A COURSE ON INTEGRATION BY PARTS

YU GU

ABSTRACT. This is the lecture note for a course I taught in spring 2022 at the University of Maryland, College Park. It is mostly about the analysis on Gaussian space and how integration by parts can be useful in different contexts. The three examples we consider are (i) a sequence of i.i.d. standard Gaussians (ii) a standard Brownian motion (iii) a spacetime white noise. The applications include e.g. superconcentration, Stein's method, stochastic PDEs, density formula, etc. The note is based on various notes and books from Sourav Chatterjee, Martin Hairer, Davar Khoshnevisan, Nourdin-Peccati, David Nualart, etc. Nothing presented here is original except for a few typos and mistakes.

CONTENTS

1. Lecture 1

1.1. **The model of directed polymer in the random environment.** Let $\{\omega_{i,j}\}$ be i.i.d. random variables and S_n be a symmetric simple random walk starting from the origin. For each realization of the random environment and each N, consider the energy of the random walk collected until N:

$$H_N(S) = \sum_{i=1}^{N} \omega_{i,S_i}$$

Let $\beta > 0$ be the parameter, consider a new "path measure" on the random walk with the weight

$$\frac{e^{\beta H_N(S)}}{\sum_S e^{\beta H_N(S)}}$$

i.e., for each given path S, the new probability is given by the above formula. Here $\sum_S$ is the sum over all possible trajectories (2^N). We are interested in the behavior of the random walk under this new path measure. Two questions: (i) distribution of the endpoint

$$\hat{\mathbb{P}}(S_N \in A) = \frac{\sum_S 1_A(S_N) e^{\beta H_N(S)}}{\sum_S e^{\beta H_N(S)}}$$

Note that under the old measure $\mathbb{P}(S_N \in A) = 2^{-N} \sum_S 1_A(S_N)$ (this is the case of $\beta = 0$). Note that it is a random probability measure, and one can consider the average. (ii) partition function $Z_N = \sum_S e^{\beta H_N(S)}$, which is a random variable depending on the i.i.d. random variables. Question: distribution of the partition function. The perspective is to view it as a complicated "smooth" function:

$$\{\omega_{ij}\} \mapsto Z_N$$

How to compute the derivative here? Simple! Regular derivative. How to use the derivative to get useful information on Z_N? Size of the variance? Mean? How do they depend on N?

1.2. **The parabolic Anderson model.** $\partial_t u = \Delta u + V(x) u$. Try to understand the equation by separating the Δ and the potential term. Diffusion and reaction. When $V(x) \equiv \lambda$, exponential growth. Model $V(x)$ as a random process (simplest one, Gaussian process which has smooth sample paths, what is the meaning of Gaussian process). Assume that $u(0, x) = \delta(x)$, one particle starts diffusion and reproduction/branching/killing, $u(t,x)$ is the density of the particle at time t and location x. Again, the same perspective:

$$\{V(x)\}_{x\in\mathbb{R}^d} \mapsto u(t,0)$$

It is not so trivial when we try to determine the dependence of $u(t,0)$ on $V(x)$. Question: nonlinear, nonlocal map? (linear equation)

1.3. **Diffusion process.** $\{B_t\}_{t\geq 0}$ is a standard Brownian motion. Consider the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \qquad X_0 = 0$$

where $b(\cdot), \sigma(\cdot)$ are given functions. Question: distribution of X_t? Long time behavior of X_t? How does it depend on the driving force? Does it depend on the driving force continuously? Question asked by Wong-Zakai, B_t is not differentiable. Consider B_t^ε which is a mollification of B_t, and the SDE on the level of the mollification

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sigma(X_t^\varepsilon)dB_t^\varepsilon$$

Question: as $\varepsilon \to 0$, do we have $X_t^\varepsilon \to X_t$? Again

$$\{B_t\}_{t\geq 0} \mapsto X_t$$

whether the mapping is continuous. (the same question can be asked in the PDE context)

1.4. **Two examples.** (1) $(1 - u''(x)) = V(x)$, where V is a smooth Gaussian process. How to solve such ODE?

$$u(x) = \int G(x-y)V(y)dy$$

Question: distribution of $u(x)$? This is a simple example where the dependence of u on V is linear ($f \mapsto G \star f$). u is a Gaussian process whose mean and variance can be computed explicitly. This is the case when the driving force is additive, typically serve as a type of linearization of complicated dynamics.

(2) $dX_t = X_t dB_t, X_0 = 1$. The solution is explicit, Geometric Brownian motion. $X_t = e^{B_t - \frac{1}{2}t}$. Pretend we do not know the explicit solution

$$X_t = 1 + \int_0^t X_s dB_s, \qquad X_s = 1 + \int_0^s X_\ell dB_\ell$$

which implies

$$X_t = 1 + \int_0^t 1 dB_t + \int_0^t (\int_0^s X_\ell dB_\ell) dB_s$$

Keep iterating, we obtain $X_t = 1 + B_t + \frac{1}{2}(B_t^2 - t) + \ldots$. In other words, we somewhat wrote the solution explicitly. Here we have iterated Ito integral:

$$\int_0^t (\int_0^s dB_\ell) dB_s$$

This type of iterated Ito integral serves as the "Fourier basis" in $L^2(\Omega)$. Any random variable $X \in L^2(\Omega)$ can be written as a sum of iterated Ito integral (Wiener chaos expansion). What's so special about Fourier basis? Orthogonal! Another important part of the course is to do Fourier analysis in $L^2(\Omega)$. Why? Easy to compute L^2 norm/variance.

1.5. **Gaussian random vectors.** Definition of multidimensional Gaussian distribution. Wick formula. Why do we want to compute $\mathbb{E}[X_1X_2\dots X_N]$ when X_j are jointly Gaussian. Example of parabolic Anderson model. Proof of Wick formula. Assume N is even and $\mathbb{E}X_j = 0$, then

$$\mathbb{E}X_1\dots X_N = \sum_\sigma \prod_{(i,j)\in\sigma} \Gamma_{ij}$$

where Γ is the covariance matrix. Using moment generating function $f(t) := \mathbb{E}e^{t\cdot X} = \mathbb{E}e^{\sum_j t_jX_j} = e^{\frac{1}{2}t'\cdot\Gamma t}$, so we have $\mathbb{E}[X_1\dots X_N] = \partial_{t_1\dots t_N} f(t)\,|_{t=0}$.

2. Lecture 2

The following is based on Davar Khoshnevisan's note Chapter 1 and 2 http://www.math.utah.edu/~davar/math7880/F18/GaussianAnalysis.pdf

2.1. **Multidimensional Gaussian distribution and Wick formula.** $X = (X_1,\dots,X_N)$ is a Gaussian random vector in $\mathbb{R}^n$ if any linear combination of X_j is of Gaussian distribution. $\mu_j = \mathbb{E}X_j$ and $\Gamma_{ij} = \mathrm{Cov}(X_i,X_j) = \mathbb{E}[(X_i-\mu_i)(X_j-\mu_j)]$. Wick formula. How to compute $\mathbb{E}[X_1X_2\dots X_N]$ when X_j are jointly Gaussian. Proof of Wick formula. Assume N is even and $\mathbb{E}X_j = 0$, then

$$\mathbb{E}X_1\dots X_N = \sum_\sigma \prod_{(i,j)\in\sigma} \Gamma_{ij}$$

where Γ is the covariance matrix. Using moment generating function $f(t) := \mathbb{E}e^{t\cdot X} = \mathbb{E}e^{\sum_j t_jX_j} = e^{\frac{1}{2}t'\cdot\Gamma t}$, so we have $\mathbb{E}[X_1\dots X_N] = \partial_{t_1\dots t_N} f(t)\,|_{t=0}$. The detailed proof can be found in the note Chapter 1 Section 5. Homework problem: parabolic Anderson model. $\partial_t u = \Delta u + V(x)u$, where V is a Gaussian process with zero mean and covariance function $R(x-y) = \mathbb{E}[V(x)V(y)]$. Using Wick formula to write $\mathbb{E}[u(t,x)]$ as an infinite series (homework).

2.2. **Gaussian-Sobolev space.** We will work with $\mathbb{R}^n$ for some n fixed at the moment. It is not hard to see later that n can be infinity. Let $\mathbb{P}$ be the standard Gaussian measure on $\mathbb{R}^n$, i.e., zero mean and the covariance matrix is identity. $p(x) = (2\pi)^{-n/2}e^{-x^2/2}$ is the density. From now on, any function $f : \mathbb{R}^n \to \mathbb{R}$ can be viewed as a random variable, $\mathbb{P}[f \in A] = \int 1_{\{f(x)\in A\}}p(x)dx$. In other words we view $x \in \mathbb{R}^n$ as the random realization sampled from $\mathbb{R}^n$ according to the measure $\mathbb{P}$. One can think of $f(x)$ as $f(Z)$ where $Z \sim N(0, I_n)$ with I_n the $n\times n$ identity matrix.

Definition of C_0^k: C^k functions whose derivatives up to order k grow slowly than $e^{\varepsilon|x|^2}$ for any $\varepsilon > 0$, i.e., smooth function that grows not crazy. Lemma: if $f \in C_0^k$, then $\mathbb{E}|\partial^\alpha f|^p < \infty$ for any $p > 0$, where α is a multiindex so that $|\alpha| \le k$. Proof: since $f \in C_0^k$ we know that $|\partial^\alpha f| \le Ce^{\varepsilon|x|^2}$ for any $\varepsilon > 0$, here $C > 0$ is a constant. Then $\mathbb{E}|\partial^\alpha f|^p \le C\int e^{p\varepsilon|x|^2}(2\pi)^{-n/2}e^{-|x|^2/2}dx < \infty$.

Inner product and Hilbert space. For $f \in C_0^1$, define $\|f\|_{1,2}^2 = \mathbb{E}f^2 + \mathbb{E}|\nabla f|^2 = \int f(x)^2p(x)dx + \int|\nabla f(x)|^2p(x)dx$. Define inner product in the same way. Definition: The Gaussian-Sobolev space $D^{1,2}(\mathbb{P})$ is the completion of C_0^1 under $\|\cdot\|_{1,2}$. For any $f \in D^{1,2}(\mathbb{P})$, there exists $f_n \in C_0^1$ so that $f_n \to f$ in $D^{1,2}$. How is ∇f defined? Note that $\int|\nabla f_n - \nabla f_m|^2p(x)dx \to 0$ as $m,n \to \infty$ and since $L^2(\mathbb{P})$ is complete, there exists an $L^2(\mathbb{P})$ limit of ∇f_n as $n \to \infty$. Define the limit as ∇f.

Example: Lipchitz function. Assume f is Lipchitz, then $f \in D^{1,2}$. Proof: define $f_n = \phi_n \star f$, where ϕ_n is an approximation of identity. Since $|f(x)| \le |f(0)| + K|x|$, where K is the Lipchitz constant, we know that f_n, f grow at most linearly. Since $f_n \to f$ everywhere, we have $\int|f_n -$

$f|p(x)dx \to 0$. We also know that f is almost everywhere differentiable (a property of Lipchitz function), denote the derivative by ∇f which is almost everywhere bounded, one can show that $\nabla f_n = \nabla\phi_n \star f = \phi_n \star \nabla f \to \nabla f$ almost everywhere, which implies that $\int |\nabla f_n - \nabla f|p(x)dx \to 0$. The proof is complete. Note that Lipchitz function is not in the usual Sobolev space because it might not be integrable at infinity, but it is in the Gaussian-Sobolev space.

2.3. **Malliavin derivative.** For $f \in D^{1,2}$, we call $Df = \nabla f$ the Malliavin derivative. This looks like the usual derivative of f with respect to the $x-$variable, but one should really view $f = f(Z)$ as a random variable, where $Z = (Z_1, \ldots, Z_n)$ with Z_j i.i.d. Gaussian, then $Df = (\nabla_1 f, \ldots, \nabla_n f)$ where $\nabla_j f$ is viewed as the derivative of f with respect to Z_j. Remark: when $\mathbb{R}^n$ is endowed with the Lebesgue measure, then we can not choose $n = \infty$. However, with the Gaussian measure, one can imagine the case of $n = \infty$ – we know there exists a sequence of i.i.d. standard Gaussian random variables $(Z_1, Z_2, \ldots,)$. For any function: $f : \mathbb{R}^\infty \to \mathbb{R}$, we can view it as a random variable $f = f(Z_1, Z_2, \ldots,)$ which depends on $Z_1, Z_2, \ldots$. To quantify how f depends on each Z_i, we can compute $\nabla_i f$, which is the derivative of f with respect to the $i-$th component. Example: $f = \sum a_i Z_i$ where a_i are constant and $\sum_i a_i^2 < \infty$. In this case $\nabla_i f = a_i$. More complicated example: directed polymer, how to compute the Malliavin derivative (homework). Even more complicated example: the case when f can depend on infinitely many Z_i in a very complicated way. Another remark: in the $\mathbb{R}^n$ case $Df = (\nabla_1 f, \ldots, \nabla_n f)$ is an $\mathbb{R}^n$ valued random variable. In the more general case, the Malliavin derivative is a Hilbert space valued random variable (for example when $n = \infty$ and $f = f(Z_1, \ldots,)$, then $Df = (\nabla_1 f, \nabla_2 f, \ldots)$ is ℓ^2 valued random variable. Why? By definition we should have $\mathbb{E}|Df|^2 = \sum_i \int |\nabla_i f|^2 p(x) < \infty = \sum_i \mathbb{E}|\nabla_i f|^2 < \infty$, which implies that $\sum_i |\nabla_i f|^2 < \infty$ almost surely, thus $\nabla f = (\nabla_1 f, \nabla_2 f, \ldots) \in \ell^2$ almost surely, so we can view it as an ℓ^2 valued random variable).

Chain rule: If ψ is a smooth function with bounded derivative and $f \in D^{1,2}$, then $X = \psi(f) \in D^{1,2}$ and we have $DX = \psi'(f)Df$. (Note in Davar's note it is Lemma 1.7, where the assumption is weaker, the proof should go by approximation but is not completely trivial!)

2.4. **Adjoint operator.** (Integration by parts is one of the most important ideas in this course) Assuming that $f, g \in C_0^1$, then $\mathbb{E}[D_j f g] = \int \partial_{x_j} f(x) g(x) p(x) dx = -\int f(x)\partial_{x_j} g(x)p(x)dx + \int f(x) x_j p(x) dx$ and the last expression can be written as $\mathbb{E}[f(-D_j g + Z_j g)]$. In other words, we view D_j as an (unbounded) operator on L^2 (which only acts on a dense subset which is $D^{1,2}$), then the adjoint of D_j is given by $A_j := -D_j + Z_j$ where we view Z_j as a multiplication operator. Question: what is the domain of A_j? Before studying the domain of A_j, let us do the following. Let $\langle \cdot, \cdot \rangle$ be the inner product in $\mathbb{R}^n$, $f \in D^{1,2}$ and $g = (g_1, \ldots, g_n)$ with $g_j \in C_0^1$. Note that Df is a random variable that takes value in $\mathbb{R}^n$, so we can compute $\mathbb{E}\langle Df, g\rangle = \sum_j \mathbb{E}[D_j f g_j] = \sum_j \mathbb{E}[f A_j g_j]$ with A_j defined before. Define the operator δ by $\delta g = \sum_j A_j g_j$, which is called the divergence operator (compare it with the real divergence) and is the adjoint of D, then we can write in a compact form $\mathbb{E}\langle Df, g\rangle = \mathbb{E}[fAg]$. Note that we view D as an operator mapping $\mathbb{R}-$valued random variables to $\mathbb{R}^n-$valued random variables, and A, which is its adjoint, as an operator mapping $\mathbb{R}^n-$valued random variable to $\mathbb{R}-$valued random variable.(Homework problem: Davar's note, page 36 problem 10). For the domain of A_j, define $\mathrm{Dom}(A_j) = \{g \in D^{1,2}, A_j g \in L^2\}$. Claim that $\mathbb{E}D_j fg = \mathbb{E}fA_j g$ for any $f \in D^{1,2}, g \in \mathrm{Dom}(A_j)$. Proof by approximation: first prove the case of $f \in C_0^1$ and $g \in \mathrm{Dom}(A_j)$, then prove the case of $f \in D^{1,2}$.

3. Lecture 3

The following is based on Davar Khoshnevisan's note Chapter 3 http://www.math.utah.edu/~davar/math7880/F18/GaussianAnalysis.pdf

3.1. **Review of derivative operator and its adjoint.** Recall that we consider the Gaussian measure on $\mathbb{R}^n$ with the density $p(x) = (2\pi)^{-n/2}e^{-x^2/2}$. For any $f : \mathbb{R}^n \to \mathbb{R}$ we view it as a random variable $f = f(Z_1, \ldots, Z_n)$ where $Z_1, \ldots, Z_n$ are i.i.d. $N(0,1)$ random variables. The Malliavin derivative of f is defined as $Df = (D_1 f, \ldots, D_n f)$ where $D_j f = \partial_{x_j} f(Z_1, \ldots, Z_n)$. The adjoint operator of D_j is $A_j = -D_j + Z_j$ where we view Z_j as a multiplication operator. Z_j comes from the structure of the Gaussian measure, what if we have a general measure of the form $e^{-V(x)}$, where V is some given deterministic function (this appears in the study of Gibbs measure associated with the so-called $\nabla\phi$ model). Homework problem (the paper of Naddaf-Spencer, maybe in future).

3.2. **1d Hermite polynomial.** Motivation: for complicated random variables, typically the solution to some PDE with random forcing or partition function of statistical physics models with disorder, to compute the derivative might not be very easy, e.g., consider the parabolic Anderson model, the solution map is highly nonlinear and non-local, how to compute the Malliavin derivative? The chaos expansion enables us to write our interested quantity in a "Fourier" domain, in which computing the derivative and doing calculus is somewhat easier. Think about the real Fourier transform, which turns differentiation into multiplication and is very useful in the study of PDE.

The simplest case is $n = 1$ and $X = f(Z)$ where $Z \sim N(0,1)$. We know that the Hermite polynomial is an ONB of $L^2(\mathbb{R}, \gamma)$, i.e., the weighted L^2 space with the weight $\gamma(x) = (2\pi)^{-1/2}e^{-x^2/2}$. How to go to high dimensions? Tensorization. Observation: γ taking derivatives is just γ multiplied by polynomials. $\gamma'(x) = -x\gamma(x)$, $\gamma''(x) = -\gamma(x) + x^2\gamma(x)$. By induction we know that $\gamma^{(k)}(x) = (-1)^k H_k(x)\gamma(x)$.

We define H_k to be the $k-$th order Hermite polynomial. Properties: (i) $H_{k+1}(x) = xH_k(x) - H_k'(x)$; (ii) $H_{k+1}'(x) = (k+1)H_k(x)$; (iii) $H_k(-x) = (-1)^k H_k(x)$. Proof: (i) by definition $\gamma^k(x) = (-1)^k H_k(x)\gamma(x)$, taking derivative on both sides, we get $\gamma^{(k+1)}(x) = (-1)^k(H_k'(x) - xH_k(x))\gamma(x)$, but this should equal to $(-1)^{k+1}H_{k+1}(x)\gamma(x)$. (ii) by induction and (i). (iii) is trivial from the definition. Note that (ii) is important, it says that if $X = \sum_n c_n H_n(Z)$, then $DX = \sum_n c_n n H_{n-1}(Z)$, i.e., from an orthogonal expansion we obtain another orthogonal expansion.

Theorem: $\{H_k/\sqrt{k!}\}_{k\geq 0}$ form an ONB of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Proof: $\mathbb{E}H_k^2 = \mathbb{E}H_k A H_{k-1} = \mathbb{E}DH_k H_{k-1} = k\mathbb{E}H_{k-1}^2$, by induction, we know $\mathbb{E}H_k^2 = k!$. $\mathbb{E}H_n H_m = \mathbb{E}H_n A H_{m-1} = \mathbb{E}DH_n H_{m-1} = n\mathbb{E}H_{n-1}H_{m-1}$, iterating, we obtain $\mathbb{E}H_k = 0$ for some $k \geq 1$ (how to prove it? by definition $\mathbb{E}H_k = \int H_k(x)\gamma(x)dx = (-1)^k \int \gamma^{(k)}(x)dx = 0$.) How to show it is complete? Suppose there exists $X \in L^2$ such that $\mathbb{E}XH_k = 0$, then $\mathbb{E}XZ^k = 0$ for all k, this (together with an exercise in real analysis) shows that $X = 0$.

Chaos expansion: for any $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, we have $X = c_n H_n(Z)/\sqrt{n!}$ with $c_n = \frac{1}{\sqrt{n!}}\mathbb{E}[XH_n(Z)]$. Question: what is DX? Formally $DX = \sum_n c_n/\sqrt{n!}DH_n(Z) = \sum_n c_n/\sqrt{n!}nH_{n-1}(Z)$. What is the problem here? $X \in L^2$ is equivalent with $\sum_n c_n^2 < \infty$, it does not mean $\sum_n c_n^2 n < \infty$ (Homework problem: Fourier characterization of $D^{k,2}$ for any k, think about how the usual Sobolev space $H^k(\mathbb{R})$ is defined through Fourier transform) Question: what is $\mathbb{E}X$?

Gaussian-Poincare inequality: If $X \in D^{1,2}$, then $\mathrm{Var}X \le \mathbb{E}|DX|^2$. By the chaos expansion $X = \sum_n c_n H_n/\sqrt{n!}$, we have $\mathrm{Var}X = \sum_{n\ge1} c_n^2$, and $DX = \sum_{n\ge1} c_n n/\sqrt{n!}H_{n-1}$, so $\mathbb{E}|DX|^2 = \sum_{n\ge1} c_n^2 n$. (another way of deriving the expansion coefficient of DX: for any n, we have $\mathbb{E}DXH_n/\sqrt{n!} = \mathbb{E}XAH_n/\sqrt{n!} = \mathbb{E}XH_{n+1}/\sqrt{n!} = \mathbb{E}XH_{n+1}/\sqrt{(n+1)!}\sqrt{n+1}$) Remark: the inequality is sharp, when does it give an equality? Remark: a different way of proving Gaussian Poincare by interpolation. Important question: when does Gaussian-Poincare give or not give you the "optimal" bound? Suppose we have a family of random variables $\{X_\varepsilon\}$, we want to derive optimal variance estimate $\mathrm{Var}X_\varepsilon$ as $\varepsilon \to 0$. Write $X_\varepsilon = \sum_n c_n^\varepsilon H_n/\sqrt{n!}$, then $\mathrm{Var}X_\varepsilon = \sum_n |c_n^\varepsilon|^2 \le \sum_n |c_n^\varepsilon|^2 n = \mathbb{E}|DX_\varepsilon|^2$. When the randomness "moves" to higher and higher order chaos. Question: what does it mean when $\mathrm{Var}X_\varepsilon \ll \mathbb{E}|DX_\varepsilon|^2$, what is a possible meaning of the ratio $\frac{\mathbb{E}|DX_\varepsilon|^2}{\mathrm{Var}X_\varepsilon}$? Question: asymptotic distribution? Fourth moment theorem. Question: how to compute high order moments? e.g. $\mathbb{E}X^4$?

Stroock formula: Suppose that $X \in D^{\infty,2}$ (think about what this means in terms of the expansion coefficients $\{c_n\}$), how to compute c_n? By definition $c_n = \mathbb{E}XH_n/\sqrt{n!}$, but this is not feasible in many cases. Sometimes computing the derivatives is easier than computing the expansion coefficient by definition. $DX = \sum_{n\ge1} c_n n/\sqrt{n!}H_{n-1}$, so $\mathbb{E}DX = c_1$, how about computing higher order derivatives?

Example: directed polymer in hw1.

3.3. **Multi-dimensional Hermite polynomial.** Lemma: if $X, Y \sim N(0,1)$ are jointly Gaussian, then $\mathbb{E}H_n(X)H_m(Y) = 0 1_{n\neq m} + n!\rho^n 1_{n=m}$ where $\rho = \mathbb{E}[XY]$. Proof: claim $F(t,x) = e^{tx-\frac12 t^2} = \sum_n t^n/n! H_n(x)$ (easily seen from Stroock formula). We have $\mathbb{E}F(s,X)F(t,Y) = \mathbb{E}e^{sX+tY-\frac12 s^2-\frac12 t^2} = e^{st\rho}$. On the other hand, we can write it as an infinite series expansion, match them to complete the proof.

Let $\{Z_k\}_{k\ge1}$ be a sequence of i.i.d. $N(0,1)$ random variables, and $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space generated by them. Define multiindex $a = (a_1, a_2, ...)$ with $a_j \ge 0$ and only finitely many nonzero a_j, and $|a| = \sum_j a_j$ and $a! = \prod_j a_j!$. Let Λ be the set of all multiindex. For any $a \in \Lambda$, define $\Phi_a = \frac{1}{\sqrt{a!}}\prod_j H_{a_j}(Z_j)$. Theorem: $\{\Phi_a\}_{a\in\Lambda}$ is an ONB of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Proof: ON is easy to show, using the properties of Hermite polynomial and independence of Z_k. let us show it is a basis. Suppose there exists $X \in L^2$ so that $\mathbb{E}X\Phi_a = 0$ for all $a \in \Lambda$. This would imply that $\mathbb{E}Xe^{i\sum_{j=1}^N \theta_j Z_j} = 0$ for any $N \ge 0$ and θ_j. This implies that X is independent of $\{Z_j\}_{j=1}^N$. Since N is arbitrary this implies X is independent of $\mathcal{F}$ but since X is measurable with respect to $\mathcal{F}$, we have $X = 0$ (note that $\mathbb{E}X = 0$).

Suppose $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, and $X = \sum_a c_a \Phi_a$, how do we compute DX? (First of all, what is DX? $DX = (D_1 X, \ldots,)$ where $D_k X$ is the derivative of X with respect to Z_k) We have (formally) $D_k X = \sum_a c_a D_k \Phi_a$. Recall that $\Phi_a = \frac{1}{\sqrt{a!}}\prod_j H_{a_j}(Z_j)$, so $D_k\Phi_a = \frac{1}{\sqrt{a!}}\prod_{j\neq k} H_{a_j}(Z_j) \times a_k H_{a_k-1}(Z_k)$. One can easily check that $\mathrm{Var}X \le \mathbb{E}|DX|^2$, where $|DX|^2 = \sum_k |D_k X|^2$.

3.4. **Wiener chaos expansion.** Let $\mathcal{H}_n$ be the closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ spanned by random variables of the form $\{\Phi_a : a \in \Lambda, |a| = n\}$. Claim (i) $\mathcal{H}_n \perp \mathcal{H}_m$ for any $n \neq m$ (ii) for any $n \ge 1$, $\mathcal{H}_n$ is infinite dimensional. We call $\mathcal{H}_n$ the $n-$th order Wiener chaos. Question: what is $\mathcal{H}_1$? Theorem: $L^2(\Omega, \mathcal{F}, \mathbb{P}) = \bigoplus_{n=0}^\infty \mathcal{H}_n$. Question: parabolic Anderson model, what should be the corresponding Wiener chaos here?

4. Lecture 4

The following is based on Davar Khoshnevisan's note Chapter 4 http://www.math.utah.edu/~davar/math7880/F18/GaussianAnalysis.pdf

4.1. **Review of Wiener chaos expansion.** Let $\mathcal{H}_n$ be the closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ spanned by random variables of the form $\{\Phi_a : a \in \Lambda, |a| = n\}$. Claim (i) $\mathcal{H}_n \perp \mathcal{H}_m$ for any $n \neq m$ (ii) for any $n \geq 1$, $\mathcal{H}_n$ is infinite dimensional. We call $\mathcal{H}_n$ the $n-$th order Wiener chaos. Question: what is $\mathcal{H}_1$? Theorem: $L^2(\Omega, \mathcal{F}, \mathbb{P}) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n$. Question: parabolic Anderson model, what should be the corresponding Wiener chaos here? Important remark: the chaos expansion for Z^2 is $Z^2 = 1 + (Z^2 - 1) = H_0(Z) + H_2(Z)$, i.e., Z^2 does not live on the 2nd order chaos. Similarly, we write $Z^3 = 3Z + (Z^3 - 3Z) = 3H_1(Z) + H_3(Z)$, so Z^3 does not live on the third chaos. In a sense, one should view $\mathcal{H}_n$ as a "recentered" $n-$th order polynomial. For the parabolic Anderson model $\partial_t u = \Delta u + uV(x)$, we can write $u(t, \cdot) = \sum_{n \geq 0} u_n(t, \cdot)$ where $u_n(t) = \mathcal{G}V\mathcal{G}V...\mathcal{G}u_0$ (we used the simplified notation), i.e., $u_n(t, \cdot)$ contains n factors of V. However, in general u_n does not live on the $n-$th order chaos (imagine that $V(\cdot)$ is constructed from a sequence of i.i.d. Gaussian so $u(t, x) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$). To see it, we can compute $\mathbb{E}u_2 \neq 0$. A very useful thing to do is to turn the expansion $\sum_n u_n$, which is obtained from a direct iteration of the mild formulation of the PDE, into a real chaos expansion. This can be done through a Feynman-Kac formula and the Stroock formula, which we will do later. (Example: $X_t' = X_t \xi_\varepsilon$ where $\xi_\varepsilon' = \frac{d}{dt} B_t^\varepsilon$ with B_t^ε a mollification of Brownian motion, how to write X_t in a chaos expansion?)

4.2. **Ornstein-Uhlenbeck operator.** Review of basics of OU process. $dX_t = -\lambda X_t dt + dB_t$ with $X_0 = 1$, the solution is explicit $X_t = e^{-\lambda t} X_0 + \int_0^t e^{-\lambda(t-s)} dB_s$. Gaussian process with explicit mean and variance. As $t \to \infty$, $X_t \Rightarrow N(0, \sigma^2)$ with $\sigma^2 = 1/(2\lambda)$. (question: strong or weak convergence?) Invariant measure is given by $N(0, \sigma^2)$. Assume $X_0 \sim N(0, \sigma^2)$ is independent of B, then $(X_0, X_t) \Rightarrow (Y_1, Y_2)$ with Y_1, Y_2 independent. We view the OU process as an interpolation between two independent Gaussian. Suppose we want to compute variance or covariance, say $\mathrm{Cov}[f(X_0)g(X_0)]$ with $X_0 \sim N(0, \sigma^2)$, claim $\mathrm{Cov}[f(X_0)g(X_0)] = \mathbb{E}f(X_0)[g(X_0) - g(X_\infty)]$. Write $g(X_\infty) - g(X_0) = \int_0^\infty ?$ (what is ?) Sometimes we need to interpolate two independent (infinite dimensional) Gaussian process, so we need to know how to define the corresponding OU process.

Consider the $\mathbb{R}^n$ case, where the Malliavin derivative and the adjoint operator were defined. Define the OU operator $L = -\sum_{j=1}^n A_j D_j$. For nice functions (random variables), $Lg(x) = -\sum_{j=1}^n (x_j - \partial_j)\partial_j g(x) = \Delta g - x \cdot \nabla g$. The operator $\Delta - x \cdot \nabla$ is the generator of the OU process. Viewed as random variables, we write $Lg(Z) = \sum_{j=1}^n D_j^2 g(Z) - \sum_{j=1}^n Z_j D_j g(Z) = \sum_{j=1}^n D_j^2 g(Z) - Z \cdot Dg(Z)$. Define $\mathrm{Dom}L = \{g \in D^{2,2} : Z \cdot Dg \in L^2\}$. For nice random variables f, g, $\mathbb{E}fLg = -\mathbb{E}f \sum_j A_j D_j g = -\sum_j \mathbb{E}D_j f D_j g = \mathbb{E}Lfg$, so the operator is self-adjoint.

