Contents

1 Stochastic heat equation with additive noise ........................... 2
  1.1 Explicit solution ................................................................... 2
  1.2 Properties of solution ......................................................... 5
  Besov regularity. Wavelets and Hölder regularity. Gaussian free field as invariant measure.
  1.3 Examples and challenges of nonlinear SPDEs .......................... 10

2 Stochastic heat equation with multiplicative noise ................. 12
  2.1 Itô solution ....................................................................... 13
  Itô integral. Existence and uniqueness of solution.
  2.2 Weak solution ..................................................................... 15

3 $\Phi^4$ model ........................................................................... 17
  3.1 $\Phi^4$ Euclidean quantum field theory ................................... 17
  3.2 Stochastic quantization and local mild solution ..................... 21
  Fixed point argument.
  3.3 $\Phi^4$ equation in $d = 2$ ..................................................... 23
  3.4 $\Phi^4$ equation in $d = 3$ ..................................................... 24

4 (Controlled) rough paths ....................................................... 26

5 Regularity structures ............................................................. 29
  5.1 Basic definitions in regularity structure theory ....................... 29
  5.2 Reconstruction theorem ....................................................... 30
  5.3 Example: controlled rough paths ......................................... 33

6 Sine-Gordon ............................................................................ 35
  6.1 Background ........................................................................ 35
  6.2 Dynamical sine-Gordon ....................................................... 37

7 Yang-Mills ............................................................................. 41
1 Stochastic heat equation with additive noise

1.1 Explicit solution

From the discussion of stochastic heat equation (SHE) – simple yet a prototype of Stochastic PDE (SPDE) – we can appreciate what problems, properties and tools we can explore for SPDEs in general. SHE with additive space-time white noise $\xi$ reads

$$\partial_t u = \Delta u + \xi.$$  

**White noise.** We call $\eta$ a random distribution if it is a continuous linear map $f \mapsto \eta(f)$ from $C_c^\infty$ (the space of compactly supported smooth test functions) into the space of square integrable random variables on some fixed probability space $(\Omega, \mathbb{P})$.

A white noise $\zeta$ on $D \subset \mathbb{R}^d$ is a random distribution, for which $\{\zeta(f)\}_{f \in C_c^\infty}$ is a collection of centered joint Gaussians on a fixed probability space $(\Omega, \mathbb{P})$ with covariance given by (“$L^2$ property”)

$$E[\zeta(f)\zeta(g)] = \int_D f(x)g(x)dx.$$  

This is often formally written as $E[\zeta(x)\zeta(y)] = \delta(x - y)$ where $\delta$ is the Dirac distribution. A space-time white noise $\xi$ is a white noise on a space-time domain $\Lambda \subset \mathbb{R} \times \mathbb{R}^d$ with

$$E[\xi(f)\xi(g)] = \int_{\Lambda} f(t, x)g(t, x)dtdx,$$  

or $E[\xi(t, x)\xi(s, y)] = \delta(t - s)\delta(x - y).$  

One example of white noise on $\mathbb{R}_+$ is $\frac{dB}{d\tau}$, the derivative of Brownian motion; in this case the above “$L^2$ property” is just the Itô isometry.

**Cylindrical Wiener process.** The space-time white noise can be also constructed as the time derivative of the so called “cylindrical Wiener process”, an infinite dimensional generalization of Brownian motion. We briefly describe this construction. For a separable Hilbert space $H$ (e.g. $L^2(\mathbb{T}^d)$ if we are interested in white noise on $\mathbb{T}^d$), let $(e_i)_{i \in \mathbb{N}}$ be an orthogonal basis. We then define

$$W(t) = \sum_{k \in \mathbb{N}} b_k(t)e_k$$  

where $\{b_k\}_{k \in \mathbb{N}}$ are independent standard Brownian motions. One has

$$E[\langle W(t), f \rangle_H \langle W(t), g \rangle_H] = \sum_{k, \ell} E[b_k(t)b_\ell(s)\langle e_k, f \rangle_H \langle e_\ell, g \rangle_H = (t \wedge s) \sum_k \langle e_k, f \rangle_H \langle e_k, g \rangle_H$$

$$= (t \wedge s) \langle f, g \rangle_H.$$
So \( W \) has the same covariance as 1D Brownian motion, but has the “\( L^2 \) property” in spatial direction. A same calculation with \( W \) replaced by its time derivative yields \( \delta(t-s) \langle f, g \rangle_H \).

