every step is properly justified. But deriving expressions with finite *dt* is also mathematically much more elaborated.
- "Formal" derivations are much simpler. But they are not rigorous.

<u>A class of colored noise</u>:

In the subsequent discussion of Ornstein-Uhlenbeck process (OU), we will see that its auto-correlation has the form:

$$R(t) = E\left(y(t)\overline{y(0)} \right) \propto \exp(-\beta |t|)$$

The corresponding power spectrum density is

$$s(\xi) = \int \exp(-i2\pi\xi t) R(t)\, dt \propto \int \exp(-i2\pi\xi t) \exp(-\beta |t|)\, dt$$

$$= \frac{2\beta}{\beta^2 + 4\pi^2 \xi^2}$$

End of discussion of white noise

Constrained Wiener process

For an unconstrained Wiener process, we have

$W(0) = 0$ and $W(t_1) \sim N(0, t_1)$

What happens if it is constrained by $W(t_1+t_2) = y$?

$W(t_1)$ is still random.

We like to know the conditional distribution $(W(t_1) \mid W(t_1+t_2) = y)$.

For that discussion, we need to introduce Bayes theorem.

Bayes Theorem

Consider two events A and B. We write Pr(A and B) in two ways.

Pr(A and B) = Pr(A | B) Pr(B)

Pr(A and B) = Pr(B | A) Pr(A)

Equating the two, we get

Pr(A | B) Pr(B) = Pr(B | A) Pr(A)

Express Pr(A | B) in terms of Pr(B | A), we arrive at

Bayes Theorem for events:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

This is Bayes theorem for events.

To derive Bayes theorem for densities, we consider

A = "$x \le X < x+\Delta x$

B = "$y \le Y < y+\Delta y$

We write probabilities in terms of densities

$$\Pr(A|B) \approx \rho(X=x|Y=y)\Delta x$$

$$\Pr(B|A) \approx \rho(Y=y|X=x)\Delta y$$

$$\Pr(A) \approx \rho(X=x)\Delta x$$

$$\Pr(B) \approx \rho(Y=y)\Delta y$$

Substituting these terms into Bayes theorem, we obtain Bayes theorem for densities.

Bayes theorem for densities

$$\rho(X=x|Y=y) = \frac{\rho(Y=y|X=x)\cdot\rho(X=x)}{\rho(Y=y)}$$

A useful trick:

In density $\rho(X=x|Y=y)$, x is the independent variable and y is a parameter. $\rho(Y=y)$ on the RHS is a function of y only, independent of x. It simply serves as a normalizing factor to make the RHS integrate to 1. Thus, we don't need to explicitly keep track of $\rho(Y=y)$. We can write Bayes theorem conveniently as

$$\boxed{\rho(X=x|Y=y) \propto \rho(Y=y|X=x)\cdot\rho(X=x)}$$

where the RHS needs a proper normalizing factor to make it integrate to 1.

This trick is especially convenient for normal distributions. Once we find

$$\rho(X=x) \propto \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right),$$

we can conclude $X \sim N(\mu, \sigma^2)$.

Conditional distribution $(W(t_1) \mid W(t_1+t_2) = y)$

The target is a probability conditioned on W at a later time. We use Bayes theorem to write it in terms of a probability conditioned on W at an earlier time.

Let $X = W(t_1)$ and $Y = W(t_1+t_2)$. We have

$$W(t_1) \sim N(0, t_1)$$

$$\Longrightarrow \quad \rho\big(W(t_1)=x\big) \sim N(0,t_1) \propto \exp\left(\frac{-x^2}{2t_1}\right)$$

$$W(t_1+t_2) = W(t_1) + \underbrace{\big(W(t_1+t_2)-W(t_1)\big)}_{\sim N(0,t_2)}$$

$$\Longrightarrow \quad \big(W(t_1+t_2)\big| W(t_1)=x\big) \sim x + N(0,t_2)$$

$$\Longrightarrow \quad \rho\big(W(t_1+t_2)=y \big| W(t_1)=x\big) \propto \exp\left(\frac{-(y-x)^2}{2t_2}\right)$$

The Bayes theorem gives us

$$\rho\big(W(t_1)=x \big| W(t_1+t_2)=y\big) \propto \rho\big(W(t_1+t_2)=y \big| W(t_1)=x\big)\cdot\rho\big(W(t_1)=x\big)$$

$$\propto \exp\left(\frac{-(y-x)^2}{2t_2}\right)\exp\left(\frac{-x^2}{2t_1}\right)$$

(We don't need to keep track of factors that are independent of x!)

$$\propto \exp\left(-\left(\left(\frac{1}{2t_1}+\frac{1}{2t_2}\right)x^2 - 2\frac{y}{2t_2}x\right)\right)$$

(Completing the square)

$$\propto \exp\left(\frac{-\left(x - \frac{t_1 y}{t_1+t_2} \right)^2}{2\frac{t_1 t_2}{t_1+t_2}} \right) \sim N\left(\frac{t_1 y}{t_1+t_2}, \frac{t_1 t_2}{t_1+t_2} \right)$$

We conclude

$$\boxed{\rho\Big(W(t_1)=x \Big| W(t_1+t_2)=y \Big) \sim N\left(\frac{t_1 y}{t_1+t_2}, \frac{t_1 t_2}{t_1+t_2} \right)}$$

<u>For the general case</u>, we have

$$\boxed{\rho\Big(W(a+t_1)=x \Big| W(a)=y_a \text{ and } W(a+t_1+t_2)=y_b \Big) \sim N\left(\frac{t_1 y_b + t_2 y_a}{t_1+t_2}, \frac{t_1 t_2}{t_1+t_2} \right)}$$

<u>A special case:</u> $t_1 = t_2 = h$

$$\rho\Big(W(a+h)=x \Big| W(a)=y_a \text{ and } W(a+2h)=y_b \Big) \sim N\left(\frac{y_a+y_b}{2}, \frac{h}{2} \right)$$