How does L act on Hermite polynomials? For multiindex k, define $H_k(x) = \prod_j H_{k_j}(x_j)$, claim $LH_k(x) = -|k|H_k(x)$. Proof: $A_j D_j H_{k_j}(x_j) = A_j[k_j H_{k_j - 1}(x_j)] = k_j H_{k_j}(x_j)$, so $LH_k(x) = \sum_j k_j H_k(x) = |k|H_k(x)$. Thus, the hermite polynomial are eigenvectors of L. For nice random variables f, write it as $f = \sum_k c_k H_k(Z)/\sqrt{k!}$, so $Lf = -\sum_k c_k |k| H_k(Z)/\sqrt{k!}$, which implies that for f to be in the domain of L, we need $\sum_k c_k^2 |k|^2 < \infty$. (when $n = 1$ compare it with the domain of D)

4.3. **Heat equation.** We should compare L with Δ (when the underlying measure is Lebesgue). Let us consider the heat equation. For any $f \in L^2$, we can write it as $f = \sum_k \langle f, H_k \rangle H_k / k!$ where

$\langle\rangle$ denotes the inner product in L^2. Define the heat flow operator $P_t f = \sum_k e^{-|k|t}\langle f, H_k\rangle H_k/k!$. Lemma: if $f \in \mathrm{Dom}(L)$, then $P_t f \in \mathrm{Dom}(L)$, and $u_t := P_t f$ is the unique L^2 valued solution to $\frac{du_t}{dt} = Lu_t$ with $u_0 = f$. Proof: $P_t f \in \mathrm{Dom}(L)$ is obvious. Let us show u_t solves the equation. $\frac{u_{t+\varepsilon}-u_t}{\varepsilon} = \sum_k \frac{e^{-|k|\varepsilon}-1}{\varepsilon} e^{-|k|t}\langle f, H_k\rangle H_k/k!$. Only need to show the rhs goes to Lu_t (by dominated convergence theorem, how?). How to show uniqueness? By linearity of the equation assume $u_0 = 0$, consider $\frac{d}{dt}\langle u_t, u_t\rangle = 2\langle Lu_t, u_t\rangle = -2\langle Du_t, Du_t\rangle \le 0$ which implies that $\langle u_t, u_t\rangle = 0$ so $u_t = 0$.

Properties of the heat semigroup: (i) for each $t > 0$, P_t is a contraction on L^2 (ii) P_t is self-adjoint on L^2 (iii) P_t is a semigroup (iv) for any $f \in L^2$, $P_t f \to \mathbb{E}f$ in L^2 as $t \to \infty$.

4.4. **Variance/covariance representation.** A general question, how to compute (not estimate) covariance/variance? Take any f, g assume WLOG that $\mathbb{E}f = \mathbb{E}g = 0$, how to compute $\mathbb{E}fg$? By hermite polynomial expansion, we have $f = \sum_k \langle f, H_k\rangle H_k/k!$ and $g = \sum\langle g, H_k\rangle H_k/k!$, which implies that $\mathbb{E}fg = \sum_k \langle f, H_k\rangle\langle g, H_k\rangle \frac{1}{k!}$. In some cases, we do not have access to all the chaos expansion coefficients. Can we find a better covariance formula? Covariance lemma: if $f, g \in D^{1,2}$, we have $\mathrm{Cov}[f, g] = \int_0^\infty e^{-t}\sum_j \mathbb{E}[D_j f P_t D_j g] = \sum_j \mathbb{E}[D_j f(1-L)^{-1}D_j g]$. Remark: it directly implies the Poincare inequality by an application of Cauchy-Schwarz. Proof: consider the simple case of $n = 1$, by chaos expansion, we have $Dg = \sum\langle g, H_k\rangle H_{k-1}/(k-1)!$, and $P_t Dg = \sum\langle g, H_k\rangle e^{-(k-1)t}H_{k-1}/(k-1)!$. Similarly we have $Df = \sum\langle f, H_k\rangle H_{k-1}/(k-1)!$. This implies $\mathbb{E}DfP_tDg = \sum\langle g, H_k\rangle\langle f, H_k\rangle e^{-(k-1)t}/(k-1)!$. Multiplying e^{-t} integrating in t completes the proof. Another proof based on interpolation: $\mathrm{Cov}[f, g] = \mathbb{E}[fg] - \mathbb{E}[f]\mathbb{E}[g] = \mathbb{E}(f(Z)[g(Z) - \mathbb{E}g])$. Run OU process to interpolate: write $\mathbb{E}g = P_\infty g$ and $g = P_0 g$, so $g(Z) - \mathbb{E}g = P_0 g - P_\infty g = -\int_0^\infty \frac{d}{dt}P_t g dt = -\int_0^\infty LP_t g dt$, so $\mathrm{Cov}[f, g] = -\int_0^\infty \mathbb{E}[fLP_t g]dt = \int_0^\infty \mathbb{E}[DfDP_t g]dt$. It is easy to show that $DP_t = e^{-t}P_t D$. This completes the proof. Remark: estimates on the covariance: $\mathrm{Cov}[f, g] \le \sum_j \sqrt{\mathbb{E}|D_j f|^2\mathbb{E}|D_j g|^2}$ (note that applying another cauchy-schwarz will lead to something useless).

Example: PDE with random coefficient: $\nabla a(x)\nabla\phi = f$ where a is random. We want to study how fact ϕ decorrelates, i.e., to estimate $\mathrm{Cov}[\phi(x), \phi(y)]$ when $|x-y| \gg 1$. By the covariance formula $\mathrm{Cov}[\phi(x), \phi(y)] = \mathbb{E}\langle D\phi(x), (1-L)^{-1}D\phi(y)\rangle$. Only need to estimate the Malliavin derivative $D\phi(x)$, which comes from differentiating with respect to the randomness on the level of the equation. Homework problem based on it.

5. Lecture 5

The following is based on Davar Khoshnevisan's note Chapter 4 http://www.math.utah.edu/~davar/math7880/F18/GaussianAnalysis.pdf

5.1. **Review of the OU and heat flow operator, Mehler's formula.** Recall that $L = -\sum_{j=1}^n A_j D_j$ (why there is a negative sign). Written it in the usual notation, $L = -\sum_{j=1}^n(-\partial_{x_j} + x_j)\partial_{x_j} = \Delta - x\cdot\nabla$, which is the generator of the OU process $dX_t = -X_t dt + \sqrt{2}dB_t$ (which is written in the vector form, here B is a $n-$dimensional Brownian motion). Remark: question, for SDE $dX_t = b(X_t)dt + \sigma(X_t)dB_t$, how to write down its generator? In the Fourier domain, how to write down L? For any Hermite polynomial H_k, we have $LH_k = -|k|H_k$. For $f = \sum_k\langle f, H_k\rangle H_k/k!$, we have $Lf = -\sum_k |k|\langle f, H_k\rangle H_k/k!$ (from which we easily find the domain of L). Another way to look at what we are doing here: we consider the space of $L^2(\mathbb{R}^n, p(x))$, i.e., the L^2 space with the Gaussian weight and $\{H_k\}$ turns to be an ONB of this space which diagonalize L. Heat equation: $\frac{d}{dt}u(t) = Lu(t)$ with $u(0) = f \in L^2$. The unique solution is given by $u(t) = \sum_k e^{-|k|t}\langle f, H_k\rangle H_k/k!$.

We can write it in the form of an usual PDE $\partial_t u(t,x) = \Delta u(t,x) - x\cdot\nabla u(t,x)$ with $u(0,x)=f(x)$. Proof of uniqueness: by linearity of the equation assume $u_0=0$, consider $\frac{d}{dt}\langle u_t,u_t\rangle = 2\langle Lu_t,u_t\rangle = -2\langle Du_t,Du_t\rangle \le 0$ which implies that $\langle u_t,u_t\rangle=0$ so $u_t=0$. Question: probabilistic representation of the solution to PDE? Simple example: heat equation $\partial_t u=\Delta u$ with $u(0,x)=f(x)$, we have $u(t,x)=\mathbb{E}_x f(\sqrt{2}B_t)=\mathbb{E}f(x+\sqrt{2}Z')=\int q_t(x-y)f(y)dy$, where q is the Green's function of $\partial_t-\Delta$, $\mathbb{E}_x$ is the expectation with respect to the Brownian motion starting at x, and $Z'\sim N(0,I_n)$. Similarly, for $\partial_t u=\Delta u - x\cdot\nabla u$ we have $u(t,x)=\mathbb{E}_x f(X_t)=\mathbb{E}f(xe^{-t}+\sqrt{1-e^{-2t}}Z')$. Remark: how about the general diffusion process? What is the corresponding probabilistic representation? backward-Kolmogorov equation!

There is a name for $u(t,x)=\mathbb{E}_x f(X_t)=\mathbb{E}f(xe^{-t}+\sqrt{1-e^{-2t}}Z')$, which is the so-called Mehler's formula. In probability space, we write it as $u(t)=\mathbb{E}f(Ze^{-t}+\sqrt{1-e^{-2t}}Z'\,|\,Z)$, where Z' is an independent copy of Z. In other words, to get the solution to the heat equation, we freeze the original randomness (which is Z here), introduce an independent copy, consider the interpolation depending on t, and average out the new randomness. This is very useful in practice, where the chaos expansion is not available. For example, parabolic Anderson model $\partial_t u=\frac{1}{2}\Delta u+V(x)u$ with $u(0,x)=1$, consider the random variable $X=u(T,x)$ for some given (T,x), how to compute $P_t u(T,x)$. We write $u(T,x)=\mathbb{E}_B e^{\int_0^T V(x+B_s)ds}$, claim $P_t u(T,x)=\mathbb{E}_{B,V'}e^{\int_0^T e^{-t}V(x+B_s)ds+\int_0^T\sqrt{1-e^{-2t}}V'(x+B_s)ds}$. Here V' is an independent copy of V and $\mathbb{E}_{B,V'}$ is the expectation with respect to both B and V'. Question: how to compute the expectation with respect to V'? Remark: combine with hypercontractivity

5.2. **Covariance representation.** A general question, how to compute (not estimate) covariance/variance? Take any f,g assume WLOG that $\mathbb{E}f=\mathbb{E}g=0$, how to compute $\mathbb{E}fg$? By hermite polynomial expansion, we have $f=\sum_k\langle f,H_k\rangle H_k/k!$ and $g=\sum\langle g,H_k\rangle H_k/k!$, which implies that $\mathbb{E}fg=\sum_k\langle f,H_k\rangle\langle g,H_k\rangle\frac{1}{k!}$. In some cases, we do not have access to all the chaos expansion coefficients. Can we find a better covariance formula? Covariance lemma: if $f,g\in D^{1,2}$, we have $\mathrm{Cov}[f,g]=\int_0^\infty e^{-t}\sum_j\mathbb{E}[D_jfP_tD_jg]=\sum_j\mathbb{E}[D_jf(1-L)^{-1}D_jg]$. Remark: it directly implies the Poincare inequality by an application of Cauchy-Schwarz. Proof: consider the simple case of $n=1$ (the general case just with heavier notations), by chaos expansion, we have $Dg=\sum\langle g,H_k\rangle H_{k-1}/(k-1)!$, and $P_tDg=\sum\langle g,H_k\rangle e^{-(k-1)t}H_{k-1}/(k-1)!$. Similarly we have $Df=\sum\langle f,H_k\rangle H_{k-1}/(k-1)!$. This implies $\mathbb{E}DfP_tDg=\sum\langle g,H_k\rangle\langle f,H_k\rangle e^{-(k-1)t}/(k-1)!$. Multiplying e^{-t} integrating in t completes the proof. Another proof based on interpolation: $\mathrm{Cov}[f,g]=\mathbb{E}[fg]-\mathbb{E}[f]\mathbb{E}[g]=\mathbb{E}(f(Z)[g(Z)-\mathbb{E}g])$. Run OU process to interpolate: write $\mathbb{E}g=P_\infty g$ and $g=P_0g$, so $g(Z)-\mathbb{E}g=P_0g-P_\infty g=-\int_0^\infty\frac{d}{dt}P_tgdt=-\int_0^\infty LP_tgdt$, so $\mathrm{Cov}[f,g]=-\int_0^\infty\mathbb{E}[fLP_tg]dt=\int_0^\infty\mathbb{E}[DfDP_tg]dt$. It is easy to show that $DP_t=e^{-t}P_tD$. This completes the proof. Remark: estimates on the covariance: $\mathrm{Cov}[f,g]\le\sum_j\sqrt{\mathbb{E}|D_jf|^2\mathbb{E}|D_jg|^2}$ (note that applying another cauchy-schwarz will lead to something useless).

5.3. **Application of covariance representation: PDE with random coefficient.** In many applications we would like to understand the correlation properties of certain random fields, which can be modeled as solutions to some PDE subjected to random perturbations. For example, the standard homogenization problem $\nabla\cdot a(x)\nabla u(x)=f(x)$ where a models the heat conductance and u represents the temperature here, f is the source. We model a as some random field and assume that we know the statistical properties of $a(\cdot)$. The goal is to study the statistical properties of $u(\cdot)$, e.g., $\mathrm{Cov}[u(x),u(y)]$ when $|x-y|\gg 1$. A direct way is to write $u(x)=\int G(x,z)f(z)dz$ where G is the Green's function, then the problem reduces to understand the correlation properties of the Green's function. Here we proceed differently, using the covariance representation. To simplify the presentation, consider a simpler setting: $(1-a(x)\Delta)u(x) = f(x), x \in \mathbb{Z}$, where Δ is the discrete Laplacian. Assume $a(x) = g(\xi_x)$ for some nice g and $\{\xi_x\}_{x\in\mathbb{Z}}$ is a sequence of i.i.d. standard Gaussian. We have

$$\mathrm{Cov}[u(x), u(y)] = \int_0^\infty e^{-t} \sum_z \mathbb{E}[D_z u(x) P_t D_z u(y)]dt = \sum_z \mathbb{E}[D_z u(x)(1-L)^{-1} D_z u(y)]$$

where D_z denotes the derivative with respect to ξ_z. One can imagine that to study the asymptotic behavior as $|x-y| \gg 1$ we need to study $D_z u(x)$ when $|x-z| \gg 1$. The key is to compute and estimate $D_z u(x)$. Directly compute the derivative on the level of equation (note there are two types of derivatives here: one "horizontal" and the other "verticle") we obtain $D_z u(x) - D_z(a(x)\Delta u(x)) = 0$. We have $D_z(a(x)\Delta u(x)) = 1_{x=z} g'(\xi_z)\Delta u(x) + a(x)\Delta D_z u(x)$ (the dependence of a on ξ is local, while the dependence of u on ξ is nonlocal), so we have $(1-a(x)\Delta)D_z u(x) = 1_{x=z} g'(\xi_z)\Delta u(x)$. In other words, for each given z, we derived the equation satisfied by $D_z u(x)$ as a function of x. Using the Green's function representation, we have $D_z u(x) = \sum_w G(x,w) 1_{w=z} g'(\xi_z)\Delta u(w) = G(x,z)g'(\xi_z)\Delta u(z)$. One can further express $\Delta u(z)$ in terms of the Green's function (e.g. if $f(x) = 1_{x=0}$, we have $u(z) = G(z,0)$ so $\Delta u(z) = \Delta G(z,0)$) In this way we can write

$$\mathrm{Cov}[u(x), u(y)] = \sum_z \mathbb{E}[G(x,z)g'(\xi_z)\Delta u(z)(1-L)^{-1} G(y,z) g'(\xi_z)\Delta u(z)]$$

and the problem reduces to studying the asymptotics of $G(x,z)$ when $|x-z| \gg 1$. If there is some averaging taking place, we can replace G by some deterministic averaging. (research project)

6. Lecture 6

The following is based on Davar Khoshnevisan's note Chapter 5 http://www.math.utah.edu/~davar/math7880/F18/GaussianAnalysis.pdf

6.1. **Covariance formula and concentration inequality.** Recall the covariance representation formula: for any $f, g \in D^{1,2}$, we have $\mathrm{Cov}[f,g] = \int_0^\infty e^{-t}\mathbb{E}\langle Df, P_t Dg\rangle$, which holds in all dimensions. Consider $f \in D^{1,2}$ and ϕ is a smooth function with bounded derivative, then we have $\mathrm{Cov}[f, \phi(f)] = \mathbb{E}[\phi'(f)\int_0^\infty e^{-t}\langle Df, P_t Df\rangle dt] = \mathbb{E}[\phi'(f)\langle Df, (1-L)^{-1}Df\rangle]$. Given a random variable f, how concentrated it is around its mean $\mathbb{E}f$? If f is Lipchitz with respect to the underlying Gaussian random variable with the Lipchitz constant C_f, then by Gaussian-Poincare inequality we know that $\mathrm{Var}X \le \mathbb{E}\langle DX, DX\rangle \le C_f^2$, so by Chebyshev we have $\mathbb{P}[|f - \mathbb{E}f| \ge t] \le t^{-2}C_f^2$. There is a stronger version of it: $\mathbb{P}[|f-\mathbb{E}f| \ge t] \le 2e^{-t^2/(2C_f^2)}$ (the decay as t increases is clearly much faster). Proof: WLOG, assume that $\mathbb{E}f = 0$ and $C_f = 1$, in which case we only need to show that $\mathbb{P}[|f| \ge t] \le 2e^{-t^2/2}$. By symmetry we consider $\mathbb{P}[f \ge t]$. For any $\lambda > 0$, we have $\mathbb{P}[f \ge t] = \mathbb{P}[e^{\lambda f} \ge e^{\lambda t}] \le e^{-\lambda t}\mathbb{E}[e^{\lambda f}]$ (in other words, we apply Chebyshev with a different test function). It suffices to estimate $\mathbb{E}e^{\lambda f}$ then optimize over $\lambda > 0$. Since $\mathbb{E}f = 0$, we have $\mathbb{E}fe^{\lambda f} = \mathrm{Cov}[f, e^{\lambda f}] = \lambda\mathbb{E}[e^{\lambda f}\langle Df, (1-L)^{-1}Df\rangle] \le \lambda\mathbb{E}[e^{\lambda f}]$ (note that $e^{\lambda f}$ has good integrability because f is Lipchitz). Denote $g(\lambda) = \mathbb{E}e^{\lambda f}$, then we have $g' \le \lambda g$ which implies that $g(\lambda) \le e^{\lambda^2/2}$. Thus, we have $\mathbb{P}[f \ge t] \le e^{-\lambda t + \lambda^2/2}$. Choose $\lambda = t$ we complete the proof. Remark: this concentration inequality is typically combined with Borel-Cantelli to prove almost sure convergence: for example, if we want to show f_n converges almost surely, (i) we first show $\mathbb{E}f_n$ converges (ii) we show $\mathbb{P}[|f_n - \mathbb{E}f_n| > \delta_n] \le \varepsilon_n$ for some ε_n so that $\sum_n \varepsilon_n < \infty$, where $\delta_n \to 0$ is chosen carefully. Homework problem: directed polymer show that $\mathbb{P}[|n^{-1}(\log Z_n - \mathbb{E}\log Z_n)| > t] \le 2e^{-nt^2/2}$

6.2. **Application of concentration of measure to spin glass.** What is the SK model in Spin glass? For any n, the configuration is $\{\pm 1\}^n$, i.e., a vector of length n with ± 1 entries. We write such vector by $\sigma = (\sigma_i)_{i=1,\ldots,n}$ with $\sigma_i = \pm 1$. We view σ_i as the spin at site i. The interaction between different spins are modeled by a Gaussian random variable x_{ij}, and the Hamiltonian is $H_n(\sigma, x) = \frac{1}{\sqrt{n}} \sum_{i<j} \sigma_i \sigma_j x_{i,j}$. We view $x = \{x_{ij}\}$ as the random environment determining the interaction between different spins. Given the realization of x, the Gibbs measure on the spin configuration is defined as

$$\mathbb{P}_n[\sigma] = \frac{e^{\beta H_n(\sigma,x)}}{\sum_\sigma e^{\beta H_n(\sigma,x)}}$$

where β is the inverse temperature. This is the so-called sherrington kirkpatrick model. The goal is to study the behavior of the Gibbs measure when n is very large. The key quantity is the partition function

$$Z_n(x) = \sum_\sigma e^{\beta H_n(\sigma,x)}$$

One of the goals is to show $\log Z_n/n$ converges almost surely as $n \to \infty$. First let us try to get some sense of the factor $1/\sqrt{n}$. Compute

$$\mathbb{E} Z_n = \sum_\sigma \mathbb{E} e^{\beta H_n(\sigma,x)} = \sum_\sigma e^{\frac{1}{2}\beta^2 \frac{1}{n} \sum_{i<j} \sigma_i^2 \sigma_j^2} = 2^n e^{\beta^2 (n-1)/4}$$

To study $\mathbb{E} \log Z_n$, an upper bound is given by $\mathbb{E} \log Z_n \le \log \mathbb{E} Z_n = n \log 2 + \frac{\beta^2(n-1)}{4}$. This indicates the right bound (and the right choice of $1/\sqrt{n}$ in the first place). (One can also get a lower bound, see Davar's note Lemma 3.5) It turns out $\mathbb{E} \log Z_n/n \to \log 2 + \beta^2/4$ (for $\beta \in (0,1)$). Question: how to compute $\mathbb{E} Z_n^2$? (replica calculation)

To show the almost sure convergence, we apply the concentration inequality. Let D_{ij} be the derivative with respect to x_{ij}. We have

$$D_{ij} \log Z_n = Z_n^{-1} D_{ij} Z_n = Z_n^{-1} \sum_\sigma \frac{\beta}{\sqrt{n}} e^{\beta H_n(\sigma,x)} \sigma_i \sigma_j$$

By the elementary fact $|\sigma_i \sigma_j| = 1$, we have $|D_{ij} \log Z_n| \le \beta/\sqrt{n}$. This implies that

$$\langle D \frac{\log Z_n}{n}, D \frac{\log Z_n}{n} \rangle = \sum_{i,j} |D_{ij} \frac{\log Z_n}{n}|^2 \le \beta^2 \frac{n(n-1)}{nn^2} \le \beta^2/n$$

so

$$\mathbb{P}[|\frac{\log Z_n}{n} - \mathbb{E} \frac{\log Z_n}{n}| \ge t] \le 2e^{-\frac{nt^2}{2\beta^2}}$$

Remark: after proving the convergence of $\log Z_n/n$, the next key question is to understand the size of fluctuations, i.e., $\mathrm{Var} \log Z_n$. The Gaussian Poincare inequality gives $\mathrm{Var} \log Z_n \le \beta^2(n-1)$ which is NOT optimal. The results in the high temperature regime $\beta \in (0,1)$ is by now well-understood: we know $\mathrm{Var} \log Z_n \sim O(1)$ and a central limit theorem is also proved $\log Z_n - \log \mathbb{E} Z_n \Rightarrow N(0, \sigma^2) - \frac{1}{2}\sigma^2$ for some $\sigma^2 > 0$ (research project). It is almost completely open in lower temperature regime. There is some soft argument proving $\mathrm{Var} \log Z_n \ll n$ (improvement of the Gaussian-Poincare inequality) (research project), but the correct size is unknown.

7. Lecture 7

The following is based on Martin Hairer's note Section 2 and 3 http://www.hairer.org/notes/Malliavin.pdf

7.1. **Malliavin derivative in the general setting.** Consider a real separable Hilbert space H with inner product denoted by $\langle , \rangle$. We say that a stochastic process $W = \{W(h), h \in H\}$ is an isonormal Gaussian process on H if W is a centered Gaussian family of random variables such that $\mathbb{E}W(h_1)W(h_2) = \langle h_1, h_2\rangle$. Examples: (i) $H = \mathbb{R}^n$, for any $x \in \mathbb{R}^n$, we have $W(x) = \sum x_j \xi_j$ where ξ_j are i.i.d standard Gaussian. In this case $W(e_j) = \xi_j$. (ii) $H = \ell_2$, for any $x \in \ell_2$, $W(x) = \sum_j x_j \xi_j$. Again $W(e_j) = \xi_j$ in this case. (iii) $H = L^2(\mathbb{R}_+)$ and for any $f \in H$, $W(f) = \int_0^\infty f(t)dB_t$, where B is a standard Brownian motion. (iv) $H = L^2(\mathbb{R}^d)$, let $\{e_j\}$ be ONB of H, and for any $h \in H$, define $W(h) = \sum_j \langle h, e_j\rangle \xi_j$. In this case W is called the white noise (recall hw1, $W(1_A)$ is independent of $W(1_B)$ iff $A \cap B$ has measure zero).

Recall that in hw1, we used a sequence of i.i.d. standard normal $\{\xi_j\}$ to construct a BM $\{B_t\}$. In a sense we only need to do calculus with respect to $\{\xi_j\}$. But in many applications, it helps to differentiate with respect to the infinitesimal increment of B_t near t, i.e., $B_{t+\Delta t} - B_t$. The difference here is whether we want to do calculus in the physical domain or the Fourier domain – typically it depends on the problem.

Consider the SDE problem $dX_t = b(X_t)dt + \sigma(X_t)dB_t$, how to compute the dependence of X_T on the noise? In this case we can not do it in the Fourier domain and it makes more sense to compute the derivative of X_T with respect to the "small increments" of B, e.g. we can start from the integral equation $X_T = X_0 + \int_0^T b(X_t)dt + \int_0^T \sigma(X_t)dB_t$ and compute derivative on both sides – similar to an example we did in class.

We define Malliavin derivative for general isonormal Gaussian process: for "nice" random variables of the form $X = F(W(h_1), \ldots, W(h_N))$ (sometimes we write it as $F(W)$), where $F : \mathbb{R}^N \to \mathbb{R}$ is a smooth function that does not grow very fast at infinity, we define

$$DF = \sum_{j=1}^N \partial_j F(W(h_1), \ldots, W(h_N))h_j,$$

i.e., DF is an $H-$valued random variable. In example (i), it just coincides with the Malliavin derivative we defined before (why?) Remark: in example (i), we can write $W(h) = \sum_{j=1}^n h_j \xi_j$ and $DW(h) = h = (h_1, \ldots, h_n)$. In example (iii), we can formally write $B_t' = \frac{dB_t}{dt}$ (which does not exist in the classical sense but exists as a random Schwarz distribution), then $W(f) = \int_0^\infty f(t)dB_t = \int_0^\infty f(t)B'(t)dt$ (still on the formal level). In this case, DF is an $L^2(\mathbb{R}_+)$ valued random variable so we can write it as $DF = \{D_tF\}_{t\geq 0}$. When $F = W(f)$, we have $D_tF = f(t)$, so one can view the Malliavin derivative at t as describing how the random variable depends on $B'(t)$. Example: $D[W(f)^2] = 2W(f)f$. $De^{W(f)} = e^{W(f)}f$. In the general case, we can view DF as the derivative in the following way: we view $W(h_j)$ as h_j integrates with respect to the white noise ξ, and if we perturb ξ by εw (with $w \in H$), then $W(h_j) \mapsto (W + \varepsilon w)(h_j) = W(h_j) + \varepsilon\langle w, h_j\rangle$, and we have

$$\frac{F(W + \varepsilon w) - F(W)}{\varepsilon} = \frac{1}{\varepsilon}\Big(F(W(h_1) + \varepsilon\langle w, h_1\rangle, \ldots, W(h_N) + \varepsilon\langle w, h_N\rangle) - F(W(h_1), \ldots, W(h_N))\Big)$$

Sending $\varepsilon \to 0$, we obtain $\langle DF, w\rangle$. In other words, we can view DF as the directional derivative. Question (i): is DF a good definition? Question (ii): Why can't we do things deterministically? For example, in the case of BM, view $B'(t)$ as a random Schwarz distribution and fix the realization, then the Malliavin derivative seems to be the Frechet derivative? (The problem is, the random variables in L^2 are only defined almost surely, so if we shift the underlying white noise, it is unclear whether the measure induced by the shifted noise is absolutely continuous with respect to the original measure, see a discussion on Martin's note page 10).

7.2. **Integration by parts.** For smooth random variables, we have $\mathbb{E}\langle DX, h\rangle = \mathbb{E}XW(h)$ for any $h \in H$. Proof: WLOG assume $X = F(W(e_1), \ldots, W(e_n))$ and $h = e_1$, where $\{e_j\}$ are orthonormal in H (why?) So we have $\langle DX, h\rangle = \sum_j \partial_j F()\langle e_j, h\rangle = \partial_1 F(W(e_1), \ldots, W(e_n))$. Since $\{W(e_j)\}_j$ are i.i.d. standard Gaussian, by an integration by parts we have $\mathbb{E}\langle DX, h\rangle = \mathbb{E}\partial_1 F(W(e_1), \ldots, W(e_n)) = \mathbb{E}F(W(e_1), \ldots, W(e_n))W(e_1) = \mathbb{E}XW(h)$ which finishes the proof. Example: how to look differently at the elementary identity $\mathbb{E}[f'(Z)] = \mathbb{E}f(Z)Z$ where $Z \in N(0,1)$. In this case $H = \mathbb{R}$ so the inner product is just multiplication and we have $W(h) = hZ$ for any $h \in \mathbb{R}$, this implies $\mathbb{E}f'(Z) = \mathbb{E}\langle Df(Z), 1\rangle = \mathbb{E}f(Z)W(1) = \mathbb{E}f(Z)Z$.

Example: assume f, g are nice deterministic functions, we have

$$\mathbb{E}f'(B_t)\int_0^t g(s)ds = \mathbb{E}[f(B_t)\int_0^t g(s)dB_s].$$

Let $X = f(B_t)$, so we have $DX = f'(B_t)1_{[0,t]}(\cdot)$, so $f'(B_t)\int_0^t g(s)ds = \langle DX, g\rangle$. By the integration by parts, we have $\mathbb{E}\langle DX, g\rangle = \mathbb{E}XW(g)$, but $W(g) = \int_0^\infty g(s)dB_s$ (how to conclude the proof?).

8. Lecture 8

The following is based on Martin Hairer's note Section 2 and 3 http://www.hairer.org/notes/Malliavin.pdf

8.1. **Definition of Malliavin derivative, review, integration by parts.** Recall the definition of isonormal Gaussian space. Note that given a sequence of i.i.d. $N(0,1)$ one can construct an isonormal Gaussian space on any separable Hilbert space (how?); and given an isonormal Gaussian space on H, one can construct a sequence of i.i.d. $N(0,1)$ (how?).

For smooth r.v. of the form $X = F(W(h_1), \ldots, W(h_N))$, we define DX as the $H-$valued random variable $DX = \sum_j \partial_j F(W(h_1), \ldots, W(h_N))h_j$. Two examples in mind: $H = \mathbb{R}^n$ and $H = L^2(\mathbb{R}_+)$.

Integration by parts formula: for smooth random variables X and $h \in H$, we have $\mathbb{E}\langle DX, h\rangle = \mathbb{E}XW(h)$. Corollary: for smooth random variables X, Y and $h \in H$, we have $\mathbb{E}[XYW(h)] = \mathbb{E}[X\langle DY, h\rangle] + \mathbb{E}[Y\langle DX, h\rangle]$.