The subtlety in this construction is that \( W \) is not an element in \( H \) when \( H \) is infinite dimensional, as can be seen from \( \mathbf{E}\{W(t), W(t)\}_H = t \sum_k \langle \epsilon_k, \epsilon_k \rangle = \infty \). In fact \( W \) can be constructed as a process in a larger Hilbert space \( H' \supset H \) via for instance \( \langle f, g \rangle_{H'} = \sum_k \frac{1}{k^2} \langle f, g \rangle_H \), so that \( \mathbf{E}\{W(t), W(t)\}_{H'} = t \sum_k \frac{1}{k^2} \langle \epsilon_k, \epsilon_k \rangle_H < \infty \). (In general, it suffices to have \( \iota : H \leftrightarrow H' \) Hilbert-Schmidt i.e. \( \iota^* \) trace class; and \( H' \) is the Cameron-Martin space for \( H' \) with a Gaussian measure of covariance \( \iota^* \) in the language of Malliavin calculus.)

**Scaling.** Scaling or “dimension analysis” will be important in our course. For \( \delta \) on \( \mathbb{R}^d \), one can check (by testing against \( f \in C_c^\infty \)) that \( \lambda^d \delta(\lambda x) = \delta(x) \). Heuristically, this means that its “scaling dimension” \( [\delta] = [-x]^{-d} \). According to this heuristic, the space-time white noise \( \xi \) then has “scaling dimension” \( [\xi] = [t]^{-\frac{d}{2}} [x]^{-\frac{d}{2}} \). In fact, one has

\[
\lambda^{k_0+k_1} \xi(\lambda^{k_0} t, \lambda^{k_1} x) \overset{\text{law}}{=} \xi(t, x) \quad (k_0, k_1) \in \mathbb{N}^2, \lambda \in \mathbb{R}_+ ,
\]

which can be shown by testing against \( f \).

The equation (1.1) also has a scaling invariance. For invariance, the variables \( t \) and \( x \) have to be scaled diffusively (i.e. parabolically) \( (t, x) \to (\lambda^2 t, \lambda x) \). A heuristic dimension counting shows that if the three terms in (1.1) have the same scaling dimension \( [u_t] = [\Delta u] = [\xi] = [t]^{-\frac{d}{2}} [x]^{-\frac{d}{2}} = [x]^{-\frac{d+2}{2}} \), then \( [u] = [x]^{-\frac{d+2}{2}} \). More precisely, if

\[
\tilde{u}(t, x) \overset{def}{=} \lambda^{-\frac{d+2}{2}} u(\lambda^2 t, \lambda x)
\]

then it satisfies \( \partial_t \tilde{u}(t, x) = \Delta \tilde{u}(t, x) + \tilde{\xi}(t, x) \) where \( \tilde{\xi}(t, x) := \lambda^{\frac{d+2}{2}} \xi(\lambda^2 t, \lambda x) \overset{\text{law}}{=} \xi(t, x) \).

**Remark 1.1** Finding the “scaling dimensions” for various objects here is useful for the discussion of their approximations. For example, \( \varepsilon^{-d} \mathbf{1}_{|x| \leq \varepsilon/2} \) is the right approximation for \( \delta \) as \( \varepsilon \to 0 \), as can be checked by testing against \( f \in C_c^\infty \). For a white noise \( \xi \) in \( \mathbb{R}^d \), a central limit theorem can be shown: namely for a smooth, not necessarily Gaussian, mean zero random field \( \zeta \) with bounded moments, \( \varepsilon^{-\frac{d}{2}} \zeta(\varepsilon^{-1} x) \) converges to \( \xi \) in law.