Last time we proved the following result $\mathbb{E}f'(B_t)\int_0^t g(s)ds = \mathbb{E}f(B_t)\int_0^t g(s)dB_s$, where f is a smooth function and $g \in L^2(\mathbb{R}_+)$. The result can be proved by a direct calculation but we simply applied the aforementioned integration by parts formula. We claim this relation holds for any g that is adapted with the filtration generated by the BM such that $\int_0^t \mathbb{E}g(s)^2 ds < \infty$. Remark: we denote the filtration generated by B by $\mathcal{F}_t$, then it is easy to see that $\mathcal{F}_t$ is the sigma-algebra generated by $W(h)$ with h supported on $[0,t]$.

Proof: we go by approximation. First we assume g is an elementary process $g(s) = \sum_j Y_j 1_{[t_j, t_{j+1})}(s)$ where Y_j is $\mathcal{F}_{t_j}$ measurable and smooth random variables. We can also assume $g(s) = 0$ if $s \geq t$. In this case, we claim the identity

$$\mathbb{E}[f'(B_t)\int_0^t g(s)ds] = \mathbb{E}[f(B_t)\int_0^t g(s)dB_s].$$

Now we write $f(B_t)\int_0^t g(s)dB_s = f(B_t)\sum_j Y_j W(1_{[t_j,t_{j+1}]})$, so we have

$$\begin{aligned}\mathbb{E}f(B_t)\int_0^t g(s)dB_s &= \sum_j \mathbb{E}f(B_t)Y_jW(1_{[t_j,t_{j+1}]})\\ &= \sum_j \mathbb{E}\langle D[f(B_t)Y_j], 1_{[t_j,t_{j+1}]}\rangle\\ &= \sum_j \mathbb{E}f'(B_t)Y_j(t_{j+1}-t_j) + \sum_j \mathbb{E}[f(B_t)\langle DY_j, 1_{[t_j,t_{j+1}]}\rangle]\end{aligned}$$

The first term is $\mathbb{E}f'(B_t)\int_0^t g(s)ds$. How about the second term? We can write it as $\langle DY_j, 1_{[t_j,t_{j+1}]}\rangle = \int_{t_j}^{t_{j+1}} D_sY_jds$. Claim $D_sY_j = 0$ for $s \geq t_j$ which completes the proof. Why? Since Y_j is F_{t_j} measurable, it can be approximated by random variables of the form $F(W(h_1),\ldots,W(h_n))$, with h_j supported on $[0,t_j]$. For such random variables, we have $D_sF(W(h_1),\ldots,W(h_n)) = \sum_j \partial_j F(W(h_1),\ldots,W(h_n))h_j(s) = 0$ if $s \geq t_j$. Thus, we proved the case when g is an elementary process. The rest is to approximate g by elementary process, take the difference and apply Ito isometry.

8.2. **Orthonormal basis and chaos expansion.** Given an isonormal Gaussian process on H, we consider the $L^2(\Omega)$ space generated by such Gaussian process, in other words, the space of square integrable random variables that is measurable to the sigma algebra generated by the Gaussian process. One can ask about the ONB of this space and given any $X \in L^2(\Omega)$, how to expand it in the orthonormal basis?

Since H is a separable Hilbert space, take $\{e_i\}$ as its ONB. Define $\xi_i = W(e_i)$, claim $\{\xi_i\}$ is i.i.d. N(0,1). For any multiindex k, define $\Phi_k = \prod_j H_{k_j}(\xi_j)$ where H_k is the $k-$th order Hermite polynomial. Claim: $\{\Phi_k\}_k$ is an ONB of $L^2(\Omega)$ (we actually proved it before, where the $L^2(\Omega)$ space is generated by $\{\xi_i\}$, but clearly the isonormal Gaussian process W contains the same information as the sequence of i.i.d. $\{\xi_i\}$, so the L^2 space generated by W is the same as the L^2 space generated by $\{\xi_i\}$). Similar as before, for any $n \geq 0$, one can define $\mathcal{H}_n$ as the closed subspace of L^2 spanned by r.v. of the form Φ_k with $|k| = n$, and it is called the $n-$th order Wiener chaos. Question: given the chaos expansion, how to compute the Malliavin derivative?

8.3. **Adjoint of D.** Recall that in the case of $H = \mathbb{R}^n$, we defined the adjoint of D_j as $A_j = -D_j + Z_j$, i.e., for any $f \in D^{1,2}$ and $g \in \mathrm{Dom}(A_j)$, we have $\mathbb{E}D_jfg = \mathbb{E}f(-D_jg + Z_jg)$. Here $Z_j = W(e_j)$ where e_j is the standard basis of $\mathbb{R}^n$. In this way, the adjoint of D can be defined as follows: for any $f \in D^{1,2}$ and a random vector $g = (g_1,\ldots,g_n)$, $\mathbb{E}\langle Df, g\rangle = \sum_j \mathbb{E}D_jfg_j = \sum_j \mathbb{E}fA_jg_j = \mathbb{E}f\sum_j A_jg_j = \mathbb{E}fAg$, where we define $Ag = \sum_j A_jg_j$, and call it the divergence operator. In many cases, the operator is denoted by δ and we will change the notation from now on. So A maps (a subset of) $L^2(\Omega,\mathbb{R}^n)$ to $L^2(\Omega,\mathbb{R})$. Similarly, in the general case of H, the adjoint of D should be mapping (a subset of) $L^2(\Omega,H)$ to $L^2(\Omega,\mathbb{R})$, and we should have the identity $\mathbb{E}\langle Df, g\rangle = \mathbb{E}f\delta(g)$. At the moment, we haven't defined the domain of D yet (clearly all smooth random variables of the form $F(W(h_1),\ldots,W(h_n))$ are in the domain). Denote $\mathcal{S}$ the set of all smooth random variables. Take some $g \in L^2(\Omega,H)$, if for all $f \in \mathcal{S}$, we have

$$\mathbb{E}[\langle Df, g\rangle] \leq C\sqrt{\mathbb{E}f^2} \tag{8.1}$$

for some constant $C > 0$, the linear mapping $f \mapsto \mathbb{E}[\langle Df, g\rangle]$ is bounded on $\mathcal{S}$, and since $\mathcal{S}$ is dense in $L^2(\Omega,\mathbb{R})$, it can be extended to all $f \in L^2(\Omega,\mathbb{R})$. Then by Riesz representation theorem, there exists $X \in L^2(\Omega,\mathbb{R})$ such that $\mathbb{E}\langle Df, g\rangle = \mathbb{E}fX$ for all smooth r.v. f. We can take (8.1) as the domain of definition of δ, i.e., those $g \in L^2(\Omega, H)$ such that there exists $C > 0$ so that (8.1) holds for all smooth f. The X obtained above is then defined as $\delta(g)$.

Recall that elementary processes take the form $\sum_j Y_j 1_{[t_j,t_{j+1})}(s)$ where the summation is finite and Y_j is $\mathcal{F}_{t_j}-$measurable. Assume $Y_j \in L^2(\Omega,\mathbb{R})$, then all elementary processes takes value in $L^2(\Omega, H)$ (where $H = L^2(\mathbb{R}_+)$ here). Denote $L^2_a(\Omega, H)$ the closure of the set of elementary processes in $L^2(\Omega, H)$. Theorem: $L^2_a(\Omega, H)$ is in the domain of δ and on it, δ coincides with the Ito integral. We have already given the proof: take any elementary process $u(s) = \sum_j Y_j 1_{[t_j,t_{j+1})}(s)$, and smooth r.v. f, consider

$$\mathbb{E}\langle Df, u\rangle = \sum_j \mathbb{E}\langle Df, Y_j 1_{[t_j,t_{j+1})}(\cdot)\rangle = -\sum_j \mathbb{E}[f\langle DY_j, 1_{[t_j,t_{j+1})}(\cdot)\rangle + \mathbb{E}[f\sum_j Y_j W(1_{[t_j,t_{j+1})})]$$

By the fact that $\langle DY_j, 1_{[t_j,t_{j+1})} = 0$ and $\sum_j Y_j W(1_{[t_j,t_{j+1})}) = \int_0^\infty u(s)dB_s$, we conclude the proof for elementary process. For the general case, we complete the proof by approximation and passing to the limit on both sides of

$$\mathbb{E}\langle Df, u\rangle = \mathbb{E}f\int_0^\infty u(s)dB_s.$$

9. Lecture 9

The following is based on Martin Hairer's note Section 2 and 3 http://www.hairer.org/notes/Malliavin.pdf

9.1. Review and an example.

Isonormal Gaussian process, Malliavin derivative, two examples, integration by parts formula, divergence operator. Recall that for $X \in D^{1,2}(\Omega,\mathbb{R})$ and $Y \in \mathrm{Dom}(\delta) \subset L^2(\Omega, H)$ we have $\mathbb{E}\langle DX, Y\rangle = \mathbb{E}X\delta(Y)$. For the case of Brownian motion and for Y that is an elementary process, we have $\delta(Y) = \int_0^\infty Y_s dB_s$ where the r.h.s. is interpreted as Ito integral. Denote $L^2_a(\Omega, H)$ the closure of the set of elementary processes in $L^2(\Omega, H)$. Theorem: $L^2_a(\Omega, H)$ is in the domain of δ and on it, δ coincides with the Ito integral.

Example: Example: consider the heat equation $\partial_t u = \frac{1}{2}\Delta u$ starting from $u(0,x) = e^{B_x}1_{x\geq 0}$ where B is a standard Brownian motion. Define $h = \log u$ which solves the HJ equation $\partial_t h = \frac{1}{2}\Delta h + \frac{1}{2}|\nabla h|^2$. Compute $\mathbb{E}[h(0,x)h(t,y)]$. We write $u(t,y) = \int_0^\infty q_t(y-z)e^{B_z}dz$, so for $x > 0$,

$$\mathbb{E}[h(0,x)h(t,y))] = \mathbb{E}[B_x \log\int_0^\infty q_t(y-z)e^{B_z}dz]$$

We write $B_x = W(1_{[0,x]})$, so we have

$$\begin{aligned}
&\mathbb{E}W(1_{[0,x]})\log\int_0^\infty q_t(y-z)e^{B_z}dz\\
&= \mathbb{E}\int_0^\infty 1_{[0,x]}(r)D_r(\log\int_0^\infty q_t(y-z)e^{B_z}dz)dr\\
&= \mathbb{E}\int_0^x \frac{\int_0^\infty q_t(y-z)e^{B_z}1_{[0,z]}(r)dz}{\int_0^\infty q_t(y-z)e^{B_z}dz}dr\\
&= \mathbb{E}\frac{\int_0^\infty q_t(y-z)e^{B_z}\min(x,z)dz}{\int_0^\infty q_t(y-z)e^{B_z}dz}
\end{aligned}$$

the last expression can be interpreted as $\hat{\mathbb{E}}\min(Y_t, x)$ where $\{Y_s\}_{s\in[0,t]}$ is a BM starting from $Y_0 = y$ with the endpoint weighted by $e^{B_{Y_t}} 1_{Y_t>0}$, and $\hat{\mathbb{E}}$ is the expectation on this weighted process. Remark: taking derivative with respect to x actually gives us the density.

9.2. **Taking derivative of Ito integral.** Motivation: consider the SDE $dX_t = b(X_t)dt + \sigma(X_t)dB_t$ so $X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s$. In most cases we do not have explicit solution. How to compute $D_r X_t$? Clearly we have $D_r X_t = 0$ for $r \geq t$. Assume that X_s is differentiable for all s, it reduces to compute the Malliavin derivative of the Ito integral. (compare it to the example of random difference equation with random coefficient). Let us try to do it first without justifying anything: we have

$$D_r X_t = D_r \int_0^t b(X_s)ds + D_r \int_0^t \sigma(X_s)dB_s$$

The first term is easy $D_r \int_0^t b(X_s)ds = \int_0^t b'(X_s)D_r X_s ds$. How about the second term? Guess, for $r \leq t$, we have $D_r \int_0^t \sigma(X_s)dB_s = \sigma(X_r) + \int_0^t D_r\sigma(X_s)dB_s = \sigma(X_r) + \int_0^t \sigma'(X_s)D_r X_s dB_s$. Recall that $D_r X_s = 0$ for $r \geq s$, so the above equation becomes

$$D_r \int_0^t \sigma(X_s)dB_s = \sigma(X_r) + \int_r^t \sigma'(X_s)D_r X_s dB_s$$

To summarize,

$$D_r X_t = \int_r^t b'(X_s)D_r X_s ds + \sigma(X_r) + \int_r^t \sigma'(X_s)D_r X_s dB_s$$

Fix r, as a process indexed by $t \geq r$, we have $Y_t := D_r X_t$ satisfies

$$Y_t = \sigma(X_r) + \int_r^t b'(X_s)Y_s ds + \int_r^t \sigma'(X_s)Y_s dB_s$$

In other words $\{Y_t\}_{t\geq r}$ solves the SDE

$$\begin{aligned} dY_t &= b'(X_t)Y_t dt + \sigma'(X_t)Y_t dB_t, \qquad t > r \\ Y_r &= \sigma(X_r) \end{aligned} \tag{9.1}$$

Here is rigorous statement for the Malliavin derivative of an Ito integral:

Proposition. Let $u \in L^2_a(\Omega, H)$ such that $u(t) \in D^{1,2}$ for almost every $t \in \mathbb{R}_+$ and

$$\int_0^\infty \mathbb{E}\langle Du(t), Du(t)\rangle dt = \int_0^\infty \left(\int_0^\infty \mathbb{E}|D_r u(t)|^2 dr \right) dt < \infty. \tag{9.2}$$

Then we have $\int_0^\infty u(t)dB_t \in D^{1,2}$ and

$$D_r \int_0^\infty u(t)dB_t = u(r) + \int_0^\infty D_r u(t)dB_t = u(r) + \int_r^\infty D_r u(t)dB_t$$

Remark: the above identity is in $L^2(\Omega, H)$.

Check that $\int_0^\infty D_. u(t)dB_t \in L^2(\Omega, H)$: note that (23.5) indicates that for almost every r, $\int_0^\infty \mathbb{E}|D_r u(t)|^2 dt < \infty$, so the Ito integral $\int_0^\infty D_r u(t)dB_t$ is well-defined (for such r), and since we have $\int_0^\infty \mathbb{E}|\int_0^\infty D_r u(t)dB_t|^2 dr < \infty$, we know that $\int_0^\infty D_. u(t)dB_t \in L^2(\Omega, H)$.

Proof: we start from elementary process. Assume $u(t) = \sum_i Y_i 1_{[t_i, t_{i+1})}(t)$ and Y_i are smooth random variables (i.e., of the form $F(W(h_1), \ldots, W(h_N))$ where F is smooth function that does not grow very fast at infinity). Note that this type of u clearly satisfies the assumption on the proposition. In this case, we have $\int_0^\infty u(t)dB_t = \sum_i Y_i W(1_{[t_i,t_{i+1})}(\cdot))$, so by product rule, we have

$$D_r \int_0^\infty u(t)dB_t = \sum_i D_r Y_i W(1_{[t_i,t_{i+1})}(\cdot)) + \sum_i Y_i 1_{[t_i,t_{i+1})}(r)$$

which completes the proof for the case of elementary process. Now we use elementary processes u_n to approximate u, we have for all $n \geq 1$,

$$D_r \int_0^\infty u_n(t)dB_t = u_n(r) + \int_0^\infty D_r u_n(t)dB_t$$

and it remains to pass to the limit as $n \to \infty$. Fact (I): we can find u_n to approximate u in the following sense (i) $u_n \to u$ in $L^2(\Omega, H)$, i.e., $\int_0^\infty \mathbb{E}|u_n(t) - u(t)|^2 dt \to 0$ (ii) $Du_n \to Du$ in $L^2(\mathbb{R}_+, L^2(\Omega, H))$, i.e.,

$$\int_0^\infty \mathbb{E}\langle Du_n(t) - Du(t), Du_n(t) - Du(t)\rangle dt = \int_0^\infty \left(\int_0^\infty \mathbb{E}|D_r u_n(t) - D_r u(t)|^2 dr \right) dt \to 0$$

Using (i) and (ii), we have $u_n(\cdot) + \int_0^\infty D.u_n(t)dB_t \to u(\cdot) + \int_0^\infty D.u(t)dB_t$ in $L^2(\Omega, H)$. Fact (II) (Proposition 3.3 in Hairer's note): D is closable, i.e., if $X_n \in D^{1,2}$ and $X_n \to X$ in $L^2(\Omega)$ and DX_n converges in $L^2(\Omega, H)$, then we can define $DX = \lim_n DX_n$.

Using these two facts, we complete the proof by sending $n \to \infty$. (Question: how to prove (19.2)?)

10. Lecture 10

10.1. **Review, differentiation of Ito integral.** For the Ito integral $\int_0^\infty u_t dB_t$, where u is a nice process ($\int_0^\infty \mathbb{E}u_t^2 dt < \infty$ and $\int_0^\infty \mathbb{E}\|Du_t\|^2_{L^2(\mathbb{R}_+)} dt < \infty$), then

$$D_r \int_0^\infty u_t dB_t = u_r + \int_0^\infty D_r u_t dB_t = u_r + \int_r^\infty D_r u_t dB_t$$

Intuition here.

10.2. **More on the divergence operator.** For elementary process of the form $X_t = \sum_i Y_i 1_{[t_i,t_{i+1})}(t)$ where Y_i is $\mathcal{F}_{t_i}$ measurable, then

$$\delta(X) = \sum_i Y_i (B_{t_{i+1}} - B_{t_i}) = \int_0^\infty X_t dB_t,$$

i.e., the divergence operator coincides with Ito integral in this case. What if X is not adapted? For example, we can assume that Y_i depends on the Brownian motion beyond t_i. Claim: in the general case we have

$$\delta(X) = \sum_i Y_i (B_{t_{i+1}} - B_{t_i}) - \sum_i \langle DY_i, 1_{[t_i,t_{i+1})}\rangle \tag{10.1}$$

In other words, we have the extra term $\sum_i \langle DY_i, 1_{[t_i,t_{i+1})}\rangle = \sum_i \int_{t_i}^{t_{i+1}} D_r Y_i dr$. To show (23.5), by linearity, we consider the case $X = Y1_{[t,s)}$, then for any smooth random variable Z, we have $D(YZ) = YDZ + ZDY$ so $\mathbb{E}YZW(1_{[t,s)}) = \mathbb{E}\langle D(YZ), 1_{[t,s)}\rangle$, which implies

$$\mathbb{E}YZ(B_s - B_t) = \mathbb{E}Z \int_t^s D_r Y dr + \mathbb{E}Y \int_t^s D_r Z dr$$

we write it as

$$\mathbb{E}Z[Y(B_s - B_t) - \int_t^s D_r Y dr] = \mathbb{E}\langle DZ, X\rangle$$

this implies that $\delta(X) = Y(B_s - B_t) - \int_t^s D_r Y dr$, so (23.5) is proved. Remark: $\mathbb{E}\delta(X) = 0$ (how to see it directly?) By the expression of the r.h.s. of (23.5), which also looks like "integral", $\delta(X)$ is called the Skorohod integral of X, which coincides with Ito integral when X is adapted (not anticipative).

10.3. **Clark-Ocone formula.** Martingale representation theorem: for square integral random variable $X \in L^2(\Omega)$, where the probability space is generated by a standard BM B, we can write

$$X - \mathbb{E}X = \int_0^\infty \phi_t dB_t$$

for some unique adapted process ϕ satisfying $\int_0^\infty \mathbb{E}\phi_t^2 dt < \infty$. Define $M_t = \mathbb{E}[X|\mathcal{F}_t]$ which is a square integrable martingale, we have $M_t = M_0 + \int_0^t \phi_s dB_s$. In other words, the martingale is written as an Ito integral. Question: given X, how to compute such ϕ?

Theorem (Clark-Ocone formula): if $X \in D^{1,2}$, we have

$$X = \mathbb{E}X + \int_0^\infty \mathbb{E}[D_r X|\mathcal{F}_r] dB_r$$

In other words, we can compute ϕ through computing the Malliavin derivative of X.

Proof: Assume $X = \mathbb{E}X + \int_0^\infty \phi_r dB_r$ for some adapted process ϕ such that $\int_0^\infty \mathbb{E}\phi_r^2 dr < \infty$. We need to show that $\phi_r = \mathbb{E}[D_r X|\mathcal{F}_r]$. To simplify the notation, denote $\psi_r = \mathbb{E}[D_r X|\mathcal{F}_r] \in L_a^2$. Take any process $u \in L_a^2$ (approximated by elementary process), we have

$$\begin{aligned}
\mathbb{E}[X\int_0^\infty u_r dB_r] &= \mathbb{E}[X\delta(u)] = \mathbb{E}\langle DX, u\rangle \\
&= \mathbb{E}\int_0^\infty D_r X u_r dr = \mathbb{E}\int_0^\infty \mathbb{E}[D_r X|\mathcal{F}_r] u_r dr = \mathbb{E}\int_0^\infty \psi_r u_r dr
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathbb{E}X\int_0^\infty u_r dB_r = \mathbb{E}(X - \mathbb{E}X)\int_0^\infty u_r dB_r = \mathbb{E}\int_0^\infty \phi_r dB_r \int_0^\infty u_r dB_r \\
&= \mathbb{E}\int_0^\infty \phi_r u_r dr
\end{aligned}$$

To summarize, we have $\int_0^\infty \mathbb{E}u_r\phi_r dr = \int_0^\infty \mathbb{E}u_r\psi_r dr$. Since u is arbitrary, we have $\phi = \psi$, which completes the proof.

10.4. **Application of Clark-Ocone formula.** Example 1: $X = B_1^3$, how to write it as an Ito integral? We have $D_r X = 3B_1^2 1_{[0,1]}(r)$, so for $r \le 1$, we have

$$\mathbb{E}[D_r X|\mathcal{F}_r] = \mathbb{E}[3B_1^2|\mathcal{F}_r] = 3\mathbb{E}[(B_r + B_1 - B_r)^2|\mathcal{F}_r] = 3B_r^2 + 3(1-r)$$

thus we have

$$B_1^3 = 3\int_0^1 (B_r^2 + 1 - r) dB_r$$

Example 2: local time of Brownian motion at the origin $L_t(0)$. Recall that the local time Brownian motion $L_t(x)$ is related to BM through $\int_0^t f(B_s)ds = \int_{\mathbb{R}} f(x)L_t(x)dx$. Question: how to write $L_t(0)$ as a stochastic integral? We know that

$$L_t(0) = \lim_{\varepsilon\to 0}\int_{\mathbb{R}} q_\varepsilon(x)L_t(x)dx = \lim_{\varepsilon\to 0}\int_0^t q_\varepsilon(B_s)ds$$

Here q_ε is an approximation of Dirac function (On the formal level we write $L_t(x) = \int_0^t \delta(B_s - x)ds$). For each ε, we have $X_\varepsilon = \int_0^t q_\varepsilon(B_s)ds$ and $X_\varepsilon - \mathbb{E}X_\varepsilon = \int_0^t \mathbb{E}[D_r X_\varepsilon|\mathcal{F}_r]dB_r$ We have

$$D_r X_\varepsilon = \int_0^t q_\varepsilon'(B_s)1_{[0,s]}(r)ds = \int_r^t q_\varepsilon'(B_s)ds$$

so the conditional expectation

$$\mathbb{E}[D_r X_\varepsilon|\mathcal{F}_r] = \mathbb{E}[\int_r^t q_\varepsilon'(B_s)ds|\mathcal{F}_r] = \int_r^t \mathbb{E}[q_\varepsilon'(B_s)|\mathcal{F}_r]ds = \int_r^t q_{\varepsilon+s-r}'(B_r)ds$$

This implies $X_\varepsilon - \mathbb{E}X_\varepsilon = \int_0^t \left(\int_r^t q_{\varepsilon+s-r}'(B_r)ds\right)dB_r$. Let $\varepsilon \to 0$, we have

$$L_t(0) - \mathbb{E}L_t(0) = \int_0^t \left(\int_r^t q_{s-r}'(B_r)ds\right)dB_r.$$

There are more complicated applications in SDE and SPDE.

11. Lecture 11

The following is based on Chapter 3 of the book *Normal Approximations with Malliavin Calculus : From Stein's Method to Universality* which can be found online through the UMD library.

Consider the following setup: given $Y = \{Z_i\}$, a sequence of i.i.d. $N(0,1)$ or $Y = \{B_t\}_{t\geq 0}$ which is a standard BM, we are interested in the distribution of $X = g(Y)$ where F is a complicated, smooth functional. If we expect the distribution is close to Gaussian, how do we show it? How do we compare the distribution of $X - \mathbb{E}X$ with $N(0,\sigma^2)$ where $\sigma^2 = \mathrm{Var}X$? Stein's method provides an efficient way of doing it, combined with some integration by parts in the Gaussian space.

11.1. Stein's equation.

Lemma: a real valued r.v. has $N(0,1)$ distribution iff for any C^1 function f such that $f \in L^1(p_1(x)dx)$ we have $\mathbb{E}f'(X) = \mathbb{E}Xf(X)$. Here $p_1(x) = (2\pi)^{-1/2}e^{-x^2/2}$. Proof: $\Rightarrow$ We have $\mathbb{E}f'(X) = \int f'(x)p_1(x)dx = \int f(x)xp_1(x)dx = \mathbb{E}Xf(X)$. $\Leftarrow$ Take $f(x) = x^n$, we have $n\mathbb{E}X^{n-1} = \mathbb{E}X^{n+1}$. We start from $\mathbb{E}X^2 = 1$ and $\mathbb{E}X = 0$, then by recursion, we have the moments of X, which is the moments of $N(0,1)$. This completes the proof. (Remark: the proof of integration by parts is not completely trivial, and the proof can be found in the reference Lemma 1.1.1)

Assume $\mathbb{E}X = 0$ and $\mathrm{Var}X = 1$, the above lemma shows that to determine if X has $N(0,1)$ distribution, we only need to compare $\mathbb{E}f'(X)$ with $\mathbb{E}Xf(X)$ for smooth functions f. It turns out the distance between the law of X and $N(0,1)$ can be controlled by $\mathbb{E}f'(X) - \mathbb{E}Xf(X)$. This is done through solving Stein's equation.

Definition: $Z \sim N(0,1)$ and $h : \mathbb{R} \to \mathbb{R}$ s.t. $\mathbb{E}|h(N)| < \infty$, the Stein's equation associated with h is the ODE

$$f'(x) - xf(x) = h(x) - \mathbb{E}h(Z).$$

Lemma: Every solution to the above equation has the form

$$f(x) = ce^{x^2/2} + e^{x^2/2}\int_{-\infty}^{x}(h(y) - \mathbb{E}h(Z))e^{-y^2/2}dy$$

where $c \in \mathbb{R}$. In particular, for $c = 0$, $f_h(x) = e^{x^2/2}\int_{-\infty}^{x}(h(y) - \mathbb{E}h(Z))e^{-y^2/2}dy$ is the unique solution s.t. $e^{-x^2/2}f(x) \to 0$ as $|x| \to \infty$.

Proof: the l.h.s. of the equation can be rewritten as $f'(x) - xf(x) = e^{x^2/2}\frac{d}{dx}(e^{-x^2/2}f(x))$, which implies

$$\frac{d}{dx}(e^{-x^2/2}f(x)) = e^{-x^2/2}(h(x) - \mathbb{E}h(Z))$$

so we have $e^{-x^2/2}f(x) = c + \int_{-\infty}^{x} e^{-y^2/2}(h(y) - \mathbb{E}h(Z))dy$ for some $c \in \mathbb{R}$. The proof is complete.

Lemma: if h is bounded, then f_h satisfies $\|f_h\|_\infty \le C$ and $\|f_h'\|_\infty \le C$ for some constant $C > 0$ only depending on $\|h\|_\infty$.

Proof: since $f_h(x) = e^{x^2/2}\int_{-\infty}^{x}(h(y) - \mathbb{E}h(Z))e^{-y^2/2}dy$. For $x \le 0$, we have

$$|f_h(x)| \le Ce^{x^2/2}\int_{-\infty}^{x} e^{-y^2/2}dy$$

which is bounded. For $x > 0$, we have $f_h(x) = e^{x^2/2}\int_{x}^{\infty}(h(y) - \mathbb{E}h(Z))e^{-y^2/2}dy$, so the proof is complete. (note that we can use the standard tail bound $\mathbb{P}(Z > x) \le e^{-x^2/2}/(\sqrt{2\pi}x)$ for $x > 0$)

11.2. **Comparing distributions.** To compare the distributions of X and Z where $Z \sim N(0,1)$, there are different ways. From now on, we assume X is normalized, i.e., $\mathbb{E}X = 0$ and $\mathrm{Var}X = 1$. Take test function h we can compare $\mathbb{E}h(X) - \mathbb{E}h(Z)$. If h is the function of the form $1_{(-\infty,x]}$, then taking sup with respect to x gives the Kolmogorov distance $\sup_x |\mathbb{P}[X \le x] - \mathbb{P}[Z \le x]|$. If h is Lipchitz function, this gives the Wasserstein distance.

By Stein's equation we have

$$\mathbb{E}h(X) - \mathbb{E}h(Z) = \mathbb{E}[f_h'(X) - Xf_h(X)]$$

thus, it reduces to study the r.h.s. (which does not concern Z!) We also have

$$\sup_{h\in H}|\mathbb{E}h(X) - \mathbb{E}h(Z)| = \sup_{h\in H}\mathbb{E}[f_h'(X) - Xf_h(X)]$$

where H can be any class of functions. By picking different H, we obtain different distances between probability laws.

11.3. **Stein's method to prove the classical CLT.** Given a sequence of i.i.d. r.v. $\{\xi_i\}$ with $\mathbb{E}\xi_i = 0$ and $\mathrm{Var}\xi_i = 1$, how to show $S_n/\sqrt{n} \Rightarrow N(0,1)$? (without characteristic function) By the previous discussion, if we are interested in the Kolmogorov distance, we only need to show that for absolutely continuous function f with $\|f\|_\infty, \|f'\|_\infty$ bounded, we have $\mathbb{E}\frac{S_n}{\sqrt{n}}f(\frac{S_n}{\sqrt{n}}) - \mathbb{E}f'(\frac{S_n}{\sqrt{n}}) \to 0$ as $n \to \infty$.

Here we consider the Wasserstein distance, for which we can actually assume that f'' is also bounded. We also assume that $\mathbb{E}|X_1|^3 < \infty$. The proof goes as follows.

$$\mathbb{E}\frac{S_n}{\sqrt{n}}f(\frac{S_n}{\sqrt{n}}) = \sqrt{n}\mathbb{E}\xi_1 f(S_n/\sqrt{n}) = \sqrt{n}\mathbb{E}\xi_1(f(S_n/\sqrt{n}) - f(\tilde{S}_n/\sqrt{n}))$$

where $\tilde{S}_n = \xi_2 + ... + \xi_n$. The above term equals to

$$\sqrt{n}\mathbb{E}\xi_1(f(S_n/\sqrt{n}) - f(\tilde{S}_n/\sqrt{n})) = \sqrt{n}\mathbb{E}\xi_1 f'(\tilde{S}_n/\sqrt{n})\frac{\xi_1}{\sqrt{n}} + err_1 = \mathbb{E}f'(\tilde{S}_n/\sqrt{n}) + err_1$$

with $err_1 \lesssim 1/\sqrt{n}$. On the other hand, we have $\mathbb{E}f'(S_n/\sqrt{n}) = \mathbb{E}f'(\tilde{S}_n/\sqrt{n}) + err_2$, with $err_2 \lesssim 1/\sqrt{n}$. Therefore, we have

$$|\mathbb{E}f'(\frac{S_n}{\sqrt{n}}) - \mathbb{E}\frac{S_n}{\sqrt{n}}f(\frac{S_n}{\sqrt{n}})| \lesssim \frac{1}{\sqrt{n}}$$

It turns out the above estimate can be turned into the estimate on the Wasserstein distance between $S_n/\sqrt{n}$ and $N(0,1)$.

12. Lecture 12

The following is based on Chapter 3 of the book *Normal Approximations with Malliavin Calculus : From Stein's Method to Universality* which can be found online through the UMD library.

12.1. **Review of Stein's method.** (i) The Gaussian distribution $N(0,1)$ can be characterized as follows: a real valued r.v. has $N(0,1)$ distribution iff for any C^1 function f such that $f \in L^1(p_1(x)dx)$ we have $\mathbb{E}f'(X) = \mathbb{E}Xf(X)$.

(ii) Stein's equation: Definition: $Z \sim N(0,1)$ and $h : \mathbb{R} \to \mathbb{R}$ s.t. $\mathbb{E}|h(Z)| < \infty$, the Stein's equation associated with h is the ODE $f'(x) - xf(x) = h(x) - \mathbb{E}h(Z)$. The ODE has an explicit solution $f_h(x) = e^{x^2/2}\int_{-\infty}^x (h(y) - \mathbb{E}h(Z))e^{-y^2/2}dy$, and if h is bounded, then f_h satisfies $\|f_h\|_\infty \le C$ and $\|f_h'\|_\infty \le C$ for some constant $C > 0$ only depending on $\|h\|_\infty$.

(iii) The setup for the quantitative central limit theorem. Let $\mathscr{H}$ be a class of functions, define

$$d_{\mathscr{H}}(X,Z) = \sup_{h\in\mathscr{H}}\{\mathbb{E}h(X) - \mathbb{E}h(Z)\}$$

then we have

$$d_{\mathscr{H}}(X,Z) = \sup_{h\in\mathscr{H}}\{\mathbb{E}f_h'(X) - \mathbb{E}Xf_h(X)\}$$

12.2. **Total variation distance, Kolmogorov distance and Wasserstein distance.** Theorem: Let $h : \mathbb{R} \to [0,1]$, then the solution to the Stein's equation satisfies $\|f_h\|_\infty \le \sqrt{\pi/2}$ and $\|f_h'\|_\infty \le 2$. In particular,

$$d_{TV}(X,Z) \le \sup_{\|f\|_\infty \le \sqrt{\pi/2}, \|f'\|_\infty \le 2}\{\mathbb{E}f'(X) - \mathbb{E}Xf(X)\}$$

where d_{TV} is the total variation distance, i.e., $d_{TV}(X,Z) = \sup_{A\subset\mathbb{R}}|\mathbb{P}(X \in A) - \mathbb{P}(Z \in A)|$.

For Kolmogorov distance, i.e., when h is of the form $h(y) = 1_{(-\infty,x]}(y)$ for some $x \in \mathbb{R}$, we clearly have $d_K \le d_{TV}$. In this case, one can check that $\|f_h\|_\infty \le \sqrt{2\pi}/4, \|f_h'\|_\infty \le 1$.

For Wasserstein distance, i.e., when h is a Lipchitz function with constant K, then $f_h \in C^1$ and $\|f_h'\|_\infty \le \sqrt{2/\pi}K$. Theorem:

$$d_W(X,Z) \le \sup_{f\in C^1, \|f'\|_\infty \le \sqrt{2/\pi}}\{\mathbb{E}f'(X) - \mathbb{E}Xf(X)\}. \tag{12.1}$$

12.3. **Second order Poincaré inequality.** Suppose we have an underlying i.i.d. standard Gaussian $\{\xi_i\}$, and $X = F(\{\xi_i\})$ is the r.v. we are interested in. The Gaussian-Poincare inequality tells us that if $X \in D^{1,2}$,

$$\mathrm{Var}X \leq \mathbb{E}\|DX\|^2 = \sum_i \mathbb{E}|\partial_i F|^2$$

In other words, the variance of X can be controlled through estimating the first derivative. There is a second order Poincare inequality which controls the distance between the law of X and that of $N(0,1)$, in terms of the second derivative. Assuming first that X is centered and rescaled so that $\mathbb{E}X = 0$ and $\mathrm{Var}X = 1$.

First assume F is smooth. Take any $\phi \in C^1$, we want to estimate $\mathbb{E}\phi'(X) - \mathbb{E}X\phi(X)$. Since X is centered, we have

$$\mathbb{E}X\phi(X) = \mathrm{Cov}[X, \phi(X)] = \mathrm{Cov}[F(\{\xi_i\}, \phi(F(\{\xi_i\}))].$$

Recall that we have the following covariance identity: for any $Y_1, Y_2 \in D^{1,2}$,

$$\mathrm{Cov}[Y_1, Y_2] = \int_0^\infty e^{-t} \sum_i \mathbb{E}[\partial_i Y_1 P_t \partial_i Y_2] = \sum_i \mathbb{E}[\partial_i Y_1 (1-L)^{-1} \partial_i Y_2]$$

Take $Y_1 = \phi(F(\{\xi_i\}))$ and $Y_2 = F(\{\xi_i\})$, we have

$$\mathrm{Cov}[Y_1, Y_2] = \sum_i \mathbb{E}[\phi'(F(\{\xi_i\}))\partial_i F(1-L)^{-1}\partial_i F] = \mathbb{E}[\phi'(X)\sum_i \partial_i F(1-L)^{-1}\partial_i F]$$

Denote

$$G = \sum_i \partial_i F(1-L)^{-1}\partial_i F,$$

we have

$$\mathrm{Cov}[X, \phi(X)] = \mathrm{Cov}[F(\{\xi_i\}, \phi(F(\{\xi_i\}))] = \mathbb{E}[\phi'(X)G]$$

Choose $\phi(x) = x$, we obtain $\mathbb{E}G = 1$. Then we write

$$\mathbb{E}\phi'(X) - \mathbb{E}X\phi(X) = \mathbb{E}\phi'(X)(1-G)$$

which implies

$$|\mathbb{E}\phi'(X) - \mathbb{E}X\phi(X)| \leq \|\phi'\|_\infty \mathbb{E}|G-1| \leq \|\phi'\|_\infty \sqrt{\mathrm{Var}G}$$

and (by Stein's method which led to (12.1))

$$d_W(X, Z) \leq \sqrt{2/\pi}\sqrt{\mathrm{Var}G}$$

It remains to estimate $\mathrm{Var}\, G$. We apply Gaussian Poincare inequality

$$\mathrm{Var}G \leq \sum_j \mathbb{E}|\partial_j G|^2$$

To compute the Malliavin derivative of G, we have

$$\partial_j G = \partial_j \sum_i \partial_i F(1-L)^{-1}\partial_i F = \sum_i \partial_{ij}F(1-L)^{-1}\partial_i F + \sum_i \partial_i F \partial_j (1-L)^{-1}\partial_i F$$

Claim

$$\partial_j (1-L)^{-1} = (2-L)^{-1}\partial_j \tag{12.2}$$

Then

$$\partial_j G = \sum_i \partial_{ij}F(1-L)^{-1}\partial_i F + \sum_i \partial_i F(2-L)^{-1}\partial_{ij}F$$

By triangle inequality, we have

$$\begin{aligned}\|\partial_j G\|_2 &\leq \sum_i \|\partial_{ij}F(1-L)^{-1}\partial_i F\|_2 + \sum_i \|\partial_i F(2-L)^{-1}\partial_{ij}F\|_2 \\ &\leq \sum_i \|\partial_{ij}F\|_4 \|(1-L)^{-1}\partial_i F\|_4 + \sum_i \|\partial_i F\|_4 \|(2-L)^{-1}\partial_{ij}F\|_4\end{aligned}$$

and in the second $\leq$, we applied Cauchy-Schwarz. We will show that P_t is a contraction from L^p to L^p (for any $p \geq 2$), and this implies that the above term is bounded by

$$\|\partial_j G\|_2 \leq 2\sum_i \|\partial_{ij}F\|_4 \|\partial_i F\|_4$$

so

$$\begin{aligned}d_W(X,Z) \leq& \sqrt{2/\pi}\sqrt{\mathrm{Var}G} \leq \sqrt{2/\pi}\sqrt{\sum_j \|\partial_j G\|_2^2} \\ \leq& \sqrt{4/\pi}\sqrt{\sum_j \big(\sum_i \|\partial_{ij}F\|_4 \|\partial_i F\|_4\big)^2}\end{aligned}$$

This is the so-called Second order Poincare inequality.

Still need to show (12.2). We write

$$\partial_j(1-L)^{-1} = \partial_j \int_0^\infty e^{-t}P_t dt = \int_0^\infty e^{-t}\partial_j P_t dt$$

For any random variable X, we write $X = \sum_k \langle X, H_k\rangle H_k/k!$ as the Wiener chaos expansion, so $P_t X = \sum_k \langle X, H_k\rangle e^{-|k|t} H_k/k!$ and

$$\partial_j P_t X = \sum_k \langle X, H_k\rangle e^{-|k|t} k_j H_{k-e_j}/k!$$

On the other hand,

$$\partial_j X = \sum_k \langle X, H_k\rangle k_j H_{k-e_j}/k!$$

which implies

$$P_t \partial_j X = \sum_k \langle X, H_k\rangle k_j e^{-(|k|-1)t} H_{k-e_j}/k!$$

Thus, $P_t\partial_j = e^t \partial_j P_t$, which implies

$$\int_0^\infty e^{-t}\partial_j P_t dt = \int_0^\infty e^{-2t} P_t \partial_j dt = (2-L)^{-1}\partial_j$$

which completes the proof of (12.2).

12.4. **Applications of the second order Poincare.** Two examples:

(i) Standard central limit theorem. $X_j = f(\xi_j)$, f nice function. Assume $\mathbb{E}X_j = 0$ and $\mathbb{E}X_j^2 = 1$. How to show $\sum_{j=1}^n X_j/\sqrt{n} \Rightarrow N(0,1)$?

Let $F_n = \frac{1}{\sqrt{n}}\sum_{j=1}^n X_j$, so we have $\partial_j F_n = \frac{1}{\sqrt{n}} f'(\xi_j)$, and $\partial_{ij}F_n = \frac{1}{\sqrt{n}} f''(\xi_j) 1_{i=j}$, so we bound

$$d_W(F_n, Z) \leq C\sqrt{\sum_j \big(\sum_i \frac{1}{n} 1_{i=j}\big)^2} = C\sqrt{\frac{1}{n}}$$

(ii) $F_n = \frac{\xi_1\xi_2+\xi_2\xi_3+\ldots+\xi_{n-1}\xi_n}{\sqrt{n-1}}$. How to show $F_n \Rightarrow N(0,1)$. We have

$$\partial_i F_n = \frac{1}{\sqrt{n-1}}(\xi_{i+1}+\xi_{i-1})$$

and we can compute $\partial_{ij}F_n$. In the end, we have

$$d_W(F_n, Z) \le C\frac{1}{\sqrt{n-1}}$$

13. Lecture 13

13.1. **Review of second order Poincare inequality.** Let $\{\xi_i\}$ be a sequence of i.i.d. $N(0,1)$, and $X = F(\{\xi_i\})$ satisfies $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$, then

$$d_W(X,Z) \le C\sqrt{\sum_j \big(\sum_i \|\partial_{ij}F\|_4\|\partial_i F\|_4}$$

where $Z \sim N(0,1)$ and d_W is the Wasserstein distance. The key point here is that one can use it to prove CLT just by carrying out some calculations (provided that $\partial_i F$ and $\partial_{ij}F$ can be computed more or less explicitly.

13.2. **Another application of Stein's method.** Recall that for random variables that are standardized $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$, to apply Stein's method, it reduces to study $\mathbb{E}\phi'(X) - \mathbb{E}X\phi(X)$ for a large class of ϕ. In the proof of the second order Poincare, we write $\mathbb{E}X\phi(X) = \mathrm{Cov}[X, \phi(X)]$ and apply the covariance relation. Sometimes when X is of special structure, it can be done more easily and we do not need to compute the second derivative.

Consider the case when the probability space is generated by a standard Brownian motion, and $X = \delta(v)$ for some $v \in L_a^2$, i.e., $X = \int_0^\infty v_s dB_s$ and v is an adapted process so that $\mathbb{E}\int_0^\infty v_s^2 ds = 1 < \infty$. We have the following result:

Proposition: suppose $X \in D^{1,2}$, $X = \delta(v)$ with $\mathbb{E}X^2 = 1$, then

$$d_W(X,Z) \le C\sqrt{\mathrm{Var}\langle v, DX\rangle}$$

Proof: it reduces to study $\mathbb{E}\phi'(X) - \mathbb{E}X\phi(X)$ for $\phi \in C^1$ with bounded derivative. We have

$$\mathbb{E}X\phi(X) = \mathbb{E}\delta(v)\phi(X) = \mathbb{E}\langle v, D\phi(X)\rangle = \mathbb{E}\langle v, \phi'(X)DX\rangle = \mathbb{E}\phi'(X)\langle v, DX\rangle$$

(recall that $\langle v, DX\rangle = \int_0^\infty v_s D_s X ds$). Therefore, we have

$$|\mathbb{E}\phi'(X) - \mathbb{E}X\phi(X)| = |\mathbb{E}\phi'(X)[1 - \langle v, DX\rangle]| \le C\mathbb{E}|1 - \langle v, DX\rangle|$$

On the other hand, we know that $\mathbb{E}X^2 = 1 = \mathbb{E}\langle v, DX\rangle$, so we can estimate the above term as

$$|\mathbb{E}\phi'(X) - \mathbb{E}X\phi(X)| \le C\sqrt{\mathrm{Var}\langle v, DX\rangle} \tag{13.1}$$

which completes the proof.

Remark 13.1. Sometimes we just directly estimate $\mathbb{E}|1 - \langle v, DX\rangle|$ to avoid the loss of applying Hölder in (13.1), or we do not assume $\mathbb{E}X^2 = 1$ (so we do not know if $\mathbb{E}\langle v, DX\rangle = 1$ and hence can't estimate it in terms of the variance).

Further extension: how to estimate the variance of the random variable $\langle v, DX\rangle$? Since $X = \int_0^\infty v_s dB_s$, we have

$$D_r X = v_r + \int_r^\infty D_r v_s dB_s$$

which implies that

$$\langle v, DX\rangle = \int_0^\infty v_r D_r X dr = \int_0^\infty v_r^2 dr + \int_0^\infty v_r (\int_r^\infty D_r v_s dB_s) dr$$

We know that $\mathbb{E}X^2 = \mathbb{E}\int_0^\infty v_s^2 ds = 1$, so we can write

$$\begin{aligned}\langle v, DX\rangle - 1 &= \left(\int_0^\infty v_r^2 dr - 1\right) + \int_0^\infty v_r (\int_r^\infty D_r v_s dB_s) dr \\ &= \left(\int_0^\infty v_r^2 dr - 1\right) + \int_0^\infty \left(\int_0^s v_r D_r v_s dr\right) dB_s\end{aligned}$$

To estimate the variance, we write

$$\sqrt{\mathrm{Var}\langle v, DX\rangle} = \|\langle v, DX\rangle - 1\|_2$$

By an application of triangle inequality, we have

$$\sqrt{\mathrm{Var}\langle v, DX\rangle} \le \left\|\int_0^\infty v_r^2 dr - 1\right\|_2 + \sqrt{\mathbb{E}\int_0^\infty \left(\int_0^s v_r D_r v_s dr\right)^2 ds}$$

To summarize, we have

$$d_W(X, Z) \le C\left(\left\|\int_0^\infty v_r^2 dr - 1\right\|_2 + \sqrt{\mathbb{E}\int_0^\infty \left(\int_0^s v_r D_r v_s dr\right)^2 ds}\right) \tag{13.2}$$

Note that the r.h.s. of the above estimate only involves the first derivative of v (we do not need to estimate the second derivative here!)

Remark 13.2. If we just estimate $\mathbb{E}|1 - \langle v, DX\rangle|$, we could write

$$d_W(X, Z) \le C\mathbb{E}|1 - \langle v, DX\rangle| \le C\left(\left\|\int_0^\infty v_r^2 dr - 1\right\|_1 + \sqrt{\mathbb{E}\int_0^\infty \left(\int_0^s v_r D_r v_s dr\right)^2 ds}\right)$$

Remark 13.3. Define $M_t = \mathbb{E}[X|\mathcal{F}_s] = \int_0^t v_s dB_s$ which is a continuous square integrable martingale. How do we connect continuous (local) martingale to a Brownian motion? Dambis–Dubins–Schwarz theorem tells that any continuous local martingale can be written as a time changed Brownian motion (with a possible enlargement of the probability space), and the time change is given by the quadratic variation. In other words, we can write

$$M_t = W_{\langle M\rangle_t}$$

where W is a standard Brownian motion, which is correlated with M in a rather complicated way. Thus, we can write

$$X = M_\infty = W_{\langle M\rangle_\infty}, \qquad \text{with } \langle M\rangle_\infty = \int_0^\infty v_s^2 ds$$

To compare the distribution of $X = W_{\langle M\rangle_\infty}$ with $N(0,1)$, it is the same as comparing the distribution with W_1, which indicates that the "error" should be related to $\langle M\rangle_\infty - 1 = \int_0^\infty v_s^2 ds - 1$. The above discussion makes it precise, i.e., in (13.2) we see that $\int_0^\infty v_s^2 ds - 1$ indeed appears in the estimate.

For martingales, there are other ways of proving central limit theorem, mostly through martingale CLT, and sometimes a quantitative martingale CLT leads to better estimates than (13.2).

13.3. **Example.** *CLT for a sequence of stochastic integral.*

Define

$$X_n = \int_0^1 \sqrt{2n} s^n e^{B_s(1-s)} dB_s$$

so $v_n(s) = \sqrt{2n} s^n e^{B_s(1-s)} 1_{[0,1]}(s)$, we first show that

$$\int_0^1 v_n(s)^2 ds = \int_0^1 2ns^{2n} e^{2B_s(1-s)} ds \to 1$$

in L^1. We can write it as $\int_0^1 2ns^{2n}[e^{2B_s(1-s)} - 1]ds + \int_0^1 2ns^{2n}ds$. The second term goes to 1 as $n \to \infty$. For the first term, we write it as

$$\int_0^1 2ns^{2n}[e^{2B_s(1-s)} - 1]ds = \int_0^{1-\delta} 2ns^{2n}[e^{2B_s(1-s)} - 1]ds + \int_{1-\delta}^1 2ns^{2n}[e^{2B_s(1-s)} - 1]ds$$

and see that it goes to zero as $n \to \infty$ (take δ small).

Now we compute

$$D_r v_n(s) = \sqrt{2n} s^n e^{B_s(1-s)}(1-s)1_{[0,s]}(r)$$

so

$$\int_0^s v_n(r) D_r v_n(s) dr = \int_0^s \sqrt{2n} r^n e^{B_r(1-r)} dr \times \sqrt{2n} s^n e^{B_s(1-s)}(1-s)$$

and we need to estimate

$$\mathbb{E} \int_0^1 \left(\int_0^s v_n(r) D_r v_n(s) dr \right)^2 ds = \mathbb{E} \int_0^1 \left(\int_0^s \sqrt{2n} r^n e^{B_r(1-r)} dr \right)^2 2ns^{2n} e^{2B_s(1-s)}(1-s)^2 ds$$

The idea is the same before, and the key here is that we have a factor of $(1-s)^2$, so the integral in $s \in [1-\delta, 1]$ will be small when δ is small.

Remark 13.4. A more direct way of seeing the "Gaussianity" from X_n is to write

$$\begin{aligned} X_n &= \int_0^{1-\delta} \sqrt{2n} s^n e^{B_s(1-s)} dB_s + \int_{1-\delta}^1 \sqrt{2n} s^n e^{B_s(1-s)} dB_s \\ &\approx \int_0^{1-\delta} \sqrt{2n} s^n e^{B_s(1-s)} dB_s + \int_{1-\delta}^1 \sqrt{2n} s^n dB_s \end{aligned}$$

The approximation is because when s is close to 1, $B_s(1-s)$ is close to zero. In the above expression, the first term goes to zero as $n \to \infty$ and the second term is a Wiener integral which is of Gaussian distribution. This is a typical example of "singular" stochastic integral, i.e., the integrand concentrates around a certain point, so the contribution to the randomness only comes from the Brownian increments near that point. Typically, in this case the limit will be independent of the original Brownian motion.

14. Lecture 14

The following is based on Sourav Chatterjee's book Chapter 5 https://link.springer.com/book/10.1007/978-3-319-03886-5

14.1. **Review of OU operator, OU semigroup, Gaussian-Poincare inequality.** Recall from lecture 4 that we have defined the OU operator L. Assuming the underlying probability space is generated by a sequence of i.i.d. standard Gaussian $\xi = \{\xi_i\}$. The OU operator L is defined as $L = -\sum_i A_i D_i$, where A_i is the adjoint of D_i. We can also write it more explicitly: $L = \sum_i (D_i^2 - \xi_i D_i)$. The associated semigroup is $P_t = e^{tL}$. As usual, $\{H_k/\sqrt{k!}\}_k$ is the ONB of $L^2(\Omega)$, and we know that $LH_k = -|k|H_k$. For any random variable $X \in L^2(\Omega)$, we decompose it according to the ONB and write $X = \sum_k \langle X, H_k\rangle H_k/k!$, then (on a formal level) we have

$$LX = -\sum_k \langle X, H_k\rangle |k| H_k/k!, \qquad P_t X = \sum_k \langle X, H_k\rangle e^{-t|k|} H_k/k!$$

It is clear from the above expression that P_t is a contraction from $L^2(\Omega)$ to $L^2(\Omega)$. We also have the Mehler's formula:

$$P_t X(\xi) = \mathbb{E}[X(e^{-t}\xi + \sqrt{1-e^{-2t}}\xi') \,|\, \xi]$$

The above formula can be interpreted as follows: X is a measure function of ξ that is square integrable, and we apply P_t to X to obtain another function of ξ, where the action of P_t is the same as the conditional expectation on the r.h.s., namely we introduce an independent copy of ξ, denoted by ξ', reduce the dependence on the original ξ by considering the interpolation $\xi_t := e^{-t}\xi + \sqrt{1-e^{-2t}}\xi'$, and then average out ξ'. The marginal distribution of ξ_t is the same as if we run independent OU processes with initial data ξ_i and evaluate them at time t.

It turns out that the OU semigroup has better contraction properties, which is the so-called hypercontractivity. This is very useful in practice. Recall the Gaussian-Poincare inequality that can be proved by the covariance estimate

$$\text{Var}X = \int_0^\infty e^{-t}\mathbb{E}\langle DX, P_t DX\rangle dt = \int_0^\infty e^{-t} \sum_i \mathbb{E}[D_i X P_t D_i X] dt \tag{14.1}$$

When we naively apply the $L^2 - L^2$ contraction, we obtain $\text{Var}X \leq \sum_i \mathbb{E}|D_i X|^2$. Note that the Gaussian-Poincare inequality is sharp (take $X = Z_i$), but in many cases, this gives the sub-optimal variance estimates. So one could ask, can we improve the Gaussian-Poincare inequality in some cases?

14.2. **Hypercontractivity of OU semigroup.** We first give a definition. A semigroup P_t is called hypercontractive if for any $p > 1$ and $t > 0$ there exists a $q = q(t, p) > p$ s such that for all $f \in L^p$, we have

$$\|P_t f\|_q \leq \|f\|_p$$

In other words the semigroup improves the integrability in a "contractive" way, with the improvement depending on the parameter t, i.e., how long we run the semigroup. Here is a theorem

Theorem 14.1. *The OU semigroup is hypercontractive with*

$$q(t, p) = 1 + (p-1)e^{2t}$$

Choose $p = 2$, we have

$$\|P_t f\|_{1+e^{2t}} \leq \|f\|_2$$

Question: what is the meaning of it when $t = \infty$?

We will give a stochastic calculus proof of the above result: note that we will have to estimate the L^q norm with $q \neq 2$, so we might have to use martingale inequalities.

Let us first check it on a simple example. Consider $X = e^{B_T}$ for some $T > 0$, we have

$$\|X\|_2 = \sqrt{\mathbb{E}e^{2B_T}} = e^T$$

and (by one of the homework problems)

$$P_t X = \mathbb{E}[e^{e^{-t}B_T + \sqrt{1-e^{-2t}}B'_T}|B] = e^{e^{-t}B_T + \frac{1}{2}(1-e^{-2t})T}$$

so

$$\begin{aligned}\|P_t X\|_q &= (\mathbb{E}e^{qe^{-t}B_T + q\frac{1}{2}(1-e^{-2t})T})^{1/q} = (e^{\frac{1}{2}q^2e^{-2t}T + q\frac{1}{2}(1-e^{-2t})T})^{1/q} \\ &= e^{\frac{1}{2}qe^{-2t}T + \frac{1}{2}(1-e^{-2t})T}\end{aligned}$$

For $\|P_t X\|_q \leq \|X\|_2$, we need

$$\frac{1}{2}qe^{-2t}T + \frac{1}{2}(1-e^{-2t})T \leq T \qquad \text{which implies} \qquad q \leq 1 + e^{2t}$$

In other words, the condition in the above theorem is sharp.

Before giving the proof, let us give an important applications of the above theorem, which is the so-called Talagrand $L^1 - L^2$ inequality.

Theorem 14.2 (Talagrand's inequality)**.** *Let X be a smooth r.v., and assume the sequence $\{A_i\}$ satisfies that $\|D_i X\|_2 \leq A_i$ for all $i \geq 1$, then*

$$\mathrm{Var}X \leq C \sum_i \frac{A_i^2}{1 + \log \frac{A_i}{\|D_i X\|_1}}$$

where $C > 0$ is some universal constant.

Remark: (i) taking $A_i = \|D_i X\|_2$, we know the above inequality implies the Gaussian-Poincare inequality since $\|D_i X\|_2 \geq \|D_i X\|_1$; (ii) if we expect an improvement of Gaussian-Poincare, then we should have $A_i \gg \|D_i X\|_1$ (or in some averaged sense after a summation in i). Still taking $A_i = \|D_i X\|_2$, if $D_i X = 1_{B_i}$ for some set B_i and $\mathbb{P}(B_i) \ll 1$, then we have $\|D_i X\|_2 = \sqrt{\mathbb{P}(B_i)} \gg \mathbb{P}(B_i) = \|D_i X\|_1$. See below the example of maximum of i.i.d. Gaussian random variables.

Proof. From (14.1),

$$\mathrm{Var}X = \int_0^\infty e^{-t}\mathbb{E}\langle DX, P_t DX\rangle dt = \int_0^\infty e^{-t} \sum_i \mathbb{E}[D_i X P_t D_i X] dt$$

we first estimate

$$|\mathbb{E}[D_i X P_t D_i X]| \leq \|D_i X\|_2 \|P_t D_i X\|_2$$

For each $t > 0$, choose $p < 2$ so that $2 = 1 + (p-1)e^{2t}$, by Hypercontractivity we have

$$\|P_t D_i X\|_2 \leq \|D_i X\|_p \leq \|D_i X\|_2^\alpha \|D_i X\|_1^\beta$$

where we applied Holder inequality in the second "$\leq$". Here $\alpha = 2 - \frac{2}{p}, \beta = \frac{2}{p} - 1$. Combine them, we have

$$\begin{aligned}|\mathbb{E}[D_i X P_t D_i X]| \leq \|D_i X\|_2^{1+\alpha}\|D_i X\|_1^\beta = \|D_i X\|_2^{3-\frac{2}{p}}\|D_i X\|_1^{\frac{2}{p}-1} &\leq A_i^{3-\frac{2}{p}}\|D_i X\|_1^{\frac{2}{p}-1} \\ &= A_i^2 \left(\frac{\|D_i X\|_1}{A_i}\right)^{\frac{2}{p}-1}\end{aligned}$$

Therefore,

$$\mathrm{Var}X \le \sum_i A_i^2 \int_0^\infty e^{-t}\left(\frac{\|D_iX\|_1}{A_i}\right)^{\frac{2}{1+e^{-2t}}-1} dt$$

Since $\frac{2}{1+e^{-2t}} - 1 = \tanh t \ge 1 - e^{-t}$, so we have

$$\mathrm{Var}X \le \sum_i A_i^2 \int_0^\infty e^{-t}\left(\frac{\|D_iX\|_1}{A_i}\right)^{1-e^{-t}} dt = \sum_i A_i^2 \int_0^1 \left(\frac{\|D_iX\|_1}{A_i}\right)^u du$$

Use the fact that $\int_0^1 e^{-\lambda u}du = \frac{1}{\lambda}(1-e^{-\lambda}) \le C\frac{1}{1+\lambda}$ for any $\lambda > 0$ (where C is some universal constant), we complete the proof. □

Remark 14.3. Typically the Talagrand's inequality leads to an improvement of log factor.

A simple application of Talagrand's inequality: define $X_n = \max_{i=1,\ldots,n} \xi_i$, what is the size of $\mathrm{Var}X_n$? We know that $D_iX_n = 1_{\xi_i = \max_j \xi_j}$, so $\|D_iX_n\|_1 = n^{-1}$ and $\|D_iX_n\|_2 = n^{-1/2}$. By Gaussian-Poincare inequality we have

$$\mathrm{Var}X_n \le \sum_i \|D_iX\|_2^2 = 1$$

By Talagrand's inequality we have

$$\mathrm{Var}X_n \le C\sum_i \frac{n^{-1}}{1+\log n^{1/2}} = C\frac{1}{1+\log n^{1/2}}$$

It is well-known that the variance of X_n is of order $1/\log n$, so Talagrand's inequality gives the sharp estimate.

15. Lecture 15

15.1. **Talagrand's inequality.** Recall the Talagrand's inequality we proved in last lecture takes the form

Theorem 15.1 (Talagrand's inequality)**.** *Let X be a smooth r.v., and assume the sequence $\{A_i\}$ satisfies that $\|D_iX\|_2 \le A_i$ for all $i \ge 1$, then*

$$\mathrm{Var}X \le C\sum_i \frac{A_i^2}{1+\log\frac{A_i}{\|D_iX\|_1}}$$

where $C > 0$ is some universal constant.

Here the underlying randomness is a sequence of i.i.d. standard normal (which basically is enough to reconstruct anything, e.g. a BM or a white noise). However, sometimes it is more convenient to use the derivative with other objects. For the Brownian motion case, we have the following version of the Talagrand's inequality: let X be a smooth functional of a standard BM and assume a function A_s satisfies that $\|D_sX\|_2 \le A_s$, then

$$\mathrm{Var}X \le C\int_0^\infty \frac{A_s^2}{1+\log\frac{A_s}{\|D_sX\|_1}} ds$$

The proof is the same: we start with the variance formula

$$\begin{aligned}\mathrm{Var}X &= \int_0^\infty e^{-t}\mathbb{E}\langle DX, P_t DX\rangle dt = \int_0^\infty e^{-t}\left(\int_0^\infty \mathbb{E}[D_s X P_t D_s X]ds\right)dt \\ &= \int_0^\infty \int_0^\infty e^{-t}\mathbb{E}[D_s X P_t D_s X]dtds\end{aligned}$$

Here $\langle\cdot,\cdot\rangle$ is the $L^2(\mathbb{R}_+)$ inner product. By following the previous proof verbatim, we have

$$\int_0^\infty e^{-t}\mathbb{E}[D_s X P_t D_s X]dt \le C\frac{A_s^2}{1+\log\frac{A_s}{\|D_s X\|_1}}$$

which completes the proof.