**Solution via heat kernel.** If \( \xi \) was a function, the linear equation (1.1) with initial condition \( u(t, \cdotp) = u_0 \) (where \( u_0 \) is deterministic) has the following explicit solution

\[
u(t, x) = \int_{\mathbb{R}^d} \int_0^t P(t-s, x-y)\xi(s, y)dsdy + \int_0^t P(t, x-y)u_0(y)dy
\]

where \( P \) is the heat kernel

\[
P(t, x) = (4\pi t)^{-\frac{d}{2}} e^{-\frac{\|x\|^2}{4t}}.
\]

We call this the **mild solution** to (1.1).\(^3\) It remains to give a suitable meaning to the first term in (1.5) (this term is sometimes called “stochastic convolution”). We assume \( u_0 = 0 \) for simplicity. Then, since \( \xi \) is centered Gaussian and (1.5) is linear in \( \xi, u \) should also be a centered Gaussian distribution.

\(^3\)The mild solution is a strong solution. The notion of “strong” and “weak” solutions in SPDE is different from those in PDE. In SPDE, a strong solution assigns each realization of \( \xi \) a function \( u \) as in (1.5), whereas a weak solution solves a process that has the required law, which we discuss later.
The subtlety is the singularity of $P$ at $(t, x) = 0$, so that $P(t \cdot, x \cdot)$ is not necessarily $L^2$. A simple calculation using (1.2) shows
\[
\mathbb{E}[u(t, x)u(\bar{t}, \bar{x})] = \int_{\mathbb{R}^d} \int_0^t P(t - s, x - y)P(\bar{t} - s, \bar{x} - y)dsdy .
\]
In particular the “variance”
\[
\mathbb{E}[u(t, x)^2] = \int_{\mathbb{R}^d} \int_0^t P(t - s, x - y)^2dsdy = \int_0^t \frac{1}{(8\pi(t - s))^\frac{d}{2}} \int_{\mathbb{R}^d} e^{-\frac{|x - y|^2}{2(t - s)}}dyds
\]

\[
= \int_0^t (8\pi(t - s))^{-\frac{d}{2}}ds < \infty \quad \text{if and only if} \quad d = 1 .
\]
This could also follow\(^4\) by Fourier transform (in $x$) and Parseval’s theorem:
\[
\int_0^t \left| \hat{P}(t-s, k) \right|^2 ds = \int_0^t e^{-2(t-s)|k|^2} ds = \frac{1 - e^{-2t|k|^2}}{2|k|^2}
\]
which is integrable as $|k| \to \infty$ if and only if $d = 1$.

This means that $u$ should not have a pointwise value when $d > 1$. On the other hand $u$ is a bona fide random distribution; in fact $u(f)$ has variance $\| \int_{\mathbb{R}^{d+1}} P(t \cdot, x \cdot) f(t, x)dt dx \|_{L^2}$ which is always finite for $f \in C^\infty_c$. Below we describe the regularity of $u$ as a distribution, but we first give another way of solving the equation using Fourier transform.