15.2. **Hypercontractivity of OU semigroup.** The goal is to show

Theorem 15.2. *The OU semigroup is hypercontractive with*

$$q(t,p) = 1+(p-1)e^{2t}$$

In other words, for any $p>1, t>0$, we have

$$\|P_t f\|_{q(t,p)} \le \|f\|_p$$

Proof. Fix p,t,q. To show that $\|P_t f\|_q \le \|f\|_p$, by duality, it is enough to prove

$$\mathbb{E}[P_t f g] \le \|f\|_p \|g\|_{q'}$$

where q' is the dual exponent of q, i.e., $\frac{1}{q}+\frac{1}{q'}=1$. By approximation, let us assume that f,g are smooth random variables taking values in $[a,b]$ with $0<a<b<\infty$ (why we can assume the r.v. is nonnegative? By Mehler's formula or comparison principle we have $|P_t f| \le P_t|f|$) We can assume f,g take the form

$$f = F(\xi_1,\ldots,\xi_N), \qquad g = G(\xi_1,\ldots,\xi_N)$$

By Mehler's formula, we can write

$$P_t f = \mathbb{E}[F(e^{-t}\xi_1+\sqrt{1-e^{-2t}}\xi_1',\ldots,e^{-t}\xi_N+\sqrt{1-e^{-2t}}\xi_N')|\xi]$$

where ξ' is an independent copy of ξ. We can also assume that $\xi_k = \int_0^1 \phi_k(s)dB_s$ and $\xi_k' = \int_0^1 \phi_k(s)dB_s'$, where $\{\phi_k\}$ is an ONB of $L^2[0,1]$, and B, B' are independent BM.

To summarize, we need to show that $\mathbb{E}XY \le \|X\|_p\|Y\|_{q'}$, with

$$X = F(e^{-t}\int_0^1 \phi_1(s)dB_s + \sqrt{1-e^{-2t}}\int_0^1 \phi_1(s)dB_s',\ldots,e^{-t}\int_0^1 \phi_N(s)dB_s+\sqrt{1-e^{-2t}}\int_0^1 \phi_N(s)dB_s')$$

$$Y = G(\int_0^1 \phi_1(s)dB_s,\ldots,\int_0^1 \phi_N(s)dB_s)$$

(Note that by Fubini we get rid of the conditional expectation). Define $\beta_s = e^{-t}B_s + \sqrt{1-e^{-2t}}B_s'$ which is a BM (note that t is fixed and the variable is s), then we can write

$$X = F(\int_0^1 \phi_1(s)d\beta_s,\ldots,\int_0^1 \phi_N(s)d\beta_s)$$

Since X^p is a functional of $\{\beta_s\}_{s\in[0,1]}$, by the martingale representation theorem we can write

$$X^p = \mathbb{E}X^p + \int_0^1 h_1(s)d\beta_s$$

where h_1 is a process adapted to the filtration generated by β. Similarly we have

$$Y^{q'} = \mathbb{E}Y^{q'} + \int_0^1 h_2(s)dB_s$$

(By Clark-Ocone formula, the h_1, h_2 can be written as conditional expectation of the Malliavin derivatives of $X^p, Y^{q'}$ but we will not need it here). Define the process (indexed by ℓ)

$$M_\ell = \mathbb{E}X^p + \int_0^\ell h_1(s)d\beta_s, \qquad N_\ell = \mathbb{E}Y^{q'} + \int_0^\ell h_2(s)dB_s$$

which are two martingales (note that by assumption of $f, g \in [a,b]$, we know that M, N are positive and bounded). The proof of $\mathbb{E}XY \le \|X\|_p\|Y\|_{q'}$ is the same as

$$\mathbb{E}M_1^{1/p}N_1^{1/q'} \le M_0^{1/p}N_0^{1/q'}$$

The idea is to apply Ito formula and show that under our assumption the drift is nonpositive. To simplify the notation. Let $\alpha = 1/p$ and $\gamma = 1/q'$.

$$\begin{aligned} d(M_\ell^\alpha N_\ell^\gamma) = &\text{ martingale terms} \\ &+ \frac{1}{2}N_\ell^\gamma \alpha(\alpha-1)M_\ell^{\alpha-2}d\langle M\rangle_\ell + \frac{1}{2}M_\ell^\alpha\gamma(\gamma-1)N_\ell^{\gamma-2}d\langle N\rangle_\ell + \alpha\gamma M_\ell^{\alpha-1}N_\ell^{\gamma-1}d\langle M,N\rangle_\ell \end{aligned}$$

Note that $d\langle M\rangle_\ell = h_1(\ell)^2d\ell, d\langle N\rangle_\ell = h_2(\ell)^2d\ell$ and $d\langle M,N\rangle_\ell = h_1(\ell)h_2(\ell)e^{-t}d\ell$ (recall that β, B are correlated Brownian motions!), then we can write the above drift term as

$$\frac{1}{2}M_\ell^\alpha N_\ell^\gamma\left[\alpha(\alpha-1)\frac{h_1(\ell)^2}{M_\ell^2} + \gamma(\gamma-1)\frac{h_2(\ell)^2}{N_\ell^2} + 2\alpha\gamma e^{-t}\frac{h_1(\ell)h_2(\ell)}{M_\ell N_\ell}\right]d\ell$$

Note that $\alpha, \gamma \in (0,1)$, by $a^2+b^2 \ge 2ab$ we have

$$\alpha(1-\alpha)\frac{h_1(\ell)^2}{M_\ell^2} + \gamma(1-\gamma)\frac{h_2(\ell)^2}{N_\ell^2} \ge 2\sqrt{\alpha\gamma(1-\alpha)(1-\gamma)}\frac{h_1(\ell)h_2(\ell)}{M_\ell N_\ell}$$

Thus, for the drift to be nonpositive, we only need

$$2\alpha\gamma e^{-t} \le 2\sqrt{\alpha\gamma(1-\alpha)(1-\gamma)}$$

Since $\alpha = 1/p$ and $\gamma = 1/q' = 1 - 1/q$, from the above inequality we derive that $q \le 1 + e^{2t}(p-1)$. This completes the proof. $\square$

15.3. **An application of hypercontractivity.** Suppose we consider random variables living in a fixed chaos, meaning that for some given $n \ge 0$,

$$X = \sum_{|k|=n}\langle X, H_k\rangle\frac{H_k}{k!}$$

We have

$$P_tX = \sum_{|k|=n}\langle X, H_k\rangle e^{-|k|t}\frac{H_k}{k!} = e^{-nt}X$$

By Hypercontractivity, we know that with $q = 1 + (p-1)e^{2t}$

$$\|P_t X\|_q = e^{-nt}\|X\|_q \le \|X\|_p$$

In other words, for any $q > p$, we have

$$\|X\|_q \le e^{\frac{n}{2}\log\frac{q-1}{p-1}}\|X\|_p = \Big(\frac{q-1}{p-1}\Big)^{n/2}\|X\|_p$$

This shows that on a fixed order of chaos, the two norms $\|\cdot\|_p$ and $\|\cdot\|_q$ are equivalent. In many applications, the estimates on the high order moments are reduced to estimating the second order moment.

16. Lecture 16

16.1. **Application of Talagrand's inequality to directed polymer in random environment.** Recall the Talagrand's inequality we proved in last lecture takes the form

Theorem 16.1 (Talagrand's inequality)**.** *Let X be a smooth r.v., and assume the sequence $\{A_i\}$ satisfies that $\|D_i X\|_2 \le A_i$ for all $i \ge 1$, then*

$$\mathrm{Var}X \le C\sum_i \frac{A_i^2}{1 + \log\frac{A_i}{\|D_i X\|_1}} \tag{16.1}$$

where $C > 0$ is some universal constant.

Here the underlying randomness is a sequence of i.i.d. standard normal (which basically is enough to reconstruct anything, e.g. a BM or a white noise). We also stated a version that involves the Malliavin derivative with respect to an underlying Brownian motion.

In this lecture we apply the above inequality to give a (partial) proof of the superconcentration of directed polymers in random environment. Here, superconcentration means that the variance of the object of interest grows sublinearly with respect to the size of the system. In contrary, the classical concentration case is when we have sum of i.i.d. random variables with finite variance $S_n = X_1 + \ldots + X_n$ then $\mathrm{Var}\,S_n \sim n$ (and the superconcentration case would just be $n^{-1}\mathrm{Var}\,S_n \to 0$ as $n \to \infty$.

Recall the model of directed polymer. Let $\omega = \{\omega_{ij}\}$ be i.i.d. $N(0,1)$ random variables which models the random environment, and let $S = \{S_n\}$ be a symmetric simple random walk. The polymer measure is the tilt of the random walk measure by the Radon-Nikodym derivative $\frac{e^{H_n(\omega,S)}}{\sum_S e^{H_n(\omega,S)}}$, where

$$H_n(\omega, S) = \sum_{j=1}^{n} \omega_{j,S_j}$$

and $\sum_S$ is the summation over all possible trajectories. Define $Z_n(\omega) = \sum_S e^{H_n(\omega,S)}$ which is the partition function. Note that Z_n inherits the randomness from ω. The interest here is on the size of the variance of $\log Z_n$ (the moments of Z_n can be computed easily but somehow that does not say anything interesting about the physics here). In $d = 1$, one expects that $\mathrm{Var}\log Z_n \sim n^{2/3}$, which has not been proved for this model. In high dimensions, almost nothing is known. The goal here is to apply (23.5) to show that

Theorem 16.2. *In all dimensions we have*

$$\mathrm{Var} \log Z_n \le C \frac{n}{\log n}$$

for some universal constant $C > 0$. In other words, $\log Z_n$ superconcentrates.

To simplify the notation, define $X_n = \log Z_n$. Use D_{ij} to denote the derivative with respect to ω_{ij}. We have

$$D_{ij} X_n = D_{ij} \log Z_n = Z_n^{-1} D_{ij} Z_n$$

Since $Z_n = \sum_S e^{H_n(\omega,S)}$, we have

$$D_{ij} Z_n = \sum_S e^{H_n(\omega,S)} D_{ij} H_n(\omega, S) = \sum_S e^{H_n(\omega,S)} 1_{S_i=j}$$

This gives

$$D_{ij} X_n = \frac{\sum_S e^{H_n(\omega,S)} 1_{S_i=j}}{\sum_S e^{H_n(\omega,S)}}$$

In other words, we can interpret $D_{ij}X_n$ as the probability of $S_i = j$ under the polymer measure, and we have the easy Gaussian-Poincare upper bound $\mathrm{Var} X_n \le \sum_{i,j} \mathbb{E}|D_{ij}X_n|^2 \le \sum_{i,j} \mathbb{E}|D_{ij}X_n| = n$. To apply (23.5), we need to estimate $\|D_{ij}X_n\|_1$ and $\|D_{ij}X_n\|_2$ which turns out to be not easy. The hope is that some upper bound of $\|D_{ij}X_n\|_2$ could be much larger than $\|D_{ij}X_n\|_1$. A naive bound would be $|D_{ij}X_n| \le 1$, which does not give anything helpful.

Here we introduce the trick of a local averaging, which in some sense makes the Malliavin derivative small artificially. For any $x \in \mathbb{Z}$, define

$$Z_n(x) = \sum_{S:S(0)=x} e^{H_n(\omega,S)}$$

In other words, we start the SSRW from x (rather than 0) so our original Z_n is just $Z_n(0)$. Since the random environment is given by i.i.d. random variables, it is easy to see that $Z_n(x) \overset{\text{law}}{=} Z_n(0)$, for any $x \in \mathbb{Z}$, and $\{Z_n(x)\}_x$ is actually a stationary random field. One could think of $\log Z_n(x)$ as the height function of an interface evolving in random environment, with n, x the temporal and spatial variables respectively.

Define a local averaging

$$\bar{X}_n = \frac{1}{(2K+1)^d} \sum_{|x| \le K} \log Z_n(x)$$

where $K \gg 1$ is some constant to be determined, and $\{|x| \le K\}$ is the box of size $(2K+1)^d$.

The idea is to estimate $\mathrm{Var}\bar{X}_n$ using Talagrand's inequality and at the same time to estimate the difference between $\bar{X}_n$ with $\log Z_n(0)$. If the interface is not too "rough", one may be able to control $\bar{X}_n - \log Z_n(0)$.

We divide the proof into several steps.

Step 1. We have

$$D_{ij}\bar{X}_n = \frac{1}{(2K+1)^d} \sum_{|x|\le K} Z_n(x)^{-1} D_{ij} Z_n(x) = \frac{1}{(2K+1)^d} \sum_{|x|\le K} \left(Z_n(x)^{-1} \sum_{S:S(0)=x} e^{H_n(\omega,S)} 1_{S_i=j} \right)$$
$$=: \frac{1}{(2K+1)^d} \sum_{|x|\le K} Y_n(x)$$

where $Y_n(x)$ is the term in the bracket. We claim that

$$Y_n(x) \stackrel{\text{law}}{=} Z_n(0)^{-1} \sum_{S:S(0)=0} e^{H_n(\omega,S)} 1_{S_i=j-x}$$

which is actually easy to see through the stationarity. Thus, we have

$$\|D_{ij}\bar{X}_n\|_1 = \frac{1}{(2K+1)^d} \sum_{|x|\le K} \|Y_n(x)\|_1 \le \frac{1}{(2K+1)^d}$$

$$\|D_{ij}\bar{X}_n\|_2^2 \le \frac{1}{(2K+1)^d} \sum_{|x|\le K} \|Y_n(x)\|_2^2 \le \frac{1}{(2K+1)^d} \sum_{|x|\le K} \|Y_n(x)\|_1 = \|D_{ij}\bar{X}_n\|_1$$

where we used the deterministic bound $|Y_n(x)| \le 1$. Take $A_{ij} = \sqrt{\|D_{ij}\bar{X}_n\|_1}$ so we have $\|D_{ij}\bar{X}_n\|_2 \le A_{ij}$. By Talagrand's inequality we have

$$\mathrm{Var}\bar{X}_n \le C \sum_{i,j} \frac{A_{ij}^2}{1+\log \frac{A_{ij}}{\|D_{ij}\bar{X}_n\|_1}} = C \sum_{i,j} \frac{A_{ij}^2}{1+\log \frac{1}{\sqrt{\|D_{ij}\bar{X}_n\|_1}}} \le C \sum_{i,j} \frac{A_{ij}^2}{1+\log \sqrt{(2K+1)^d}}$$

On the other hand, we have

$$\sum_{i,j} A_{ij}^2 = \sum_{i,j} \|D_{ij}\bar{X}_n\|_1 = \sum_{i,j} \left(\frac{1}{(2K+1)^d} \sum_{|x|\le K} \|Y_n(x)\|_1 \right) = n$$

and this implies that $\mathrm{Var}\bar{X}_n \le Cn/\log K$.

Step 2. We need to estimate the variance of $\bar{X}_n - \log Z_n(0)$. We can write it as

$$\bar{X}_n - \log Z_n(0) = \frac{1}{(2K+1)^d} \sum_{|x|\le K} [\log Z_n(x) - \log Z_n(0)]$$

First we have $\mathbb{E}[\log Z_n(x) - \log Z_n(0)] = 0$, so the l.h.s. has mean zero. Then

$$\sqrt{\mathrm{Var}[\bar{X}_n - \log Z_n(0)]} = \|\bar{X}_n - \log Z_n(0)\|_2 \le \frac{1}{(2K+1)^d} \sum_{|x|\le K} \|\log Z_n(x) - \log Z_n(0)\|_2$$

Suppose we can estimate $\|\log Z_n(x) - \log Z_n(0)\|_2$, which could come from some PDE regularity estimate, and the estimates we are looking for could take the form

$$\|\log Z_n(x) - \log Z_n(0)\|_2 \le C|x| \tag{16.2}$$

(for example, provided we can estimate the moments of the derivative of $\log Z_n(x)$ with respect to x, uniformly in n). Assuming (19.2), we obtain

$$\sqrt{\mathrm{Var}[\bar{X}_n - \log Z_n(0)]} \le \frac{C}{(2K+1)^d} \sum_{|x|\le K} |x| \le CK$$

Note that if we view $\log Z_n(x)$ as the height of an evolving interface at time n and spatial location x, then (19.2) basically says that the interface can not be too rough in the sense that the transversal height grows at most linearly. This is the so-called subroughness introduced in the recent paper of Chatterjee "Superconcentration in surface growth", see it here.

Step 3. Recall that the goal is to estimate the variance of $\log Z_n(0)$. We have

$$\begin{aligned}\mathrm{Var}\log Z_n(0) &= \mathrm{Var}[\log Z_n(0) - \bar{X}_n + \bar{X}_n] \le 2\mathrm{Var}[\log Z_n(0) - \bar{X}_n] + 2\mathrm{Var}\bar{X}_n \\ &\le 2C^2K^2 + 2C\frac{n}{\log K}\end{aligned}$$

Choose $K = n^\alpha$ for some $\alpha < 1/2$, we derive that

$$\mathrm{Var}\log Z_n(0) \le C\frac{n}{\log n}$$

To summarize, modulo the estimate (19.2), we proved the superconcentration result for the free energy of directed polymer in random environment. The proof of (19.2) in a continuous setting can be found in Propoistion 3.2 of this paper here, and in the discrete setting here, see Theorem 2.4.

17. Lecture 17

The following is based on Sourav Chatterjee's book Chapter 6 https://link.springer.com/book/10.1007/978-3-319-03886-5

17.1. **The spectral approach to improve the Gaussian-Poincare inequality.** First, let us recall that the Talagrand's inequality improves the Gaussian-Poincare inequality (typically by a log factor), and the underlying driving force is the hypercontractivity of the OU semigroup. Here we will present another way of improving it by the Hermite polynomial expansion.

Let us first recall how one proves the Gaussian-Poicare inequality using the Hermite polynomial expansion. Let $\xi = \{\xi_i\}$ be a sequence of i.i.d. $N(0,1)$, and $H_k(\xi) = \prod_j H_{k_j}(\xi_j)$ is the Hermite polynomial with multiindex $k = (k_1, k_2, \ldots)$. We have proved that $\{\frac{H_k(\xi)}{\sqrt{k!}}\}_k$ is an ONB of $L^2(\Omega)$.

For any $X \in L^2(\Omega)$, we have

$$X = \sum_k \mathbb{E}[XH_k]\frac{H_k}{k!} = \sum_k c_k\frac{H_k}{k!}$$

where $c_k := \mathbb{E}[XH_k]$. Recall that for any $m \ge 0$, we call

$$X_m := \sum_{k:|k|=m} c_k\frac{H_k}{k!}$$

the $m-$th chaos of X, and we have $X = \sum_{m\ge 0} X_m$.

For any j, we have

$$D_jX = \sum_{k\neq 0} c_k\frac{D_jH_k}{k!} = \sum_{k\neq 0} c_k\frac{k_jH_{k-e_j}}{k!}$$

Here e_j is the index with the $j-$th component equaling to 1 and zero otherwise. Now we can compute L^2 norms and prove the Gaussian-Poincare inequality

$$\mathrm{Var}X = \mathbb{E}(X - \mathbb{E}X)^2 = \sum_{k\neq 0} c_k^2 \frac{1}{k!}$$

$$\mathbb{E}(D_j X)^2 = \sum_{k\neq 0} c_k^2 k_j^2 \frac{1}{k_j k!} = \sum_{k\neq 0} c_k^2 \frac{k_j}{k!}$$

$$\sum_j \mathbb{E}(D_j X)^2 = \sum_{k\neq 0} c_k^2 \frac{|k|}{k!} \geq \mathrm{Var}X$$

If we rewrite the summation according to the value of $|k|$, then we have

$$\mathrm{Var}X = \sum_{m\geq 1} \mathrm{Var}X_m = \sum_{m\geq 1} \sum_{k:|k|=m} c_k^2 \frac{1}{k!}$$

$$\mathbb{E}\|DX\|^2 = \sum_j \mathbb{E}(D_j X)^2 = \sum_{m\geq 1} m \sum_{k:|k|=m} c_k^2 \frac{1}{k!} = \sum_{m\geq 1} m\mathrm{Var}X_m$$

Thus, the problem comes from multiplying the variance of the $m-$th order chaos by the number m. If the random variable X lives in high order chaos, the upper bound obtained by Gaussian-Poincare inequality deteriorates.

The spectral approach to improve the Gaussian-Poincare inequality is easy to describe: we choose some $N > 0$, and write

$$\mathrm{Var}X = \sum_{m=1}^{N-1} \mathrm{Var}X_m + \sum_{m\geq N} \sum_{k:|k|=m} c_k^2 \frac{1}{k!}$$

Since

$$\mathbb{E}\|DX\|^2 \geq N \sum_{m\geq N} \sum_{k:|k|=m} c_k^2 \frac{1}{k!}$$

the variance can be bounded by

$$\mathrm{Var}X \leq \sum_{m=1}^{N-1} \mathrm{Var}X_m + \frac{1}{N}\mathbb{E}\|DX\|^2 \tag{17.1}$$

which is the "improved" Gaussian-Poincare inequality. Note that if $N = 1$, we recover the usual Gaussian-Poincare inequality. The point here is, if we can compute or estimate $\mathrm{Var}X_m$ for those $m = 1, \ldots, N-1$ and at the same time choose $N \gg 1$ to obtain a balance of the two terms on the r.h.s. of the above equation, it might be possible to beat the upper bound given by $\mathbb{E}\|DX\|^2$.

Still, we need to compute $\mathrm{Var}X_m = \sum_{k:|k|=m} c_k^2 \frac{1}{k!}$. Here

$$c_k = \mathbb{E}XH_k$$

We claim that

$$c_k = \mathbb{E}\prod_j D_j^{k_j} X \tag{17.2}$$

which comes from integration by parts (recall that in the 1-d case we have $\langle X, H_k\rangle = \langle DX, H_{k-1}\rangle = \ldots = \langle D^k X, 1\rangle = \mathbb{E}D^k X$). Here one should view $\prod_j D_j^{k_j}$ as a high order partial differential operator.

Given $m \geq 1$, let us write

$$
\begin{aligned}
\sum_{k:|k|=m} c_k^2 \frac{1}{k!} &= \sum_{k:|k|=m} (\mathbb{E} \prod_j D_j^{k_j} X)^2 \frac{1}{k!} \\
&= \frac{1}{m!} \sum_{k:|k|=m} (\mathbb{E} \prod_j D_j^{k_j} X)^2 \frac{(\sum_j k_j)!}{\prod_j k_j!}.
\end{aligned} \tag{17.3}
$$

If we write $X = f(\xi_1, \xi_2, ...)$, and use the usual notations of the partial differential operator, then

$$
\sum_{k:|k|=m} (\mathbb{E} \prod_j D_j^{k_j} X)^2 \frac{(\sum_j k_j)!}{\prod_j k_j!} = \sum_{i_1,\ldots,i_m} (\mathbb{E} \partial^m_{i_1,\ldots,i_m} f)^2 \tag{17.4}
$$

Here the summation $\sum_{i_1,\ldots,i_m}$ is over all possible $i_1, \ldots, i_m \geq 1$, and $\partial^m_{i_1,\ldots,i_m} f$ is the derivative of f with respect to $\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_m}$ (e.g. i_1 could equal to i_2). In other words, we can write

$$
\mathrm{Var} X_m = \sum_{k:|k|=m} \frac{c_k^2}{k!} = \frac{1}{m!} \sum_{i_1,\ldots,i_m} (\mathbb{E} \partial^m_{i_1,\ldots,i_m} f)^2 \tag{17.5}
$$

We summarize the spectral approach one could try to apply:

Step 1: get an upper bound $\mathbb{E}\|DX\|^2$.

Step 2: for all multiindex k such that $|k| \leq N-1$, compute/estimate $c_k = \mathbb{E} X H_k = \mathbb{E} \prod_j D_j^{k_j} X$.

Step 3: choose N to optimize the upper bound obtained from (17.1).

17.2. **Application to the SK model.** Here we use the spectral approach to prove the superconcentration in the SK model. Let us recall what the model is (see the introduction in lecture 6).

For any n, the configuration space is $\{\pm 1\}^n$, i.e., a vector of length n with ± 1 entries. We write such vector by $\sigma = (\sigma_i)_{i=1,\ldots,n}$ with $\sigma_i = \pm 1$. We view σ_i as the spin at site i. The interaction between different spins are modeled by a Gaussian random variable x_{ij}, and the Hamiltonian is $H_n(\sigma, x) = \frac{1}{\sqrt{n}} \sum_{i<j} \sigma_i \sigma_j x_{ij}$. We view $x = \{x_{ij}\}$ as the random environment determining the interaction between different spins. Given the realization of x, the Gibbs measure on the spin configuration is defined as

$$
\mathbb{P}_n[\sigma] = \frac{e^{\beta H_n(\sigma,x)}}{\sum_\sigma e^{\beta H_n(\sigma,x)}}
$$

where β is the inverse temperature. This is the so-called Sherrington-Kirkpatrick model. The goal is to study the behavior of the Gibbs measure when n is very large. The key quantity is the partition function

$$
Z_n(x) = \sum_\sigma e^{\beta H_n(\sigma,x)}
$$

To simplify the notation, we will omit the dependence on x from now on. Our goal is to give a (partial) proof of the fact

Theorem 17.1. *For all values of* $\beta > 0$, *we have* $\mathrm{Var} \log Z_n \ll n$.

Remark 17.2. In the high temperature regime $\beta \in (0,1)$, more precise results are known, in particular, the variance is bounded uniformly in n and a central limit theorem holds for $\log Z_n$ (after centering). Almost nothing is known in the low temperature regime $\beta > 1$.

Recall that by Gaussian-Poincare inequality we get an upper bound of order $O(n)$: let D_{ij} be the derivative with respect to x_{ij}, we have

$$D_{ij}\log Z_n = Z_n^{-1}D_{ij}Z_n = Z_n^{-1}\sum_\sigma \frac{\beta}{\sqrt{n}}e^{\beta H_n(\sigma,x)}\sigma_i\sigma_j$$

By the elementary fact $|\sigma_i\sigma_j| = 1$, we have $|D_{ij}\log Z_n| \le \beta/\sqrt{n}$. This implies that

$$\sum_{i,j}|D_{ij}\log Z_n|^2 \le \beta^2\frac{n(n-1)}{2n} \le Cn$$

so we have $\mathrm{Var}\log Z_n \le \sum_{i,j}\mathbb{E}|D_{ij}\log Z_n|^2 \le Cn$.

To improve the $O(n)$ bound, we need to compute the variance of the first few chaos explicitly. Let us introduce the notation of ensemble average $\langle\cdot\rangle$, which is the expectation taking under the Gibbs measure (for each realization of the random environment), e.g. we have

$$Z_n^{-1}\sum_\sigma e^{\beta H_n(\sigma,x)}\sigma_i\sigma_j = \frac{\sum_\sigma e^{\beta H_n(\sigma,x)}\sigma_i\sigma_j}{\sum_\sigma e^{\beta H_n(\sigma,x)}} = \langle\sigma_i\sigma_j\rangle$$

thus

$$D_{ij}\log Z_n = \frac{\beta}{\sqrt{n}}\langle\sigma_i\sigma_j\rangle$$

Taking expectation, we have $\mathbb{E}D_{ij}\log Z_n = \frac{\beta}{\sqrt{n}}\mathbb{E}\langle\sigma_i\sigma_j\rangle$ (by (17.2), we know that this equals to $\mathbb{E}[x_{ij}\log Z_n]$). One needs to think about it, but after taking into account the randomness from the environment, the distribution of σ_i,σ_j are independent and takes value ±1 with probability $1/2$. More precisely, under the "total measure", $\sigma = (\sigma_1,\dots,\sigma_n)$ is uniformly distributed on $\{\pm1\}^n$. Thus we actually have $\mathbb{E}\langle\sigma_i\sigma_j\rangle = 0$. We need to compute higher order coefficient. For example, take any ℓ, k, we have

$$\begin{aligned}D_{\ell k}D_{ij}\log Z_n &= \frac{\beta}{\sqrt{n}}D_{\ell k}\langle\sigma_i\sigma_j\rangle = \frac{\beta}{\sqrt{n}}D_{\ell k}\frac{\sum_\sigma e^{\beta H_n(\sigma,x)}\sigma_i\sigma_j}{\sum_\sigma e^{\beta H_n(\sigma,x)}}\\ &= \frac{\beta^2}{n}\left(\frac{e^{\beta H_n(\sigma,x)}\sigma_i\sigma_j\sigma_\ell\sigma_k}{\sum_\sigma e^{\beta H_n(\sigma,x)}} - \frac{\sum_\sigma e^{\beta H_n(\sigma,x)}\sigma_i\sigma_j}{\sum_\sigma e^{\beta H_n(\sigma,x)}}\frac{\sum_\sigma e^{\beta H_n(\sigma,x)}\sigma_\ell\sigma_k}{\sum_\sigma e^{\beta H_n(\sigma,x)}}\right)\\ &= \frac{\beta^2}{n}[\langle\sigma_i\sigma_j\sigma_k\sigma_\ell\rangle - \langle\sigma_i\sigma_j\rangle\langle\sigma_k\sigma_\ell\rangle]\end{aligned} \tag{17.6}$$

(note that if $(ij) = (\ell k)$, we have $\mathbb{E}[D_{\ell k}D_{ij}\log Z_n] = \mathbb{E}(x_{ij}^2-1)\log Z_n$, otherwise we have $\mathbb{E}[D_{\ell k}D_{ij}\log Z_n] = \mathbb{E}[x_{ij}x_{\ell k}\log Z_n]$). Then one can imagine that a similar calculation leads to

$$D_{i_1j_1}D_{i_2j_2}\dots D_{i_mj_m}\log Z_n = \frac{\beta^m}{n^{m/2}}\sum[\dots]$$

and what is inside the summation is something similar to (17.6). Through induction (for details, see Chatterjee's book page 61-62), one can show that (which is a nontrivial calculation)

$$\sum_{(i_1,j_1),\dots,(i_m,j_m)}\Big(\mathbb{E}D_{i_1j_1}D_{i_2j_2}\dots D_{i_mj_m}\log Z_n\Big)^2 \le \beta^{2m}(m-1)!m^{2m}$$

The key here is that one needs to utilize the cancellation coming from the summation over all $\sigma_j = \pm1$.

If we denote the $m-$th order chaos of $\log Z_n$ by X_m, then by (17.5), we have

$$\mathrm{Var} X_m = \frac{1}{m!}\sum_{(i_1,j_1),\ldots,(i_m,j_m)} \Big(\mathbb{E} D_{i_1j_1}D_{i_2j_2}\ldots D_{i_mj_m}\log Z_n\Big)^2 \leq \beta^{2m}m^{2m-1}$$

This implies that for any N, we have

$$\mathrm{Var}\log Z_n \leq \sum_{m=1}^{N-1}\mathrm{Var}X_m + \frac{Cn}{N} \leq \sum_{m=1}^{N-1}\beta^{2m}m^{2m-1} + \frac{Cn}{N}$$

where the Cn factor comes from the Gaussian-Poincare bound. Take $N = \delta\frac{\log n}{\log\log n}$ for some δ sufficiently small, one can check that $\mathrm{Var}\log Z_n \leq Cn\frac{\log\log n}{\log n} \ll n$.

18. Lecture 18

18.1. **Review.** What we did so far can be summarized as follows:

(i) in the general Gaussian setting, introduce the concept of Malliavin derivative, adjoint operator, OU operator;

(ii) applications to estimating fluctuations: concentration inequality, Gaussian-Poincare inequality, Hypercontractivity, Talagrand's inequality, improved Gaussian-Poincare inequality;

(iii) applications of integration by parts: Stein's method, second order Poincare inequality, Clark-Ocone formula.

We mostly focused on two classical examples: (i) a sequence of i.i.d. standard Gaussian random variables (ii) the standard Brownian motion. In the rest of the lectures, we will move on to another infinite dimensional setting: the spacetime white noise and the study of stochastic PDE.

18.2. **Isonormal Gaussian process.** Recall the definition of isonormal Gaussian process over some Hilbert space H: $\{W(h) : h \in H\}$ is a family of centered Gaussian random variables so that $\mathbb{E}W(h_1)W(h_2) = \langle h_1, h_2\rangle$. The two classical examples: (i) $H = \ell_2$, then the isonormal Gaussian process over H is "equivalent" with a sequence of i.i.d. standard Gaussian $\{\xi_k\}_{k\geq 1}$ where $\xi_k = W(e_k)$ and e_k is the canonical basis of ℓ_2. We have $W(h) = \sum_{k\geq 1} h_k\xi_k$. (ii) $H = L^2(\mathbb{R}_+)$, then the isonormal Gaussian process over H is "equivalent" with a standard Brownian motion B with $W(h) = \int_0^\infty h_s dB_s$.

In both cases, we view $W(h)$ as h integrate against the underlying white noise. Consider the case $H = L^2(\mathbb{R}_+\times\mathbb{R}^d)$, the spacetime white noise is the isonormal Gaussian process over H. Being white simply means that for any $A, B \subset [0,\infty)\times\mathbb{R}^d$ such that $A\cap B = \emptyset$, we have $W(1_A)$ being independent of $W(1_B)$. If we write $W(A) = W(1_A)$, one can view $W(\cdot)$ as an $L^2(\Omega)$ valued random measure, i.e., for each set A with finite Lebesgue measure, $W(A) \in L^2(\Omega)$, and $W(\cup_i A_i) = \sum_i W(A_i)$ if A_i are mutually disjoint and $\sum_i |A_i| < \infty$ (however, it is not a signed measure almost surely, why?) For deterministic function $f \in H$, we write $W(f) = \int_0^\infty\int_{\mathbb{R}^d} f(t,x)\xi(t,x)dxdt$ (which is only a symbol), and view $\xi(t,x)$ as the density of the aforementioned measure, i.e., heuristically we have $\xi(t,x)dxdt = W(dxdt)$. In this way, one can approximate the function f by simple functions $f \approx \sum_i f_i 1_{A_i}$ and the integral $\int_0^\infty\int_{\mathbb{R}^d} f(t,x)\xi(t,x)dxdt$ is approximated by $\sum_i f_i W(A_i)$, which is a weighted sum of independent Gaussians.

Another way to look at it is to define the so-called Brownian sheet: for any $t \geq 0, x \in \mathbb{R}^d$

$$B(t,x) = W(1_{[0,t]\times[0,x]}) = \int_{[0,t]\times[0,x]} \xi(s,y)dyds$$

which is a continuous Gaussian process. Then $\xi(t,x)$ can be defined as the derivative of B (in the sense of distribution):

$$\xi(t,x) = \frac{\partial^{1+d}}{\partial_t \partial_{x_1} \ldots \partial_{x_d}} B(t,x)$$

In many cases, the physical systems we are interested in are subjected to random perturbations that can be modeled by the spacetime white noise (or spatial white noise depending on the problem), so it is helpful to define stochastic integral with respect to the spacetime white noise.

Example of stochastic heat equation (SHE)

(i) additive SHE $\partial_t u = \Delta u + \xi$. If ξ is a smooth function, the solution takes the form

$$u(t,x) = q_t \star u_0(x) + \int_0^t \int_{\mathbb{R}^d} q_{t-s}(x-y)\xi(s,y)dyds.$$

For random ξ, the key is to ensure the stochastic integral makes sense. We need $\int_0^t \int_{\mathbb{R}^d} q_{t-s}(x-y)^2 dyds < \infty$, which only holds in $d = 1$. Thus, the additive SHE only makes sense in 1d. In high dimensions, one can view $\int_0^t \int_{\mathbb{R}^d} q_{t-s}(\cdot - y)\xi(s,y)dyds$ as a random Schwarz distribution, and test it with a test function $f(\cdot)$, which gives us

$$\int_0^t \int_{\mathbb{R}^d} f \star q_{t-s}(y)\xi(s,y)dyds$$

It is easy to check that if f is a nice function, then the above Wiener integral is well-defined.

(ii) multiplicative SHE $\partial_t u = \Delta u + u\xi$. Similarly in this case we expect the solution takes the form

$$u(t,x) = q_t \star u_0(x) + \int_0^t \int_{\mathbb{R}^d} q_{t-s}(x-y)u(s,y)\xi(s,y)dyds \tag{18.1}$$

Now the problem becomes more complicated as we have multiplication of u with ξ.

There are different ways of defining the above stochastic integral. One way leads to the so-called Skorohod integral: we write

$$\int_0^t \int_{\mathbb{R}^d} q_{t-s}(x-y)u(s,y)\xi(s,y)dyds = \int_0^\infty \int_{\mathbb{R}^d} 1_{[0,t]}(s) q_{t-s}(x-y)u(s,y)\xi(s,y)dyds$$

and view the r.h.s. as $\delta(1_{[0,t]}(\cdot)q_{t-\cdot}(x-\cdot)u(\cdot,\cdot))$ where $\delta(\cdot)$ is the divergence operator, i.e., the adjoint of the Malliavin derivative. In this way, we can deal with spatial white noise as well.

Another approach leads to the Ito-type integral, and the key is to use the "time direction" and define filtration, and for u adapted to the filtration, the multiplication of $u(s,y)$ and $\xi(s,y)dyds$ is the multiplication of u with the "future" random increment. (In the case when the noise is only spatial and there is no time direction, this approach does not exist)

18.3. **Ito integration with respect to white noise.** One way to define the Ito integral is through a more functional analytic perspective and to make use of the Ito integral with respect to standard BM. We briefly sketch it here. Let $\{e_k(x)\}_{k\geq 1}$ be an ONB of $L^2(\mathbb{R}^d)$, it is easy to check that $B_k(t) := W(1_{[0,t]}(\cdot)e_k(\cdot))$ is a sequence of independent standard Brownian motions, and we can formally write $\xi(t,x) = \sum_k B_k'(t)e_k(x)$ (since $B_k(t) = \int_0^t \int_{\mathbb{R}^d} e_k(x)\xi(s,x)dxds$) in the sense that for any $f \in L^2([0,\infty)\times\mathbb{R}^d)$, we have

$$\int_0^\infty \int_{\mathbb{R}^d} f(t,x)\xi(t,x)dxdt = \sum_k \int_0^\infty \left(\int_{\mathbb{R}^d} f(t,x)e_k(x)dx\right) dB_k(t)$$

When f is a random process adapted to the filtration generated by the Brownian motions, the integral can be defined in the same way, i.e., we first project f in each direction, then for each component which is again an adapted process, we perform the Ito integration with respect to the standard Brownian motion.

Here we will adopt another (slightly different) definition through a simple process approximation. Before going to the definition, let us digress to discuss a separable subject.

18.4. **The case of a spatial white noise.** Consider the equation written formally as

$$\partial_t u = \Delta u + u\xi(x)$$

where ξ is a white noise on $L^2(\mathbb{R}^d)$ (so it only depends on the spatial variable and we call it a spatial white noise). This is the so-called parabolic Anderson model. If we write the PDE in the mild form

$$u(t,x) = q_t \star u_0(x) + \int_0^t \int_{\mathbb{R}^d} q_{t-s}(x-y)u(s,y)\xi(y)dyds \tag{18.2}$$

it is unclear how it should be interpreted. Again one way is to define the stochastic integral on the r.h.s. in the Skorohod sense, i.e., to write it as $\delta(1_{[0,t]}(\cdot)q_{t-\cdot}(x-\cdot)u(\cdot))$ where $\delta(\cdot)$ is the divergence operator. But it is unclear whether this definition is "physical" or not. Here the meaning of being "physical" is simply about whether the resulting equation describes the real physical phenomenon.

One way to check whether the definition of the solution to SPDE is physical or not is as follows. Take a mollifier $\phi_\varepsilon(x) = \varepsilon^{-d}\phi(x/\varepsilon)$ where $\phi \in C_c^\infty(\mathbb{R}^d)$ and $\int \phi = 1$. Define $\xi_\varepsilon(x) = \int_{\mathbb{R}^d} \phi_\varepsilon(x-y)\xi(y)dy = W(\phi_\varepsilon(x-\cdot))$. One can check that ξ_ε is a smooth Gaussian process. So the following "mollified" equation admits classical solution

$$\partial_t u_\varepsilon = \Delta u_\varepsilon + u_\varepsilon \xi_\varepsilon(x)$$

By sending $\varepsilon \to 0$, one can ask whether $u_\varepsilon(x)$ converges or not and if it does converge, does the limit solve the SPDE? If the limit indeed solves the SPDE (in the way we defined solution), then it means that the solution is physical, i.e., it is an approximate model of u_ε for $\varepsilon \ll 1$.

In this case, the definition of the stochastic integral $\int_0^t \int_{\mathbb{R}^d} q_{t-s}(x-y)u(s,y)\xi(y)dyds$ is still unclear. In some sense, one uses the limiting procedure to define the stochastic integral in the Stratonovitch sense:

$$\int_0^t \int_{\mathbb{R}^d} q_{t-s}(x-y)u(s,y)\xi(y)dyds := \lim_{\varepsilon\to 0} \int_0^t \int_{\mathbb{R}^d} q_{t-s}(x-y)u_\varepsilon(s,y)\xi_\varepsilon(y)dyds$$

if the r.h.s. converges.

Remark 18.1. One could ask whether one can pick a certain direction as the direction of time and simply define Ito integral as we did before. The problem is that the mild formulation (18.2) is not compatible with it.

Remark 18.2. In many cases, the approximate solution u_ε does not converge, but if we "renormalize" the equation, it does. This falls into the subject of singular SPDE, and basically started ten years ago with the work of Martin Hairer.

19. Lecture 19

19.1. **Ito integral with respect to spacetime white noise.** Let us give some definitions. For any $h \in L^2(\mathbb{R}_+ \times \mathbb{R}^d)$, define $X_t(h) = \int_0^\infty \int_{\mathbb{R}^d} h(s,y) 1_{[0,t]}(s)\xi(s,y)dyds$, which one can check is a Gaussian process with independent increments. Let $\mathcal{F}_t(h)$ be the filtration generated by $X_t(h)$, and define $(\mathcal{F}_t)_{t\geq 0}$ such that $\mathcal{F}_t$ is the $\sigma-$algebra generated by $\mathcal{F}_t(h), h \in L^2(\mathbb{R}_+ \times \mathbb{R}^d)$. It is easy to check that $\mathcal{F}_t$ is just the $\sigma-$algebra generated by random variables of the form $W(B)$ such that $B \subset [0,t] \times \mathbb{R}^d$.

Define elementary random field as follows: $\{\Phi(t,x) : t \geq 0, x \in \mathbb{R}^d\}$ is elementary if it can be written as

$$\Phi(t,x) = X 1_{(a,b]}(t) 1_A(x)$$

where X is $\mathcal{F}_a$ measurable and $A \subset \mathbb{R}^d$ is some bounded set. A random field Φ is called simple if there exists elementary fields Φ_i with disjoint supports such that $\Phi = \sum_i \Phi_i = \sum_i X_i 1_{(a_i,b_i]}(t) 1_{A_i}(x)$, where X_i is $\mathcal{F}_{a_i}$ measurable. One should view simple random field as "piecewise constant" which can be used to approximated more complicated random field.

For any elementary field Φ, define

$$\int_0^\infty \int_{\mathbb{R}^d} \Phi(t,x)\xi(t,x)dxdt = X \int_0^\infty \int_{\mathbb{R}^d} 1_{(a,b]}(t) 1_A(x)\xi(t,x)dxdt = XW((a,b] \times A)$$

For simple random field $\Phi = \sum_i \Phi_i$, define

$$\int_0^\infty \int_{\mathbb{R}^d} \Phi(t,x)\xi(t,x)dxdt = \sum_i \int_0^\infty \int_{\mathbb{R}^d} \Phi_i(t,x)\xi(t,x)dxdt$$

To simplify the notation, from now on, we write $\int_0^\infty \int_{\mathbb{R}^d} \Phi(t,x)\xi(t,x)dxdt = \int \Phi(t,x)\xi(t,x)dxdt$. Note that with $\Phi = \sum_i \Phi_i = \sum_i X_i 1_{(a_i,b_i]} 1_{A_i}$, then the above definition gives

$$\int \Phi(t,x)\xi(t,x)dxdt = \sum_i X_i W((a_i,b_i] \times A_i)$$

which can be viewed as a Riemann sum.

Proposition 19.1. *For simple process Φ, we have*

(i) $\mathbb{E} \int \Phi(t,x)\xi(t,x)dxdt = 0$

(ii)

$$\mathbb{E}\Big(\int \Phi(t,x)\xi(t,x)dxdt\Big)^2 = \int_0^\infty \int_{\mathbb{R}^d} \mathbb{E}\Phi(t,x)^2 dxdt \tag{19.1}$$

Proof. (i) For elementary field $\Phi(t,x) = X1_{(a,b]}(t)1_A(x)$, we use the fact that X is independent of $W((a,b] \times A)$ to complete the proof.

(ii) Define the set $B_i = (a_i, b_i] \times A_i \subset \mathbb{R}_+ \times \mathbb{R}^d$, and assume $\Phi(t,x) = \sum_i X_i 1_{B_i}(t,x)$ and $B_i \cap B_j = \emptyset$. We have $\int \Phi(t,x)\xi(t,x)dxdt = \sum_i A_i$ with

$$A_i = \int X_i 1_{(a_i,b_i]}(t) 1_{A_i}(x)\xi(t,x)dxdt = X_i W(B_i)$$

We first show that if $i \neq j$, $\mathbb{E}A_iA_j = 0$. WLOG assume $a_i \leq a_j$. We have

$$\mathbb{E}A_iA_j = \mathbb{E}X_iX_jW(B_i)W(B_j)$$

By definition we know that $B_i \cap B_j = \emptyset$ and since $a_i \leq a_j$, X_i, X_j is $\mathcal{F}_{a_j}$ measurable, we conclude that $W(B_j)$ is independent of $X_iX_jW(B_i)$, thus $\mathbb{E}A_iA_j = \mathbb{E}[X_iX_jW(B_i)]\mathbb{E}W(B_j) = 0$. On the other hand, we have

$$\mathbb{E}A_i^2 = \mathbb{E}X_i^2W(B_i)^2 = \mathbb{E}X_i^2|B_i|$$

so

$$\sum_i \mathbb{E}A_i^2 = \sum_i \mathbb{E}X_i^2|B_i| = \int \mathbb{E}\Phi(t,x)^2 dxdt$$

which completes the proof. □

Remark 19.2. The equation (23.5) should be viewed as an Ito isometry (one should compare the above proof with the case of BM, the only difference is notational, i.e., we have the extra dependence on x). One can use (23.5) to extend the definition of Ito integral to more general integrands.

Using (23.5), we can extend the Ito integral defined above to general predictable field Φ such that

$$\int_0^\infty \int_{\mathbb{R}^d} \mathbb{E}\Phi(t,x)^2 dxdt < \infty \tag{19.2}$$

we do not go into the detail here.

Here is an expected result

Proposition 19.3. *For predictable random field Φ such that (19.2) holds, we define*

$$M_t = \int_0^t \int_{\mathbb{R}^d} \Phi(s,y)\xi(s,y)dyds,$$

then M is a continuous square-integrable martingale and $\langle M\rangle_t = \int_0^t \int_{\mathbb{R}^d} \Phi(s,y)^2 dyds$.

Proof. The proof is very similar to the Brownian motion case so we only sketch the argument here. First, by approximation, it suffices to show the martingale property when Φ is an elementary random field, in which case

$$M_t = \int_0^t \int_{\mathbb{R}^d} X1_{(a,b]}(s)1_A(y)\xi(s,y)dyds = \int_0^\infty \int_{\mathbb{R}^d} X1_{(a,b]\cap[0,t]}(s)1_A(y)\xi(s,y)dyds$$

Now we can write M_t explicitly (depending on the value of t) and check martingale property explicitly. The general case is proved by an approximation (note that one needs to first check it when $\Phi = \sum_i \Phi_i$ is a simple field). □

19.2. **SHE with multiplicative noise.** Now we can sketch the construction of the solution to SHE with multiplicative noise. We do it in the more general setting and consider the nonlinear case:

$$\partial_t u = \Delta u + \sigma(u)\xi(t,x), \qquad u(0,x) = u_0(x) \tag{19.3}$$

Here σ is a Lipchitz function. We only consider $d = 1$ and the case when u_0 is bounded. The mild solution to the above equation is a random field that is adapted to the filtration generated by ξ and the following equation holds almost surely

$$u(t,x) = q_t \star u_0(x) + \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)\sigma(u(s,y))\xi(s,y)dyds$$

Theorem 19.4. *There exists a unique random field u solving (20.1) such that for any $T > 0$, we have $\sup_{t\in[0,T],x\in\mathbb{R}} \mathbb{E}u(t,x)^2 < \infty$.*

Proof. We first show uniqueness. Assume both u, v solve the equation, we have

$$u(t,x) - v(t,x) = \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)[\sigma(u(s,y)) - \sigma(v(s,y))]\xi(s,y)dyds$$

By Ito isometry, we have

$$\mathbb{E}|u(t,x) - v(t,x)|^2 = \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)^2\mathbb{E}[\sigma(u(s,y)) - \sigma(v(s,y))]^2 dyds$$

Since σ is Lipchitz, we have

$$\mathbb{E}|u(t,x) - v(t,x)|^2 \leq C\int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)^2\mathbb{E}|u(s,y) - v(s,y)|^2 dyds$$

Define $f(s) = \sup_{y\in\mathbb{R}} \mathbb{E}|u(s,y) - v(s,y)|^2$, then we have

$$\mathbb{E}|u(t,x) - v(t,x)|^2 \leq C\int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)^2 f(s)dyds = C'\int_0^t \frac{1}{\sqrt{t-s}}f(s)ds$$

thus $f(t) \leq C' \int_0^t \frac{1}{\sqrt{t-s}}f(s)ds$. First, we apply a Holder inequality to obtain

$$f(t)^p \leq C'^p \left(\int_0^t \frac{1}{(t-s)^{q/2}}ds\right)^{p/q} \int_0^t f(s)^p ds$$

Choose $q < 2$, we have for all $t < T$, $f(t)^p \leq C_{p,T} \int_0^t f(s)^p ds$. Further applying Gronwall's inequality and using the fact that $f(0) = 0$, we conclude that $f(t) = 0$.

Now we show existence. Define u_n recursively by

$$u_n(t,x) = q_t \star u_0(x) + \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)\sigma(u_{n-1}(s,y))\xi(s,y)dyds \tag{19.4}$$

we write the difference as

$$u_{n+1}(t,x) - u_n(t,x) = \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)[\sigma(u_n(s,y)) - \sigma(u_{n-1}(s,y))]\xi(s,y)dyds$$

Again by Ito isometry and the fact that σ is Lipchitz, we have

$$\mathbb{E}|u_{n+1}(t,x) - u_n(t,x)|^2 \leq C\int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)^2|u_n(s,y) - u_{n-1}(s,y)|^2 dyds$$

Define $z_n(s) = \sup_y |u_n(s,y) - u_{n-1}(s,y)|^2$, so we have (by the same calculation as before)

$$z_{n+1}(t) \le C' \int_0^t \frac{1}{\sqrt{t-s}} z_n(s) ds$$

and again (by Holder inequality)

$$z_{n+1}(t)^p \le C_{p,T} \int_0^t z_n(s)^p ds \le C_{p,T}^2 \int_0^t \int_0^s z_{n-1}(\ell)^p d\ell ds \le \ldots \le \frac{(Ct)^n}{n!}$$

for some constant C (depending on T, p). The fast decay in n ensures that e.g.

$$\sum_n \sup_{s \in [0,T]} z_n(s) < \infty$$

which implies the convergence of u_n. Sending $n \to \infty$ in (19.4), we complete the proof. □

20. Lecture 20

20.1. **Review of Ito integral with respect to spacetime white noise and the multiplicative stochastic heat equation.** Recall that for simple processes of the form $\Phi(t,x) = \sum_i X_i 1_{(a_i,b_i]}(t) 1_{A_i}(x)$ where X_i is $\mathcal{F}_{a_i}$ measurable, we have defined the Riemann sum as the Ito integral

$$\int \Phi(t,x)\xi(t,x) dx dt = \sum_i X_i W((a_i,b_i] \times A_i)$$

and extended the definition to all random fields that is adapted to the filtration and satisfies $\int \mathbb{E}\Phi(t,x)^2 dx dt < \infty$.

For the equation of the form $\partial_t u = \frac{1}{2}\Delta u + u\xi(t,x)$, we have defined the solution as the unique random field such that

$$u(t,x) = q_t \star u_0(x) + \int_0^t \int q_{t-s}(x-y) u(s,y) \xi(s,y) dy ds \tag{20.1}$$

with the stochastic integral interpreted in the Ito sense.

20.2. **Chaos expansion for SHE.** One can iterate the mild formulation (20.1):

$$\begin{aligned} u(t,x) &= q_t \star u_0(x) + \int_0^t \int q_{t-s}(x-y) \left(q_s \star u_0(y) + \int_0^s \int q_{s-\ell}(y-z) u(\ell,z)\xi(\ell,z) dz d\ell \right) \xi(s,y) dy ds \\ &= q_t \star u_0(x) + \int_0^t \int q_{t-s}(x-y) q_s \star u_0(y) \xi(s,y) dy ds + \ldots \end{aligned}$$

By computing the $L^2(\Omega)$ norm, one can check the series converges and also different terms in the above series are orthogonal to each other. In fact, the above series expansion is the chaos expansion of $u(t,x) \in L^2(\Omega)$ (how to do it at least heuristically? Wick theorem)

20.3. **A Wong-Zakai theorem.** It is actually unclear in the equation $\partial_t u = \frac{1}{2}\Delta u + u\xi(t,x)$, what is the meaning of the product $u\xi(t,x)$. You can ask the same question for SDE. Consider the toy example $dX_t = X_t dB_t, X_0 = 1$, we know that the solution is $X_t = e^{B_t - t/2}$. We can also write the Stratonovitch equation $dX_t = X_t \circ dB_t$ whose solution is $X_t = e^{B_t}$. Consider a smooth perturbation of B_t constructed in the following way: write $B_t = \int 1_{[0,t]}\xi(s)ds$ where ξ is a (time) white noise, and define $\xi_\varepsilon(t) = \int \phi_\varepsilon(t-s)\xi(s)ds$ which is a smooth Gaussian process (for each $\varepsilon > 0$), then define $B_t^\varepsilon = \int 1_{[0,t]}(s)\xi_\varepsilon(s)ds$ which is an approximation of B. The ODE $dX_t^\varepsilon = X_t^\varepsilon dB_t^\varepsilon$ can be interpreted in the classical sense, as B^ε has smooth trajectory, and the solution is $X_t^\varepsilon = e^{B_t^\varepsilon}$. So we see that the Stratonovitch solution is the "physical" one in the sense that it is stable under smooth perturbations. On the other hand, if we want to get the Ito limit, it is necessary to "renormalize" the equation and consider $dX_t^\varepsilon = X_t^\varepsilon dB_t^\varepsilon - \frac{1}{2}X_t^\varepsilon dt$ so that $X_t^\varepsilon = e^{B_t^\varepsilon - \frac{1}{2}t}$. It is unclear at this stage what this renormalization means.

One way to understand the difference between the Ito and Stratonovitch solution is as follows. The approximate differential equation takes the form $dX_t^\varepsilon = X_t^\varepsilon dB_t^\varepsilon$, so X_t^ε depends on B_t^ε, hence on ξ_t^ε, up to time t. One can check that it does not matter how we choose the mollifier ϕ_ε, there is always some overlap between X_t^ε and ξ_t^ε: for example, suppose we pick $\phi_\varepsilon(\cdot)$ to support only on $\mathbb{R}_-$, then for each t, ξ_t^ε depends on $\xi(s)$ for $s \geq t$, but in this case X_t^ε is not $\mathcal{F}_t-$measurable, so the product $X_t^\varepsilon(B_t^\varepsilon)'$ still contributes nontrivially in the limit.

Remark 20.1. There are more general results on SDE of the form $dX_t = b(X_t)dt + \sigma(X_t)dB_t$, and the corresponding result is the classical Wong-Zakai theorem, with the renormalization term $-\frac{1}{2}X_t^\varepsilon dt$ called the Wong-Zakai correction term (or Ito-Stratonovitch correction). It is intimately related to the rough path theory.

We can ask the same question for SPDE. Define the smoothed white noise $\xi_\varepsilon(t,x) = \int \phi_\varepsilon(t-s, x-y)\xi(s,y)dyds$, and consider the PDE $\partial_t u_\varepsilon = \frac{1}{2}\Delta u_\varepsilon + u_\varepsilon \xi_\varepsilon(t,x)$. Here we assume the mollifier takes the form $\phi_\varepsilon(t,x) = \frac{1}{\varepsilon^3}\phi(t/\varepsilon^2, x/\varepsilon)$. The following is an infinite dimensional version of the Wong-Zakai theorem:

Theorem 20.2. *There exists two positive constants c_1, c_2 such that $u_\varepsilon(t,x)e^{-c_1 t/\varepsilon - c_2 t} \Rightarrow u(t,x)$, where u solves the equation interpreted in the Ito sense.*

We will not give a full proof here but only focus on two aspects: (i) how to determine the renormalization constant c_1, c_2? (ii) how to show the convergence of the first chaos? The main tool will be the Feynman-Kac formula. Recall that for heat equation with a potential $\partial_t u = \frac{1}{2}\Delta u + uV(t,x)$, the solution can be written as

$$u(t,x) = \mathbb{E}_B u_0(x+B_t)e^{\int_0^t V(t-s, x+B_s)ds}$$

(i) *Determining the renormalization constants.* Starting from (20.1), we know that $\bar{u}(t,x) := \mathbb{E}u(t,x)$ solves the standard heat equation. Suppose we have some uniform integrability then we must choose c_1, c_2 so that $\mathbb{E}u_\varepsilon(t,x)e^{-c_1 t/\varepsilon - c_2 t} \to \mathbb{E}u(t,x) = \bar{u}(t,x)$. The expectation $\mathbb{E}u_\varepsilon(t,x)$ can be computed through the Feynman-Kac formula.

First we write

$$u_\varepsilon(t,x) = \mathbb{E}_B u_0(x+B_t)e^{\int_0^t \xi_\varepsilon(t-s, x+B_s)ds}$$

so we have

$$\mathbb{E}u_\varepsilon(t,x) = \mathbb{E}_B u_0(x+B_t)\mathbb{E}e^{\int_0^t \xi_\varepsilon(t-s, x+B_s)ds}$$

For each realization of B, we have $X_\varepsilon := \int_0^t \xi_\varepsilon(t-s, x+B_s)ds$ is of Gaussian distribution of mean zero and variance

$$\mathrm{Var} X_\varepsilon = \int_0^t \int_0^t R_\varepsilon(s_1 - s_2, B_{s_1} - B_{s_2}) ds_1 ds_2$$

Here R_ε is the covariance function of the Gaussian process ξ_ε. This implies

$$\mathbb{E} u_\varepsilon(t,x) = \mathbb{E}_B u_0(x + B_t) e^{\frac{1}{2}\mathrm{Var} X_\varepsilon}$$

The r.h.s. only depends on the Brownian motion. Recall that $\xi_\varepsilon = \phi_\varepsilon \star \xi$ and $\phi_\varepsilon(t,x) = \varepsilon^{-3}\phi(t/\varepsilon^2, x/\varepsilon)$ for some ϕ. We have $R_\varepsilon(t,x) = \varepsilon^{-3}R(t/\varepsilon^2, x/\varepsilon)$ with $R(t,x) = \int \phi(t+s, x+y)\phi(s,y)dyds$. By the scaling property of BM, one can check that

$$\mathrm{Var} X_\varepsilon \stackrel{law}{=} \varepsilon \int_0^{t/\varepsilon^2} \int_0^{t/\varepsilon^2} R(s_1 - s_2, B_{s_1} - B_{s_2}) ds_1 ds_2$$

One can write the r.h.s. as

$$\begin{aligned} & 2\varepsilon \int_0^{t/\varepsilon^2} \left(\int_0^{s_1} R(s_1 - s_2, B_{s_1} - B_{s_2}) ds_2 \right) ds_1 \\ & = 2\varepsilon \int_0^{t/\varepsilon^2} \left(\int_0^{s_1} R(s_2, B_{s_1} - B_{s_1 - s_2}) ds_2 \right) ds_1 \end{aligned}$$

Assume R is compactly supported (if ϕ is), e.g. $R(s,\cdot) = 0$ if $s > M$ for some fixed M, then the above term

$$2\varepsilon \int_0^{t/\varepsilon^2} \left(\int_0^{s_1} R(s_2, B_{s_1} - B_{s_1 - s_2}) ds_2 \right) ds_1 \approx 2\varepsilon \int_0^{t/\varepsilon^2} \left(\int_0^{M} R(s_2, B_{s_1} - B_{s_1 - s_2}) ds_2 \right) ds_1 = 2\varepsilon \int_0^{t/\varepsilon^2} Y_{s_1} ds_1$$

The process $Y_{s_1} := \int_0^M R(s_2, B_{s_1} - B_{s_1 - s_2}) ds_2$ is stationary and has a finite range of dependence. One can show that

$$\varepsilon \int_0^{t/\varepsilon^2} Y_s ds - c_1 t/\varepsilon \Rightarrow \sigma W_t$$

where W is a Brownian motion independent of B. Here $c_1 = \mathbb{E} Y_s$. Then we have

$$\begin{aligned} \mathbb{E} u_\varepsilon(t,x) e^{-c_1 t/\varepsilon - c_2 t} &= \mathbb{E}_B u_0(x + B_t) e^{\frac{1}{2}\mathrm{Var} X_\varepsilon - c_1 t/\varepsilon - c_2 t} \\ &\approx \mathbb{E}_B u_0(x + B_t) e^{\sigma W_t - c_2 t} = \mathbb{E}_B u_0(x + B_t) = \bar{u}(t,x) \end{aligned}$$

if we choose $c_2 = \frac{1}{2}\sigma^2$.

Remark 20.3. In the SDE example, one simply compute $\mathbb{E} e^{\frac{1}{2}B_t^\varepsilon} \approx e^{\frac{1}{2}t}$ to determine the renormalization.

21. Lecture 21

21.1. **Chaos expansion of SHE.** Recall that the goal is to show $u_\varepsilon(t,x)e^{-c_1 t/\varepsilon - c_2 t} \to u(t,x)$, where u solves the SHE $\partial_t u = \frac{1}{2}\Delta u + u\xi$ and u_ε solves a mollified version $\partial_t u_\varepsilon = \frac{1}{2}\Delta u_\varepsilon + u_\varepsilon \xi_\varepsilon$. From last lecture, we determined the two constants c_1, c_2 so that $\mathbb{E} u_\varepsilon(t,x) e^{-c_1 t/\varepsilon - c_2 t} \to \mathbb{E} u(t,x)$. It remains to show the convergence. The distribution of $u(t,x)$ is difficult to characterize: it is not Gaussian and the moments grow to fast to determine the law (one actually has $\mathbb{E} u(t,x)^n \sim e^{n^3 t}$).