**Solution via Fourier transform.** For a Fourier analysis, assuming for simplicity that our underlying space is the torus $\mathbb{T}^d$. Recall that for any real function $u = \sum_{k \in \mathbb{Z}^d} \hat{u}(k) e^{ikx}$, the Fourier coefficients satisfy $\hat{u}(k) = \overline{u(-k)}$. The cylindrical Wiener process is given by $W(t) = \sum_{k \in \mathbb{Z}^d} \beta_k(t)e^{ikx}$ where, since $e^{ikx}$ are complex, $\{\beta_k\}_{k \in \mathbb{Z}^d}$ are independent complex valued Brownian motions (whose real and imaginary parts are independent, each being a standard Brownian motion divided by $\sqrt{2}$), with $\beta_k = \overline{\beta_{-k}}$. Applying (spatial) Fourier transform to (1.1), we get $\partial_t \hat{u} = -k^2 \hat{u} + \xi$, and this is actually a system of decoupled\(^5\) SDEs
\[
d\hat{u}(k) = -k^2 \hat{u}(k)dt + d\beta_k . \quad (1.7)
\]
The solution is now given by
\[
\hat{u}(t, k) = \int_0^t e^{-(t-s)|k|^2}d\beta_k(s) + e^{-t|k|^2} \hat{u}_0(k) .
\]
Focusing on the random term by assuming $u_0 = 0$, then $\hat{u}$ is centered Gaussian by Gaussianity of $u$ and linearity of Fourier transform. Its covariance is given by (for $k_1, k_2 \neq 0$)
\[
\mathbb{E}[\hat{u}(t, k_1)\hat{u}(t, -k_2)] = \mathbb{1}_{k_1 = k_2} \int_0^t e^{-2(t-s)|k_1|^2} ds = \mathbb{1}_{k_1 = k_2} \int_0^t e^{-2(t-s)|k_1|^2} ds = \mathbb{1}_{k_1 = k_2} \frac{1 - e^{-2|k_1|^2}}{2|k_1|^2} \quad (1.8)
\]
where we have used Itô isometry.

---

\(^4\)This could be also “guessed” by dimension counting: $[P] = [t]^{-d/2} = [x]^{-d}$, so $[P^2] \cdot [t] \cdot [x]^d > 0$ if and only if $d = 1$.

\(^5\)decoupled except for the constraint $\hat{u}(k) = \overline{u(-k)}$.
1.2 Properties of solution

**Besov Space regularity.** Let \( \chi, \varrho \in C_c^\infty \) be nonnegative radial functions on \( \mathbb{R}^d \), such that \( \text{supp} \chi \) is contained in a ball and \( \text{supp} \varrho \) is contained in an annulus, satisfying “partition of unity” \( \chi(z) + \sum_{j \geq 0} \varrho(2^{-j} z) = 1 \) for all \( z \in \mathbb{R}^d \), with \( \text{supp}(\chi) \cap \text{supp}(\varrho(2^{-j})) = \emptyset \) for \( j \geq 1 \) and \( \text{supp}(\varrho(2^{-i})) \cap \text{supp}(\varrho(2^{-j})) = \emptyset \) for \( |i - j| \geq 2 \). See \cite{[1]} (Proposition 2.10) for existence of such functions. We will write \( \varrho_{-1} = \chi \) and \( \varrho_j = \varrho(2^{-j} \cdot) \) for \( j \geq 0 \).

The Littlewood–Paley blocks are now defined as

\[
\Delta_j u = \mathcal{F}^{-1}(\varrho_j \mathcal{F} u) \quad \text{for } j \geq -1.
\]

Then one has \( u = \sum_{j \geq -1} \Delta_j u \). For \( p, q \in [1, \infty) \) we define

\[
B_{p,q}^\alpha = \left\{ u \in S'(\mathbb{R}^d) : \|u\|_{B_{p,q}^\alpha} = \left( \sum_{j \geq -1} (2^{j\alpha} \|\Delta_j u\|_{L^p})^q \right)^{1/q} < \infty \right\},
\]

with the usual interpretation as \( \ell^\infty \) norm in case \( q = \infty \). The space \( B_{p,q}^\alpha \) does not depend on \( (\chi, \varrho) \). We write \( C^\alpha = B_{\infty,\infty}^\alpha \). For \( \alpha \in \mathbb{R} \setminus \mathbb{Z} \), it can be shown that \( C^\alpha \) are identical to the Hölder spaces. We write \( u_t \overset{\text{def}}{=} u(t, \cdot) \).