We will use the chaos expansion, which in a sense writes the random variable explicitly. For the limit, we have

$$u(t,x) = q_t \star u_0(x) + \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)u(s,y)\xi(s,y)dyds \tag{21.1}$$

By iterating the above formulation, we have $u(t,x) = \sum_{n\geq 0} I_n(t,x)$, with $I_0(t,x) = q_t \star u_0(x)$, $I_1(t,x) = \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)q_s \star u_0(y)\xi(s,y)dyds$, etc. One can check that each $I_n(t,x)$ lives in the $n-$th order Wiener chaos associated with the spacetime white noise.

The idea is to write $u_\varepsilon(t,x)$ in chaos expansion as well and prove the corresponding chaos converges. This turns out to be not easy. One can write down the mild form of the PDE satisfied by u_ε:

$$u_\varepsilon(t,x) = q_t \star u_0(x) + \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)u_\varepsilon(s,y)\xi_\varepsilon(s,y)dyds \tag{21.2}$$

and try to iterate. However, the resulting series is very different from the one obtained before. At this point one should think about the difference between (21.1) and (21.2): one is Ito integral and the other is Lebesgue integral.

21.2. **Stroock formula.** We use the so-called Stroock formula. Recall that in the 1d case when $X = f(Z)$ and $Z \sim N(0,1)$, we can expand X in Hermite polynomials: $X = \sum_{n\geq 0}\langle X, H_n(Z)\rangle \frac{H_n(Z)}{n!}$, and the expansion coefficients can be computed through integration by parts

$$\langle X, H_n(Z)\rangle = \mathbb{E}[f(Z)H_n(Z)] = \mathbb{E}[D^n f(Z)]$$

In other words, to compute the expansion coefficients, one only needs to compute the Malliavin derivatives and take the average. This so-called Stroock formula holds in a very general setting. For example, when we have a spacetime white noise ξ, and $X = f(\xi)$ is a complicated functional of ξ that is infinitely Malliavin differentiable (in this case $H = L^2(\mathbb{R}_+ \times \mathbb{R})$ and DX is $H-$valued random variable, D^2X is $H \otimes H$ valued random variable etc), then the expansion coefficient of X can be obtained by computing the multi-order Malliavin derivative and taking the average.

Consider the first two chaos as an example. Assume $X \in L^2(\Omega)$ is decomposed into Wiener chaos: $X = \sum_{n\geq 0} X_n$. We know that if $X \in D^{1,2}$, then DX is an $H-$valued random variable and $\mathbb{E}\|DX\|_H^2 < \infty$. Since $H = L^2(\mathbb{R}_+ \times \mathbb{R})$, we can write $DX = (D_{s,y}X)_{s\geq 0, y\in\mathbb{R}}$ and interpret $D_{s,y}X$ as the derivative of X with respect to the infinitesimal increment of the noise at (s,y). It turns out that $f_1(s,y) := \mathbb{E}D_{s,y}X$ is the expansion coefficient of the first chaos: $X_1 = \int_0^\infty \int_{\mathbb{R}} f_1(s,y)\xi(s,y)dyds$. Similarly, if $X \in D^{2,2}$ then D^2X is an $H \otimes H-$valued random variable, and we can write it as $D^2X = (D_{s_1,y_1}D_{s_2,y_2}X)_{s_1,s_2\geq 0, y_1,y_2\in\mathbb{R}}$. If we define $f_2(s_1,y_1,s_2,y_2) = \mathbb{E}[D_{s_1,y_1}D_{s_2,y_2}X]$, then

$$X_2 = \int_0^\infty \int \left(\int_0^{s_2} \int_{\mathbb{R}} f_2(s_1,y_2,s_2,y_2)\xi(s_1,y_1)dy_1ds_1 \right) \xi(s_2,y_2)dy_2ds_2$$

We will not go into the proof of this "fact" here. This can be found in the classical book of David Nualart.

Remark 21.1. A good exercise to convince oneself the above "fact" is to consider the 1d BM case and a random variable $X = \sum_n X_n$ with each X_n written as Iterated Ito integral with respect to BM: $X_n = \int 1_{s_1<s_2<\ldots<s_n} f_n(s_1,\ldots,s_n)dB_{s_1}dB_{s_2}\ldots dB_{s_n}$. To obtain the coefficient f_n, one can proceed as follows: (i) compute $D_{s_1}D_{s_2}\ldots D_{s_n}X$ (with $s_1 < \ldots < s_n$), and observe that only X_k with $k \geq n$ contributes; (ii) taking expectation, all terms with $k \geq n+1$ vanish, and what remains is $f_n(s_1,\ldots,s_n)$.

21.3. **Expansion coefficient for $u_\varepsilon(t,x)$.** With the above Stroock formula, one can try to compute the expansion coefficient of $u_\varepsilon(t,x)$. Again, we do it through the Feynman-Kac formula

$$u_\varepsilon(t,x)e^{-c_1t/\varepsilon^2-c_2t} = \mathbb{E}_B[u_0(x+B_t)e^{\int_0^t \xi_\varepsilon(t-s,x+B_s)ds-c_1t/\varepsilon-c_2t}]$$

Recall that $\xi_\varepsilon(t,x) = \int \psi_\varepsilon(t-s,x-y)\xi(s,y)dyds$, so for each realization of the Brownian motion, one can write

$$\begin{aligned} X_\varepsilon = \int_0^t \xi_\varepsilon(t-s,x+B_s)ds &= \int \left(\int_0^t \phi_\varepsilon(t-s-\ell,x+B_s-y)ds\right)\xi(\ell,y)dyd\ell \\ &=: \int \Psi^\varepsilon_{B,t,x}(\ell,y)\xi(\ell,y)dyd\ell = W(\Psi^\varepsilon_{B,t,x}) \end{aligned}$$

To compute the Malliavin derivative, we write

$$\begin{aligned} D_{\ell,y}u_\varepsilon(t,x)e^{-c_1t/\varepsilon^2-c_2t} &= D_{\ell,y}\mathbb{E}_B[u_0(x+B_t)e^{W(\Psi^\varepsilon_{B,t,x})-c_1t/\varepsilon-c_2t}] \\ &= \mathbb{E}_B[u_0(x+B_t)D_{\ell,y}e^{W(\Psi^\varepsilon_{B,t,x})-c_1t/\varepsilon-c_2t}] \\ &= \mathbb{E}_B[u_0(x+B_t)e^{W(\Psi^\varepsilon_{B,t,x})-c_1t/\varepsilon-c_2t}\Psi^\varepsilon_{B,t,x}(\ell,y)] \end{aligned}$$

The high order Malliavin derivatives can be computed in the same way. Let us focus on the first order for the moment. By Stroock formula, the expansion coefficient should be

$$\begin{aligned} f_1(\ell,y) &= \mathbb{E}D_{\ell,y}u_\varepsilon(t,x)e^{-c_1t/\varepsilon^2-c_2t} = \mathbb{E}\mathbb{E}_B[u_0(x+B_t)e^{W(\Psi^\varepsilon_{B,t,x})-c_1t/\varepsilon-c_2t}\Psi^\varepsilon_{B,t,x}(\ell,y)] \\ &= \mathbb{E}_B[u_0(x+B_t)e^{\frac12\mathrm{Var}X_\varepsilon-c_1t/\varepsilon-c_2t}\Psi^\varepsilon_{B,t,x}(\ell,y)] \end{aligned}$$

Here $\mathrm{Var}X_\varepsilon$ is the second moment of X_ε conditioning on B. By the discussion in last lecture, we have

$$\frac12\mathrm{Var}X_\varepsilon - c_1t/\varepsilon - c_2t \Rightarrow \sigma W_t - c_2t$$

where W is a Brownian motion independent of B. On the other hand, since $\Psi^\varepsilon_{B,t,x}(\ell,y) = \int_0^t \phi_\varepsilon(t-s-\ell,x+B_s-y)ds$ and ϕ_ε approximates the Dirac function, one expects something like $\Psi^\varepsilon_{B,t,x}(\ell,y) \to \delta(x+B_{t-\ell}-y)$ (note that this is only formal and makes no rigorous mathematical sense). Therefore, on the formal level, we expect that

$$f_1(\ell,y) \to \mathbb{E}_B[u_0(x+B_t)e^{\sigma W_t-c_2t}\delta(x+B_{t-\ell}-y)] = \mathbb{E}_B[u_0(x+B_t)\delta(x+B_{t-\ell}-y)] = q_{t-\ell}(x-y)q_\ell\star u_0(y)$$

which is the coefficient of the first order chaos of the limiting SHE.

The above heuristics can be made rigorous, see this paper here. The same discussion applies to all chaos.

22. Lecture 22-23

The following is based on this and this paper.

22.1. **A central limit theorem for SHE.** Consider the 1d multiplicative stochastic heat equation $\partial_t u = \Delta u + \sigma(u)\xi$, where ξ is a spacetime white noise and we start from constant initial data $u(0,x) \equiv 1$. Since the initial data is a constant, and the noise is a stationary random field, one can check that, for each $t > 0$, $\{u(t,x)\}_{x\in\mathbb{R}}$ is a stationary random field. The goal here is to understand the mixing property of this random field. Since $t > 0$ is fixed, we write $u(t,x) = U(x)$.

We know that

$$u(t,x) = 1 + \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)\sigma(u(s,y))\xi(s,y)dyds$$

so $\mathbb{E}u(t,x) = 1$. A natural question is whether we have LLN of the form $\frac{1}{N}\int_0^N U(x)dx \to 1$ and CLT of the form $\frac{1}{\sqrt{N}}\int_0^N (U(x)-1)dx \Rightarrow N(0,\sigma^2)$. Here is the theorem

Theorem 22.1. *There exists a positive constant $\sigma^2 > 0$ such that*

$$\frac{1}{\sqrt{N}}\int_0^N (U(x)-1)dx \Rightarrow N(0,\sigma^2)$$

as $N \to \infty$.

The result shows that to some extent one can view $\int_0^N U(x)dx$ as sum of N almost independent random variables. Note that they are not strictly independent, because by the mild formulation, we see that $U(x)$ is actually a measurable function of the noise everywhere in $[0,t]\times\mathbb{R}$. Let us first try to compute the two-point covariance function in the linear case

$$\mathbb{E}u(t,x)u(t,y) = 1 + \int_0^t \int_{\mathbb{R}} q_{t-s}(x-z)q_{t-s}(y-z)\mathbb{E}u(s,z)^2 dzds$$

Denote $f(s) = \mathbb{E}u(s,z)^2 > 0$, we have

$$\mathbb{E}u(t,x)u(t,y) - 1 = \int_0^t \int_{\mathbb{R}} q_{2(t-s)}(x-y)f(s)ds$$

We know that $\sup_{s\in[0,t]} f(s) < C_t$, so the covariance is bounded from above by

$$\text{Cov}[u(t,x),u(t,y)] \le C_t \int_0^t \int_{\mathbb{R}} q_{2(t-s)}(x-y)ds = C_t \int_0^t \int_{\mathbb{R}} q_{2s}(x-y)ds \tag{22.1}$$

When $|x-y| \gg 1$, it is easy to see the r.h.s. is small, which means we have fast decorrelation.

The fast decorrelation actually implies the convergence of the variance. Let us state the following lemma:

Lemma 22.2. *If U is a stationary random field with zero mean and an integrable covariance function $R(\cdot)$, then we have*

$$\text{Var}[\frac{1}{\sqrt{N}}\int_0^N U(x)dx] \to \int_{\mathbb{R}} R(x)dx$$

Proof. We compute

$$\text{Var}\frac{1}{\sqrt{N}}\int_0^N U(x)dx = \mathbb{E}(\frac{1}{\sqrt{N}}\int_0^N U(x)dx)^2 = \frac{1}{N}\int_0^N\int_0^N \mathbb{E}U(x)U(y)dxdy$$

Denote the covariance function of U by R, so we have

$$\frac{1}{N}\int_0^N\int_0^N R(x-y)dxdy = 2\frac{1}{N}\int_0^N dx\left(\int_0^x R(y)dy\right) = 2\int_0^1 dx\left(\int_0^{Nx} R(y)dy\right)$$

which converges to $2\int_0^\infty R(x)dx = \int_{\mathbb{R}} R(x)dx$. □

By the above lemma, to show the convergence of the variance we only need to show $U(x)$ has an integrable covariance function. In the case of a linear equation, this was done in (23.5). In the nonlinear case, one can do a similar calculation.

Define $X_N = \frac{1}{\sqrt{N}}\int_0^N (U(x)-1)dx$. We divide the proof of the theorem into two steps

(i) Show $\text{Var}X_N \to \sigma^2 \neq 0$; (ii) Show $\frac{X_N}{\sqrt{\text{Var}X_N}} \Rightarrow N(0,1)$.

Step (i) is almost done (except for proving the limit is nonzero). For step (ii) we use Stein's method.

Remark 22.3. Another maybe more traditional way of proving the Gaussianity is to modify $U(x)$ in a way that $\int_0^N U(x)dx$ is indeed sum of independent random variables to which we can apply the standard CLT and the error induced by the modification can be controlled. This appears to be more complicated than Stein's method.

22.2. **Review of Stein's method.** Let us first realize that

$$X_N = \frac{1}{\sqrt{N}}\int_0^N (U(x)-1)dx = \frac{1}{\sqrt{N}}\int_0^t \int_{\mathbb{R}} (\int_0^N q_{t-s}(x-y)dx)\sigma(u(s,y))\xi(s,y)dyds$$

where we applied Fubini theorem. Therefore, our interested random variable is written as an Ito integral

$$Y_N := \frac{X_N}{\sqrt{\text{Var}X_N}} = \int_0^\infty \int_{\mathbb{R}} f_N(s,y)\xi(s,y)dyds$$

for some adapted random field f_N which is explicit:

$$f_N(s,y) = \frac{1}{\sqrt{N\text{Var}X_N}} 1_{[0,t]}(s)(\int_0^N q_{t-s}(x-y)dx)\sigma(u(s,y)). \tag{22.2}$$

The idea of Stein's method is to characterize the Gaussian distribution through an integration by parts. Suppose for nice function ϕ (say Lipchitz 1), we can show

$$|\mathbb{E}Y_N\phi(Y_N) - \mathbb{E}\phi'(Y_N)| \to 0$$

as $N\to\infty$, then Stein's equation tells us that $Y_N \Rightarrow N(0,1)$.

From now on, fix the test function ϕ, we use the fact that Y_N is written as an Ito integral, which is the adjoint of the Malliavin derivative in this case: we write $Y_N = \delta(f_N)$ where $\delta(\cdot)$ is the adjoint operator, so

$$\mathbb{E}Y_N\phi(Y_N) = \mathbb{E}\delta(f_N)\phi(Y_N) = \mathbb{E}\langle f_N, D\phi(Y_N)\rangle = \mathbb{E}[\phi'(Y_N)\langle f_N, DY_N\rangle]$$

Here the bracket $\langle\cdot,\cdot\rangle$ is the inner product in $L^2(\mathbb{R}_+\times\mathbb{R})$. Thus,

$$|\mathbb{E}Y_N\phi(Y_N) - \mathbb{E}\phi'(Y_N)| = |\mathbb{E}\phi'(Y_N)[\langle f_N, DY_N\rangle - 1]| \le \mathbb{E}|\langle f_N, DY_N\rangle - 1|$$

where we used the fact that $|\phi'|\le 1$. We also see that $\mathbb{E}\langle f_N, DY_N\rangle = 1$ by choosing $\phi(x)=x$. An application of Holder inequality gives

$$|\mathbb{E}Y_N\phi(Y_N) - \mathbb{E}\phi'(Y_N)| \le \sqrt{\text{Var}\langle f_N, DY_N\rangle}$$

so we only need to show the r.h.s. goes to zero as $N\to\infty$.

By the definition of Y_N, we have

$$D_{s,y}Y_N = \frac{1}{\sqrt{\text{Var}X_N}} D_{s,y}X_N = \frac{1}{\sqrt{\text{Var}X_N}} D_{s,y}\frac{1}{\sqrt{N}}\int_0^N (u(t,x)-1)dx$$
$$= \frac{1}{\sqrt{\text{Var}X_N}}\frac{1}{\sqrt{N}}\int_0^N D_{s,y}u(t,x)dx$$

so we can write

$$\langle f_N, DY_N\rangle = \int_0^t \int_{\mathbb{R}} f_N(s,y) D_{s,y} Y_N dyds$$
$$= \frac{1}{\sqrt{\mathrm{Var} X_N}} \frac{1}{\sqrt{N}} \int_0^N \int_0^t \int_{\mathbb{R}} f_N(s,y) D_{s,y} u(t,x) dydsdx$$

with f_N given in (22.2). It remains to show that

$$\mathrm{Var}\langle f_N, DY_N\rangle \to 0 \tag{22.3}$$

as $N \to \infty$. We will not give the full detail of proving (22.3) here. An essential ingredient in the proof is the following estimates on the Malliavin derivative

Lemma 22.4. *There exists $C = C(t,p)$ such that*

$$(\mathbb{E}|D_{s,y} u(t,x)|^p)^{1/p} \le C q_{t-s}(x-y) \tag{22.4}$$

22.3. **Malliavin derivative estimates.** Now we sketch a proof of Lemma 22.4. Using the mild formulation

$$u(t,x) = 1 + \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)\sigma(u(s,y))\xi(s,y) dyds$$

we have

$$D_{s,y} u(t,x) = q_{t-s}(x-y)\sigma(u(s,y)) + \int_s^t \int_{\mathbb{R}} q_{t-\ell}(x-z)\sigma'(u(\ell,z)) D_{s,y} u(\ell,z)\xi(\ell,z) dz d\ell$$

Let $\|\cdot\|_p$ denote the $L^p(\Omega)$ norm. We write the above relation as $X = Y + Z$, so

$$\|X\|_p^2 \le 2\|Y\|_p^2 + 2\|Z\|_p^2$$

By the fact that $|\sigma(u(s,y))| \le |\sigma(0)| + |u(s,y)|$ and the moment bound on $u(s,y)$ (with $s \le t$), we have

$$\|Y\|_p^2 \le C q_{t-s}(x-y)^2$$

Let us consider $\|Z\|_p$. Note that Z is an Ito integral (not a martingale yet). Define

$$Z_v = \int_s^v \int_{\mathbb{R}} q_{t-\ell}(x-z)\sigma'(u(\ell,z)) D_{s,y} u(\ell,z)\xi(\ell,z) dz d\ell, \qquad v \in [s,t]$$

so $Z = Z_t$ and $\{Z_v\}_{v\in[s,t]}$ is a martingale. By the BDG inequality we have

$$\begin{aligned}
\|Z\|_p^2 &= \|Z_t\|_p^2 \le C\|\langle Z\rangle_t\|_{p/2} \\
&\le C\left\| \int_s^t \int_{\mathbb{R}} q_{t-\ell}(x-z)^2 \sigma'(u(\ell,z))^2 (D_{s,y} u(\ell,z))^2 dz d\ell \right\|_{p/2} \\
&\le C \int_s^t \int_{\mathbb{R}} q_{t-\ell}(x-z)^2 \|(D_{s,y} u(\ell,z))^2\|_{p/2} dz d\ell \\
&= C \int_s^t \int_{\mathbb{R}} q_{t-\ell}(x-z)^2 \|D_{s,y} u(\ell,z)\|_p^2 dz d\ell
\end{aligned}$$

where we used triangle inequality and the fact that $|\sigma'| \le 1$. To summarize, we have

$$\|D_{s,y} u(t,x)\|_p^2 \le C q_{t-s}(x-y)^2 + C\int_s^t \int_{\mathbb{R}} q_{t-\ell}(x-z)^2 \|D_{s,y} u(\ell,z)\|_p^2 dz d\ell$$

Note that by fixing s, y, we can view the above as an integral inequality for $D_{s,y}u(\cdot,\cdot)$. Iterate the above integral inequality, we obtain (22.4) (for details on the iteration see Lemma 5.1 of this paper).

Remark 22.5. We have not actually proved the fact that $M_t := \int_0^t \int_{\mathbb{R}} f(s,y)\xi(s,y)dyds$ is a martingale, for f that is adapted to the filtration and satisfies the integrability condition $\mathbb{E}\int_0^t\int_{\mathbb{R}} f(s,y)^2 dyds < \infty$. The intuition is pretty clear: for any $t_1 < t_2$, $\int_{t_1}^{t_2} f(s,y)\xi(s,y)dyds$ is f multiplying the noise increment beyond t_1, so by conditioning, we should have

$$\mathbb{E}\left[\int_{t_1}^{t_2} f(s,y)\xi(s,y)dyds\middle|\mathcal{F}_{t_1}\right] = 0$$

which implies that M is a martingale by definition. The actual proof goes through approximation by simple processes.

22.4. **Strict positivity of σ^2.** By Lemma 22.2, we know that, for each fixed $t > 0$, $\mathrm{Var}\frac{1}{\sqrt{N}}\int_0^N u(t,x)dx$ converges to

$$\sigma^2 = \int_{\mathbb{R}} R(x)dx$$

where $R(\cdot)$ is the covariance function of the stationary random field $\{u(t,x)\}_{x\in\mathbb{R}}$. In principle it is possible that $\sigma^2 = 0$: suppose $U(x) = F'(x)$ with F a stationary random field with smooth trajectory, then

$$\frac{1}{\sqrt{N}}\int_0^N U(x)dx = \frac{1}{\sqrt{N}}[F(N) - F(0)] \to 0$$

as $N \to \infty$. Thus, it is important to exclude the possible degenerate case of $\sigma = 0$ – if indeed this were true, it simply says that the central limit scaling is not the right scaling and we should rescale the interested quantity differently.

For the example we considered here, it is not hard to show $\sigma^2 > 0$ through a direct calculation: we know that

$$u(t,x) - 1 = \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y)\sigma(u(s,y))\xi(s,y)dyds$$

so

$$R(x) = \mathrm{Cov}[u(t,x), u(t,0)] = \int_0^t\int_{\mathbb{R}} q_{t-s}(x-y)q_{t-s}(-y)\mathbb{E}[\sigma(u(s,y))^2]dyds$$

and

$$\int_{\mathbb{R}} R(x)dx = \int_0^t f(s)ds$$

where $f(s) = \mathbb{E}\sigma(u(s,y))^2$ (is independent of y by stationarity). Thus, consider e.g. the case of $\sigma(x) = x$, we know that $f(s) > 0$ which implies that $\sigma^2 = \int_{\mathbb{R}} R(x)dx > 0$.

22.5. **A slightly different problem.** In Theorem 24.1, we considered the random field $\{u(t,x)\}_{x\in\mathbb{R}}$ and proved a central limit theorem for the spatial average. A related problem is to study $\{g(u(t,x))\}_{x\in\mathbb{R}}$ where g is some smooth function. The reason why this is a natural problem comes from the connection to the so-called KPZ equation. Consider the linear case

$$\partial_t u = \Delta u + u\xi, \qquad u(0,x) \equiv 1.$$

and define $h = \log u$. Note that there is a highly nontrivial result saying that $u(t,x)$ is strictly positive, which justifies taking the log and defining h (we will accept this result here, which is due to Carl Mueller). If we formally follow the chain rule, then we get the following SPDE (which is only a symbol at this stage)

$$\partial_t h = \Delta h + |\nabla h|^2 + \xi, \qquad h(0,x) = 0. \tag{22.5}$$

This is the Kardar-Parisi-Zhang equation and considered as a default model of surface growth subjected to random perturbation (think of $h(t,x)$ as the height of some interface at time t and spatial location x). One can ask the same question, do we have a central limit theorem for

$$\frac{1}{\sqrt{N}} \int_0^N [h(t,x) - \mathbb{E}h(t,x)]dx.$$

There is not much we can do besides viewing h as a functional of u. Directly making sense of (22.5) is the Fields Medal work of Martin Hairer.

The convergence of the variance is proved in the same way. We know that $\{h(t,x)\}_{x\in\mathbb{R}}$ is a stationary random field, so as long as the covariance function is integrable, the variance converges. We need to estimate

$$R(x) = \mathrm{Cov}[h(t,x), h(t,0)] = \mathrm{Cov}[\log u(t,x), \log u(t,0)]$$

The tool is the Gaussian Poincare covariance inequality:

$$|\mathrm{Cov}[\log u(t,x), \log u(t,0)]| \le \int_0^t \int_{\mathbb{R}} \|D_{s,y} \log u(t,x)\|_2 \|D_{s,y} \log u(t,0)\|_2 dyds \tag{22.6}$$

By chain rule we have

$$D_{s,y} \log u(t,x) = u(t,x)^{-1} D_{s,y} u(t,x)$$

so

$$\|D_{s,y} \log u(t,x)\|_2 \le \|u(t,x)^{-1}\|_4 \|D_{s,y} u(t,x)\|_4$$

It turns out that the negative moments of u can be controlled, i.e., $\sup_{t\in[0,T],x\in\mathbb{R}} \|u(t,x)^{-1}\|_4 \le C_T$, so we have by Lemma 22.4

$$\|D_{s,y} \log u(t,x)\|_2 \le C q_{t-s}(x-y)$$

which implies

$$|\mathrm{Cov}[\log u(t,x), \log u(t,0)]| \le C \int_0^t \int_{\mathbb{R}} q_{t-s}(x-y) q_{t-s}(0-y) dyds$$

which is integrable as a function of x. Thus we know that

$$\mathrm{Var} \frac{1}{\sqrt{N}} \int_0^N h(t,x)dx \to \sigma^2 = \int_{\mathbb{R}} R(x)dx$$

Here proving $\sigma^2 > 0$ is more complicated than the case of u. When we study u, the covariance can be written more or less explicitly by Ito isometry, but the above calculation using the Gaussian-Poincare covariance inequality only gives the upper bound. In a sense we need a lower bound here. To achieve this, we use the Clark-Ocone formula:

$$\log u(t,x) - \mathbb{E} \log u(t,x) = \int_0^t \int_{\mathbb{R}} \mathbb{E}[D_{s,y} \log u(t,x)|\mathcal{F}_s] \xi(s,y) dyds \tag{22.7}$$

which in some sense plays the role of mild formulation here. Using Ito isometry, we can write

$$\mathrm{Cov}[\log u(t,x), \log u(t,0)] = \int_0^t \int_{\mathbb{R}} \mathbb{E}\Big[\mathbb{E}[D_{s,y} \log u(t,x)|\mathcal{F}_s] \mathbb{E}[D_{s,y} \log u(t,0)|\mathcal{F}_s]\Big] dyds$$

Actually the Gaussian-Poincare covariance inequality can be directly proved from the above identity by applying Cauchy-Schwarz inequality. In the linear case, one can actually show that $D_{s,y} \log u(t,x)$ is strictly positive (we will not go into the detail here, but this is intimately connected to the free energy of directed polymer), which implies $\int_{\mathbb{R}} \mathrm{Cov}[\log u(t,x), \log u(t,0)]dx > 0$.

Next we need to show the central limit theorem. Denote $X_N = \frac{1}{\sqrt{N}} \int_0^N [h(t,x) - \mathbb{E}h(t,x)]dx$, it suffices to show that

$$\frac{X_N}{\sqrt{\mathrm{Var} X_N}} \Rightarrow N(0,1)$$

The approach we presented for u itself relies very much on the fact that $u(t,x) - 1$ is written as an Ito integral. One can try to start from (22.7) (which also gives an Ito integral on the r.h.s.) and proceed in a similar way. We do not go into the detail here, which can be found in this paper.

23. Lecture 24-25

We discuss another application of Malliavin calculus in these two lectures, i.e., the study of densities. In other words, what kind of information does the Malliavin derivative contain to determine if the random variable is absolute continuous with respect to Lebesgue measure or what kind of properties does the corresponding density have. The problem we have in mind is to estimate the density of the solution to the stochastic heat equation. We will introduce several tools and discuss some applications found in this paper.

23.1. **A density formula.** Let us start from the simple setting: $X = f(Z)$ where $Z \sim N(0,1)$. What can we say about the density of X? Clearly it is not the degree of the smoothness of f that matters: if $f = const$, then X has no density. A simple calculation leads to the density of X if it exists:

$$\mathbb{P}[X \le x] = \mathbb{P}[f(Z) \le x] = \mathbb{P}[Z \le f^{-1}(x)] = F(f^{-1}(x))$$

where we assumed that f is monotone and continuous and F here is the c.d.f. of $N(0,1)$. Taking the derivative with x we have the density equals to

$$F'(f^{-1}(x)) \frac{1}{f'(f^{-1}(x))}$$

so the derivative f' can not be zero at "too many points". The f' here is a special example of the Malliavin derivative, and, no surprisingly, what matters is the non-degeneracy of DX. A general result says that in an isonormal Gaussian space (using our usual notation, the Hilbert space is H and the inner product is $\langle \cdot, \cdot \rangle$), if $X \in D^{1,2}$ and $\|DX\|_H > 0$ a.s., then the law of X is absolutely continuous with respect to the Lebesgue measure. We will not present a proof of this result here, and the interested readers should go the Chapter 2 of Nualart's book. Under a stronger assumption, we actually have an "explicit" formula stated in the following proposition.

Remark 23.1. A very natural question is actually to consider the random vector $X = (X_1, \ldots, X_n)$ and ask when there exists a density on $\mathbb{R}^n$. In this case what matters is the Malliavin matrix: define $\gamma_{ij} = \langle DX_i, DX_j \rangle$, then the $n \times n$ matrix $\Gamma = (\gamma_{ij})_{ij}$ is the Malliavin matrix. If the matrix is invertible a.s., then the law of X is absolutely continuous with respect to the Lebesgue measure. One can consider the simplest case of linear transformation of i.i.d. Gaussian, then Γ is just the covariance matrix.

Proposition 23.2. *Let $X \in D^{1,2}$. Suppose that $\frac{DX}{\|DX\|_H^2}$ is in the domain of $\delta(\cdot)$, then X has a continuous and bounded density $p(x)$ given by*

$$p(x) = \mathbb{E}\left[1_{X>x}\delta(\frac{DX}{\|DX\|_H^2})\right]$$

Remark 23.3. On a purely formal level, we see why the above identity should be true. Fix any $x \in \mathbb{R}$, let $\phi(y) = 1_{y>x}$ so $\phi'(y) = \delta(y-x)$, then through an integration by parts we have

$$\mathbb{E}\left[1_{X>x}\delta(\frac{DX}{\|DX\|_H^2})\right] = \mathbb{E}\phi(X)\delta(\frac{DX}{\|DX\|_H^2}) = \mathbb{E}\langle D\phi(X), \frac{DX}{\|DX\|_H^2}\rangle = \mathbb{E}\phi'(X) = \mathbb{E}\delta(X-x)$$

Then the last expression is the density of X evaluated at x. As one could easily imagine, the rigorous following proof is through an approximation. (One should be careful with the notations here: $\delta(\cdot)$ could refer to either the divergence operator or the Dirac function)

Proof. First we note that the p given by the r.h.s. is bounded.

Take any smooth, non-negative, compacted supported function ψ, define $\phi(x) = \int_{-\infty}^{x}\psi(y)dy$, we know that $\phi(X) \in D^{1,2}$ and $D\phi(X) = \psi(X)DX$, so

$$\langle D\phi(X), DX\rangle = \psi(X)\|DX\|_H^2$$

Thus, we can write

$$\mathbb{E}\psi(X) = \mathbb{E}\langle D\phi(X), \frac{DX}{\|DX\|_H^2}\rangle$$

Since $DX/\|DX\|_H^2$ is in the domain of $\delta(\cdot)$ which is the adjoint of D, we have

$$\mathbb{E}\psi(X) = \mathbb{E}\left[\phi(X)\delta(\frac{DX}{\|DX\|_H^2})\right]$$

By an approximation argument, we can take $\psi(x) = 1_{[a,b]}(x)$ for any $a < b$. The above becomes

$$\mathbb{P}[a \le X \le b] = \mathbb{E}\int_{-\infty}^{X} 1_{[a,b]}(y)dy\delta(\frac{DX}{\|DX\|_H^2})$$

One can write

$$\int_{-\infty}^{X} 1_{[a,b]}(y)dy = \int_a^b 1_{X>y}dy$$

which implies (by Fubini)

$$\mathbb{P}[a \le X \le b] = \int_a^b \mathbb{E}\left[1_{X>y}\delta(\frac{DX}{\|DX\|_H^2})\right]dy$$

and completes the proof. □

Remark 23.4. Consider a simple example $X = W(h)$ so $DX = h$ and $DX/\|DX\|_H^2 = h/\|h\|_H^2$, this implies

$$p(x) = \|h\|_H^{-2}\mathbb{E}1_{W(h)>x}W(h) = \|h\|_H^{-2}\int_x^\infty yp(y)dy$$

which can be checked from integration by parts.

Remark 23.5. The above density formula is not unique. For example, take any random variable $F \in L^2(\Omega, H)$ such that $F/\langle DX, F\rangle$ belongs to the domain of $\delta(\cdot)$, we have (by exactly the same proof)

$$p(x) = \mathbb{E}\Big[1_{X>x}\delta(\frac{F}{\langle DX, F\rangle})\Big] \tag{23.1}$$

This flexibility turns out to be useful in the study of stochastic heat equation: if X is the solution to SHE, one can choose F to be the Ito integral term in the mild formulation – we will come back to this later.

With the density formula, an immediate question is how to obtain the upper/lower bound of $p(x)$, and probably more importantly, how to check if $DX/\|DX\|_H^2$ is in the domain of $\delta(\cdot)$ and how to rewrite $\delta(DX/\|DX\|_H^2)$ more explicitly. We knew that if there is a filtration and f is adapted, then $\delta(f)$ is the Ito integral of f with respect to the underlying Gaussian process. In the general case, this is not true, but we first recall the following lemma.

Lemma 23.6. *Suppose $X = \sum_i F_i h_i$ with F_i smooth random variables and $h_i \in H$, then*

$$\delta(X) = \sum_i F_i W(h_i) - \sum_i \langle DF_i, h_i\rangle$$

Proof. Consider a smooth random variable G and use the fact that $\mathbb{E}\langle D(F_iG), h_i\rangle = \mathbb{E}F_iGW(h_i)$ and the product rule to complete the proof. □

The same integration by parts proof leads to

$$\delta(Fh) = F\delta(h) - \langle DF, h\rangle \tag{23.2}$$

for more general F, h (where F is $\mathbb{R}-$valued and h is $H-$valued, and we will not go into the detailed assumptions on F, h here): actually for smooth random variable G, $\mathbb{E}\langle D(FG), h\rangle = \mathbb{E}FG\delta(h)$, and since $D(FG) = FDG + GDF$, we have

$$\mathbb{E}\langle DG, Fh\rangle = \mathbb{E}F\langle DG, h\rangle = \mathbb{E}G[F\delta(h) - \langle DF, h\rangle]$$

which implies $\delta(Fh) = F\delta(h) - \langle DF, h\rangle$. Therefore, under certain integrability assumptions, one has

$$\delta(\frac{DX}{\|DX\|_H^2}) = \|DX\|_H^{-2}\delta(DX) - \langle D\|DX\|_H^{-2}, DX\rangle. \tag{23.3}$$

One can further apply chain rule to simplify the expression of the second term

$$D\|DX\|_H^{-2} = -\|DX\|_H^{-4}D\langle DX, DX\rangle = -2\|DX\|_H^{-4}\langle D^2X, DX\rangle_H$$

then

$$\langle D\|DX\|_H^{-2}, DX\rangle = -2\|DX\|_H^{-4}\langle D^2X, DX \otimes DX\rangle_{H\otimes H}$$

Note that when there is no confusion we just wrote $\langle\cdot,\cdot\rangle_H = \langle\cdot,\cdot\rangle$. When there is, we will stress which Hilbert space we take inner product in.

23.2. **Meyer's estimate.** The above lemma indicates that in the general setting, if we want to estimate $\delta(X)$, the upper bound may involve both X and DX. There is a general theorem saying that if $u \in D^{1,p}(H)$ for some $p > 1$, then

$$\begin{aligned} \|\delta(u)\|_p &\le C_p\Big(\|u\|_{L^p(\Omega,H)} + \|Du\|_{L^p(\Omega,H\otimes H)}\Big) \\ &\le C_p\Big(\|\mathbb{E}u\|_H + \|Du\|_{L^p(\Omega,H\otimes H)}\Big) \end{aligned} \tag{23.4}$$

Note that $u \in D^{1,p}(H)$ means that u is an $H-$valued r.v. such that $\mathbb{E}\|u\|_H^p + \mathbb{E}\|Du\|_{H\otimes H}^p < \infty$, so the first "$\le$" in (23.4) says that δ is bounded from $D^{1,p}(H)$ to $L^p(\Omega)$. The proof of (23.4) is complicated and we will not go into the detail here. We will use it in the following section to estimate the density obtained in Proposition 23.2.

23.3. **Upper bound on the density.** Using the Meyer's estimate, we can proceed as follows.

For the upper bound, we pick some p, q to be determined and apply Holder's inequality to obtain

$$\begin{aligned} p(x) = \mathbb{E}1_{X>x}\delta(DX/\|DX\|_H^2) \le& (\mathbb{E}1_{X>x})^{1/q}\|\delta(DX/\|DX\|_H^2)\|_p \\ =& \mathbb{P}(X>x)^{1/q}\|\delta(DX/\|DX\|_H^2)\|_p \end{aligned}$$

For the $p-$norm of the divergence, we apply (23.4) to derive

$$\|\delta(DX/\|DX\|_H^2)\|_p \le C\Big(\|\mathbb{E}(DX/\|DX\|_H^2)\|_H + \|D(DX/\|DX\|_H^2)\|_{L^p(\Omega,H\otimes H)}\Big)$$

For the first term, apply triangle inequality we have

$$\|\mathbb{E}(DX/\|DX\|_H^2)\|_H \le \mathbb{E}\|DX\|_H^{-1}$$

For the second term, by a similar calculation as before, we have

$$D(DX/\|DX\|_H^2) = D^2X/\|DX\|_H^2 - 2\|DX\|_H^{-4}DX\langle D^2X, DX\rangle_H$$

which implies

$$\|D(DX/\|DX\|_H^2)\|_{H\otimes H} \le \|DX\|_H^{-2}\|D^2X\|_{H\otimes H} + 2\|DX\|_H^{-4}(*)$$

Here $(*)$ is the $H \otimes H$ norm of $DX\langle D^2X, DX\rangle_H$. Let us consider a specific example to see what the upper bound should be: assume $H = L^2(\mathbb{R}_+)$ so we can write

$$DX\langle D^2X, DX\rangle_H = (DX\langle D^2X, DX\rangle_H)_{s_1,s_2} = D_{s_1}X\int_0^\infty D_{s_2,s_3}XD_{s_3}Xds_3$$

so we have

$$\begin{aligned} (*)^2 &= \int_0^\infty\int_0^\infty\left(D_{s_1}X\int_0^\infty D_{s_2,s_3}XD_{s_3}Xds_3\right)^2 ds_1ds_2 \\ &= \int_{[0,\infty)^4}(D_{s_1}X)^2D_{s_2,s_3}XD_{s_2,s_4}XD_{s_3}XD_{s_4}Xds_1ds_2ds_3ds_4 \\ &= \|DX\|_H^2\int_{[0,\infty)^3}D_{s_2,s_3}XD_{s_2,s_4}XD_{s_3}XD_{s_4}Xds_2ds_3ds_4 \\ &\le \|DX\|_H^2\|DX\|_H^2\int_0^\infty\left(\int_0^\infty(D_{s_2,s_3}X)^2ds_3\right)ds_2 = \|DX\|_H^4\|D^2X\|_{H\otimes H}^2 \end{aligned}$$

Here in the last $\le$, we applied Cauchy-Schwarz $|\int_0^\infty D_{s_2,s_3}XD_{s_3}Xds_3| \le \|DX\|_H\sqrt{\int_0^\infty(D_{s_2,s_3}X)^2ds_3}$.

To summarize, we have

$$\|D(DX/\|DX\|_H^2)\|_{H\otimes H} \leq 3\|DX\|_H^{-2}\|D^2X\|_{H\otimes H}$$

which implies

$$\|\delta(DX/\|DX\|_H^2)\|_p \leq C\mathbb{E}\|DX\|_H^{-1} + C\big(\mathbb{E}\|DX\|_H^{-2p}\|D^2X\|_{H\otimes H}^p\big)^{1/p}$$

For any exponents $\alpha, \beta > 0$ such that $\alpha^{-1} + \beta^{-1} = p^{-1}$, we apply Holder again to bound the second term

$$\big(\mathbb{E}\|DX\|_H^{-2p}\|D^2X\|_{H\otimes H}^p\big)^{1/p} \leq (\mathbb{E}\|DX\|_H^{-2\alpha})^{1/\alpha}(\mathbb{E}\|D^2X\|_{H\otimes H}^{\beta})^{1/\beta}$$

This leads to the following theorem on the upper bound of the density

Theorem 23.7. *Let $q, \alpha, \beta > 0$ satisfy $1/q + 1/\alpha + 1/\beta = 1$, and $X \in D^{2,\beta}$ with $\mathbb{E}\|DX\|_H^{-2\alpha} < \infty$, then the density of X satifies*

$$p(x) \leq C\mathbb{P}(X > x)^{1/q}\bigg(\mathbb{E}\|DX\|_H^{-1} + (\mathbb{E}\|DX\|_H^{-2\alpha})^{1/\alpha}(\mathbb{E}\|D^2X\|_{H\otimes H}^{\beta})^{1/\beta}\bigg)$$

23.4. **Lower bound on the density.** *For the lower bound, it is more like a case by case analysis. Starting from the identity*

$$p(x) = \mathbb{E}1_{X>x}\delta(DX/\|DX\|_H^2)$$

or, more generally, the identity

$$p(x) = \mathbb{E}\bigg[1_{X>x}\delta(\frac{F}{\langle DX, F\rangle})\bigg]$$

and using (23.2), we can write

$$\delta(\frac{F}{\langle DX, F\rangle}) = \frac{\delta(F)}{\langle DX, F\rangle} + \langle DX, F\rangle^{-2}\langle D\langle DX, F\rangle, F\rangle.$$

Then one needs to study the two terms separately.

23.5. **Application to SPDE.** For the SHE $\partial_t u = \frac{1}{2}\Delta u + u\xi(t,x)$, started from $u(0,x) = 1$. By stationarity we know that the distribution of $u(t,x)$ does not depend on x. Here is a result on the upper tail of the density:

Theorem 23.8. *For each $t \in (0,T]$, $u(t,0)$ has a density which we denote by $\rho(t,\cdot)$, and we have*

$$\frac{c_1}{t^{\beta/2}}e^{-\frac{c_2}{t^\beta}(\log(c_3 y))^\gamma} \leq \rho(t,y) \leq \frac{c_4}{t^{\beta/2}}e^{-\frac{c_5}{t^\beta}(\log(c_6 y))^\gamma}, \qquad \text{for } y \gg 1 \tag{23.5}$$

Here the constants c_i depend on T, and $\gamma = \frac{3}{2}$, $\beta = \frac{1}{2}$.

To get the bounds on the density, we first need bounds on the probability. In the following, we will first explain how to obtain the upper and lower bound for the probability $\mathbb{P}[u(t,0) > y]$ for $y \gg 1$, then we will use the tools developed in the previous sections to derive the estimates on the density.

Estimates on the probability. The estimates on the probability requires bounds on the moments. Define $\kappa_p = \mathbb{E}u(t,0)^p$, assume that we have sharp upper and lower bounds on κ_p for any $p > 1$. (This should be obtained from other properties of the stochastic heat equation)

The upper bound is always achieved by Chebyshev: for any $p > 0$, we have

$$\mathbb{P}[u(t,0) > y] \leq y^{-p}\mathbb{E}u(t,0)^p = e^{\log\kappa_p - p\log y} \tag{23.6}$$

Suppose we have an appropriate upper bound of κ_p, one can choose p (depending on y) to try to minimize the exponent $\log \kappa_p - p \log y$.

For the lower bound, we use the Paley-Zygmund inequality:

Lemma 23.9. *If $X \geq 0$ has a finite variance, then for any $\theta \in (0,1)$, we have (denote $\mu = \mathbb{E}X$)*

$$\mathbb{P}[X \geq \theta\mu] \geq (1-\theta)^2 \frac{\mu^2}{\mathbb{E}X^2}$$

Proof. It is based on Cauchy-Schwarz:

$$(\mathbb{E}X 1_{X>\theta\mu})^2 \leq \mathbb{E}X^2 \mathbb{P}[X > \theta\mu]$$

We also have $\mathbb{E}X1_{X>\theta\mu} + \mathbb{E}X1_{X\leq\theta\mu} = \mu$, which implies

$$\mathbb{E}X1_{X>\theta\mu} \geq \mu - \theta\mu$$

The proof is complete. □

Using the above lemma, one can try to derive a lower bound for the probability $\mathbb{P}[u(t,0) > y]$. Note that since we are interested in the regime of $y \gg 1$ and we know that $\mathbb{E}u(t,0) \equiv 1$, one can not just simply apply the Paley-Zygmund to $u(t,0)$ itself: instead, suppose one has κ_p large when p is large, and for $y \gg 1$, one can choose p so that $\frac{1}{2}\kappa_p^{1/p} > y$, we have

$$\mathbb{P}[u(t,0) > y] \geq \mathbb{P}[u(t,0) \geq \frac{1}{2}\kappa_p^{1/p}] = \mathbb{P}[u(t,0)^p \geq \frac{1}{2^p}\kappa_p] \geq (1-\frac{1}{2^p})^2 \frac{\kappa_p^2}{\kappa_{2p}} \tag{23.7}$$

Combine (23.6) and (23.7), we obtain upper and lower bounds on the tail probability for $y \gg 1$.

Estimates on the density. Denote the density of $u(t,0)$ by $\rho_t(\cdot)$, which is supported on $[0,\infty)$ (Since u is strictly positive). For the upper bound, we apply Theorem 24.1 to derive that

$$\rho_t(y) \leq C\mathbb{P}(u(t,0) > y)^{1/q} \Big(\mathbb{E}\|Du(t,0)\|_H^{-1} + (\mathbb{E}\|Du(t,0)\|_H^{-2\alpha})^{1/\alpha} (\mathbb{E}\|D^2u(t,0)\|_{H\otimes H}^{\beta})^{1/\beta} \Big)$$

and we just use the upper tail probability bound (23.6) in the above estimate.

For the lower bound, it is more complicated. We will use the more general identity (23.1):

$$\rho_t(y) = \mathbb{E}\left[1_{u(t,0)>y} \delta\Big(\frac{F}{\langle Du(t,0), F\rangle} \Big) \right]$$

Here F is any process so that $F/\langle Du(t,0), F\rangle$ is in the domain of $\delta(\cdot)$. We know that

$$\delta\Big(\frac{F}{\langle Du(t,0),F\rangle}\Big) = \langle Du(t,0), F\rangle^{-1}\delta(F) - \langle D\langle u(t,0), F\rangle^{-1}, F\rangle$$

For the moment, let us just consider the first term. Recall that

$$u(t,0) = 1 + \int_0^t \int_{\mathbb{R}} q_{t-s}(0-y)u(s,y)\xi(s,y)dyds$$

The trick here is to pick $F(s,y) = q_{t-s}(0-y)u(s,y)1_{[0,t]}(s)$, in which case we know that

$$\delta(F) = \int_0^t \int_{\mathbb{R}} q_{t-s}(0-y)u(s,y)\xi(s,y)dyds = u(t,0) - 1$$

Recall that we have an "integral" equation satisfied by $D_{s,y}u(t,0)$, so one can write $\langle Du(t,0), F\rangle$ explicitly as

$$\langle Du(t,0), F\rangle = \int_0^t \int_{\mathbb{R}} D_{s,y}u(t,0)F(s,y)dyds =: A_t$$

In this way we have

$$\langle Du(t,0), F\rangle^{-1}\delta(F) = \frac{u(t,0)-1}{A_t}$$

and the first term in the density formula is

$$\mathbb{E}1_{u(t,0)>y}\langle Du(t,0), F\rangle^{-1}\delta(F) = \mathbb{E}\left[1_{u(t,0)>y}\frac{u(t,0)-1}{A_t}\right]$$

Since $y \gg 1$, we know that under the event of $u(t,0) > y$, we must have $u(t,0)-1 > 1$, so we have

$$\mathbb{E}1_{u(t,0)>y}\langle Du(t,0), F\rangle^{-1}\delta(F) \geq \mathbb{E}\left[1_{u(t,0)>y}\frac{1}{A_t}\right]$$

Apply Cauchy-Schwarz, we have

$$\mathbb{P}[u(t,0) > y] = \mathbb{E}[\frac{1_{u(t,0)>y}}{\sqrt{A_t}}\sqrt{A_t}] \leq \sqrt{\mathbb{E}\frac{1_{u(t,0)>y}}{A_t}\mathbb{E}A_t}$$

which implies

$$\mathbb{E}[1_{u(t,0)>y}\langle Du(t,0), F\rangle^{-1}\delta(F)] \geq \mathbb{P}[u(t,0) > y]^2\frac{1}{\mathbb{E}A_t}$$

and then one can apply the lower bound on the tail probability obtained in (23.7). We will not go into the detail here. The argument can be found in Section 6 of the paper here.

24. Lecture 26-27

The following is based on this paper, and one can also find relevant discussions in Davar's note Section 4.3.

24.1. Another density formula.

In these two lectures, we will introduce another density formula and study its applications. Recall that from the last two lectures, we learnt that if $X \in D^{1,2}$ is such that $DX/\|DX\|^2$ is in the domain of $\delta(\cdot)$, then X has a bounded, continuous density given by

$$p(y) = \mathbb{E}[1_{X>y}\delta(DX/\|DX\|^2)]$$

Sometimes it would be nice to have a formula without any reference to the divergence operator.

Assume the general setting an isonormal Gaussian process, with the underlying Hilbert space H and the inner product $\langle\cdot,\cdot\rangle$. For a given real-valued random variable $Z \in D^{1,2}$, define the function $G : \mathbb{R} \to \mathbb{R}$

$$G(x) = \mathbb{E}[\langle DZ, (1-L)^{-1}DZ\rangle | Z = x] \tag{24.1}$$

where L is the OU operator.

It is clear that if $Z \in D^{1,2}$ we have

$$\mathcal{X}_Z := \langle DZ, (1-L)^{-1}DZ\rangle \in L^1,$$

so the conditional expectation $\mathbb{E}[\mathcal{X}_Z|Z]$ is well-defined and is a *deterministic* function of Z, and we used G to represent this deterministic function. Note that $G(\cdot)$ is only defined on the support of

Z. For convenience we extend the domain of definition of G so that it is ∞ outside the support, and this does not affect the identity (24.1).

The density formula is given by the following theorem

Theorem 24.1. *Assume $\mathbb{E}Z = 0$. The law of Z has a density ρ if and only if the random variable $G(Z)$ is strictly positive almost surely. In this case, the support of ρ is a closed interval of $\mathbb{R}$ containing zero and we have, for almost all $x \in \text{supp}\rho$,*

$$\rho(x) = \frac{\mathbb{E}|Z|}{2G(x)} e^{-\int_0^x yG(y)^{-1}dy} \tag{24.2}$$

The above theorem says that (i) the non-degeneracy of the random variable $G(Z)$ is equivalent with the existence of the density (ii) there is an explicit formula for the density in this case. One should compare the random variable $G(Z)$ with $\|DZ\|^2 = \langle DZ, DZ\rangle$. We will show that $G(Z) \geq 0$ a.s. in the Lemma 24.5 below. So in a sense both $G(Z)$ and $\|DZ\|^2$ measures the nondegeneracy of DZ.

Remark 24.2. Consider the trivial case of $Z = \xi_1$, it gives the density for the standard Gaussian.

Remark 24.3. If there is a constant upper and lower bound for $G(Z)$, then we get upper/lower bound for $\rho(\cdot)$ in terms of the Gaussian density.

24.2. **Review of the covariance representation.** Let us first review a covariance representation we learnt some time ago. Assuming we have a sequence of i.i.d. $N(0,1)$, denoted by $\{\xi_i\}_{i\geq 1}$,(so that the underlying Hilbert space is ℓ^2), and for any $Z \in D^{1,2}$ we can write $DZ = (D_iZ)_{i\geq 1}$. Let A_i be the adjoint of D_i, so

$$A_i = \xi_i - D_i,$$

where the ξ_i above is interpreted as the multiplicative operator. For smooth vectors $Y = (Y_1, Y_2, \ldots,)$, define

$$\delta(Y) = \sum_{i\geq 1} A_i Y_i = \sum_{i\geq 1} Y_i \xi_i - \sum_{i\geq 1} D_i Y_i$$

which is the divergence operator (and we interpret it as a type of stochastic integral). The OU operator is defined as

$$L = -\delta D = -\sum_{i\geq 1} A_i D_i = \sum_{i\geq 1} (D_i^2 - \xi_i D_i)$$

and the heat semigroup is formally written as $P_t = e^{tL}$.

The following covariance representation holds: for any $X, Y \in D^{1,2}$, we have

$$\text{Cov}[X,Y] = \int_0^\infty e^{-t}\mathbb{E}\langle DX, P_t DY\rangle dt = \mathbb{E}\langle DX, (1-L)^{-1}DY\rangle$$

Take $Y = X$ we have

$$\text{Var}X = \int_0^\infty e^{-t}\mathbb{E}\langle DX, P_t DY\rangle dt = \mathbb{E}\langle DX, (1-L)^{-1}DY\rangle$$

Recall that in the proof of the above covariance/variance formula, we used the OU semigroup to interpolate two independent Gaussians and that is how the OU operator/semigroup shows up.

It turns out the above formula holds in the very general setting. In principle, all Gaussian processes can be constructed from $\{\xi_i\}_{i\geq 1}$, so we only need to use the above version, where the derivative is with respect to each ξ_i and the inner product is in ℓ^2. However, sometimes it is more convenient to compute the Malliavin derivative in other "coordinates". We state it as the following proposition

Proposition 24.4. *Let $X, Y \in D^{1,2}$, then*

$$\mathrm{Cov}[X,Y] = \int_0^\infty e^{-t}\mathbb{E}\langle DX, P_t DY\rangle dt = \mathbb{E}\langle DX, (1-L)^{-1}DY\rangle$$

where $L = -\delta D$ is the OU operator and $P_t = e^{tL}$ is the OU semigroup. In particular, one can apply Mehler's formula to rewrite

$$P_t DY(\omega) = \mathbb{E}[DY(e^{-t}\omega + \sqrt{1-e^{-2t}}\omega')|\omega]$$

With above proposition, we see that $\mathbb{E}\mathcal{X}_Z = \mathrm{Var}Z$.

24.3. **Derivation of the density formula.** We first have the following lemma

Lemma 24.5. *$G(Z) \geq 0$ a.s.*

Proof. Take any $g \geq 0$ continuous, compactly supported, and define $f(x) = \int_0^x g(y)dy$, so we have

$$\mathbb{E}Zf(Z) = \mathbb{E}g(Z)\langle DZ, (1-L)^{-1}DZ\rangle = \mathbb{E}g(Z)G(Z)$$

Note that when $x \geq 0$, $f(x) \geq 0$ and when $x \leq 0$, we have $f(x) \leq 0$. In other words, $xf(x) \geq 0$ for all $x \in \mathbb{R}$. Thus, we have

$$\mathbb{E}g(Z)G(Z) \geq 0, \qquad \text{for all } g \in C_c(\mathbb{R}, \mathbb{R}_+)$$

This implies that $G(Z) \geq 0$ a.s. □

Let us present a few lemmas.

Lemma 24.6. *(i) If $\langle DZ, DZ\rangle > 0$ a.s., then Z has a density*

(ii) If $G(Z) = \mathbb{E}[\langle DZ, (1-L)^{-1}DZ\rangle|Z] > 0$ a.s., then Z has a density.

Proof. The proofs of (i) and (ii) are very similar. Take any bounded Borel set B, define $F(x) = \int_{-\infty}^x 1_B(y)dy$ which is a bounded Lipschitz function. Take a sequence of bounded continuous functions f_n such that $f_n \to 1_B$ a.e. with respect to the Lebesgue measure and also a.e. with respect to μ, where μ is the measure on $\mathbb{R}$ induced by Z. Define $F_n(x) = \int_\infty^x f_n(y)dy$. By chain rule we have

$$DF_n(Z) = f_n(Z)DZ$$

From now on we assume B has zero Lebesgue measure so $F \equiv 0$.

(i) Since $F_n(Z) \to 0$ and $DF_n(Z) \to 1_B(Z)DZ$, we have $0 = 1_B(Z)DZ$ hence $0 = 1_B(Z)\langle DZ, DZ\rangle$, which implies that $\mathbb{P}(Z \in B) = 0$ and completes the proof.

(ii) We have $\mathbb{E}ZF_n(Z) = \mathbb{E}f_n(Z)\langle DZ, (1-L)^{-1}DZ\rangle = \mathbb{E}f_n(Z)G(Z)$. Sending $n \to \infty$, we have $0 = \mathbb{E}1_B(Z)G(Z)$ and since $G(Z) > 0$ a.s., we have $\mathbb{P}(Z \in B) = 0$, which completes the proof. □

Lemma 24.7. *If $Z \in D^{1,2}$, then the support of Z is connected.*

Proof. Suppose the support of Z intersects both $(-\infty, a]$ and $[b, \infty)$ with some $a < b$, and does not intersect (a, b). Pick $\varepsilon > 0$ so that $a + 2\varepsilon < b$ and define function f_ε so that $f_\varepsilon(x) = 1$ when $x \leq a + \varepsilon$, $f_\varepsilon(x) = 0$ when $x \geq a + 2\varepsilon$, and is linear in $[a+\varepsilon, a+2\varepsilon]$. We have $Df_\varepsilon(Z) = f'_\varepsilon(Z)DZ$, and since f'_ε is only nonzero when $x \in [a+\varepsilon, a+2\varepsilon]$, we have $Df_\varepsilon(Z) = 0$, thus $f_\varepsilon(Z) = \mathbb{E}f_\varepsilon(Z)$ is a constant (by Gaussian-Poincare inequality). But this is a contradiction because with a positive probability $f_\varepsilon(Z) = 1$ and with a positive probability $f_\varepsilon(Z) = 0$. □

Proof of Theorem 24.1. Now we can complete the proof of Theorem 24.1.

(i) Let us first show that if Z has a density, then $G(Z)$ is positive almost surely. Take any f continuous with compact support and define $F(x) = \int_{-\infty}^x f(y)dy$ (so F is bounded), we have

$$\begin{aligned}\mathbb{E}f(Z)G(Z) = \mathbb{E}F(Z)Z &= \int F(y)y\rho(y)dy \\ &= \int f(y)\left(\int_y^\infty w\rho(w)dw\right)dy = \int f(y)\rho(y)\frac{\left(\int_y^\infty w\rho(w)dw\right)}{\rho(y)}dy \\ &= \mathbb{E}f(Z)\frac{\left(\int_Z^\infty w\rho(w)dw\right)}{\rho(Z)}\end{aligned} \tag{24.3}$$

Here we used integration by parts, and the fact that $\int_y^\infty z\rho(z)dz \to 0$ as $|y| \to \infty$ (because $\mathbb{E}Z = 0$ and $Z \in L^1$). Also note that $\mathbb{E}1_{\rho(Z)=0} = \int 1_{\rho(z)=0}\rho(z)dz = 0$, so $\rho(Z) > 0$ a.s. (and the last expression makes sense).

Since $Z \in D^{1,2}$, by Lemma 24.7, we know that the support of ρ is $[\alpha, \beta]$, but since $\mathbb{E}Z = 0$, we necessarily have $\alpha < 0, \beta > 0$. Consider the function

$$\phi(z) = \int_z^\infty w\rho(w)dw, \qquad z \in [\alpha, \beta].$$

We claim that ϕ is strictly increasing in $z \in [\alpha, 0]$ and strictly decreasing in $[0, \beta]$, which can be seen from the sign of the derivative and the fact that the support of ρ is the whole $[\alpha, \beta]$. Since $\mathbb{E}Z = 0$, we have $\phi(\alpha) = \int_\alpha^\infty w\rho(w)dw = 0$, and we also have that $\phi(\beta) = \int_\beta^\infty w\rho(w)dw = 0$, thus, $\phi(z) > 0$ when $z \in (\alpha, \beta)$.

By (24.3), since f is arbitrary, we know that

$$G(Z) = \frac{\left(\int_Z^\infty w\rho(w)dw\right)}{\rho(Z)} = \frac{\phi(Z)}{\rho(Z)}, \qquad a.s. \tag{24.4}$$

thus $G(Z) > 0$ a.s.

(ii) Now let us prove the density formula. First, suppose that $\rho(w) > 0$ a.e. in $[\alpha, \beta]$ (note that the support of Z being $[\alpha, \beta]$ does not necessarily imply $\rho(w) > 0$ a.e. in $[\alpha, \beta]$), then by (24.4), we have

$$G(w) = \phi(w)/\rho(w), \qquad \text{a.e. in } [\alpha, \beta]$$

On the other hand, by the expression of ϕ, we have

$$\phi'(w) = -w\rho(w), \qquad \text{a.e. in } [\alpha, \beta]$$

The above two equations implies

$$\frac{\phi'(w)}{\phi(w)} = -\frac{w}{G(w)}, \qquad \text{a.e. in } [\alpha, \beta]$$

so we integrate to obtain

$$\log\phi(x) - \log\phi(0) = -\int_0^x \frac{w}{G(w)}dw$$

so

$$\phi(x) = \phi(0)e^{-\int_0^x w/G(w)dw}$$

On the other hand, we know that $\mathbb{E}Z = 0$ and $\phi(0) = \int_0^\infty w\rho(w)dw = \frac{1}{2}\mathbb{E}|Z|$. This implies that

$$\rho(w) = \frac{\phi(w)}{G(w)} = \frac{\mathbb{E}|Z|}{2G(w)} e^{-\int_0^x w/G(w)dw}$$

which completes the proof.

In the "strange" case when $\{w : \rho(w) = 0\}$ has a positive Lebesgue measure. We still define $G(w) = \phi(w)/\rho(w)$, which takes value ∞ on the set of $\rho(w) = 0$. In this case we still have $\phi'(w)/\phi(w) = -w/G(w)$ a.e. on $[\alpha, \beta]$ (note that on the set of $\rho(w) = 0$, we have $\phi'(w) = 0$ a.e.) The rest of the proof is the same. $\square$