**Lemma 1.2** Let \( \gamma = -\frac{d - 2}{2} \). For any \( \varepsilon > 0, \delta \in (0, 1), p \in \mathbb{N} \) we have

\[
E\|u_t - u_s\|_{B_{\delta-\varepsilon}^{\alpha-p}}^p \leq C|t - s|^{\delta p/2}.
\]

From this, together with Kolmogorov continuity theorem, and a continuous Besov imbedding \( B_{p,p}^{\alpha} \hookrightarrow B_{\infty,\infty}^{\alpha - \frac{d}{2}} \), by taking \( p \) large enough we have\(^6\)

\[
u \in C([0, T], B_{\infty,\infty}^{\alpha - \frac{d}{2}}) \quad \text{a.s.} \quad \forall \varepsilon > 0 \tag{1.9}
\]

To illustrate the main idea, we assume that the underlying space is the torus \( \mathbb{T}^d \). We will only prove a simpler version of Lemma 1.2

\[
E\|u(t, \cdot)\|_{B_{p,p}^{\alpha}}^p \leq C \quad \forall \alpha < \gamma. \tag{1.10}
\]

**Proof.** By definition, \( \|u\|_{B_{p,p}^{\alpha}}^p = \sum_{j \geq -1} 2^{j\alpha p} \|\Delta_j u\|_{L^p}^p \). Note that if we started by trying to bound moments of \( B_{\infty,\infty}^{\alpha} \) it would be inconvenient to deal with expectation of supremum; however now with \( p < \infty \) we only need to compute

\[
E\|\Delta_j u_t\|_{L^p}^p = E \int_{\mathbb{T}^d} \left| \sum_{k \in \mathbb{Z}^d} \varrho_j(k) \hat{u}_t(k) e^{ikx} \right|^p dx.
\]

Since \( u \) (and thus \( \hat{u} \) since Fourier transform is linear) is Gaussian, the above is bounded by

\[
\leq C \int_{\mathbb{T}^d} \left[ \left( \sum_{k \in \mathbb{Z}^d} \varrho_j(k) \hat{u}_t(k) e^{ikx} \right)^2 \right]^\frac{p}{2} dx.
\]

\(^6\)more precisely, we can find a version of \( u \) which is continuous in \( t \)
Now everything boils down to a second moment calculation of Gaussian: by (1.8)
\[
E \left| \sum_{k \in \mathbb{Z}^d} \varrho_j(k) \hat{u}(k)e^{ikx} \right|^2 = \sum_{k_1, k_2 \in \mathbb{Z}^d} \varrho_j(k_1) \varrho_j(k_2) E[\hat{u}(k_1)\hat{u}(-k_2)]e^{i(k_1-k_2)x} \\
= \sum_{k \in \mathbb{Z}^d} \varrho_j(k)^2 \left( 1 - \frac{e^{-2|k|^2}}{2|k|^2} \right) \asymp 2^{-2j} 2^{jd} = 2^{j(d-2)}
\]
where \( \asymp \) denotes ‘bounded above and below up to proportional constants’, since \( \varrho_j \) is supported on an annulus of width \( 2^j \) (thus it contains \( \sim 2^j \) terms). Here \( 1 - e^{-2|k|^2} \) is asymptotically \( 1 \) as \( k \to \infty \). The summability then requires
\[
\sum_{j \geq -1} 2^{j\alpha r} E \| \Delta_j u \|^2 \|_{L^p} \asymp \sum_{j \geq -1} 2^{j\alpha r} (2^{j(d-2)})^\frac{p}{2} < \infty \quad \iff \quad \alpha < \gamma = -\frac{d-2}{2}
\]
as required by (1.10). \( \square \)

**Hölder regularity.** In (1.9) we view \( u \) as a process in time taking values in a Besov space. Here we explore the other viewpoint, which is, viewing \( u \) as a random distribution over space-time, and we measure regularity of \( u \) in real space (rather than Fourier). We again assume that the underlying space is \( \mathbb{T}^d \). Write \( \Lambda = [0, T] \times \mathbb{T}^d \). We first introduce some notation. For \( z \in \Lambda \) we define a parabolic distance \( \| z \| = \sqrt{t} + |x| \). For \( \varphi \in C^\alpha_r(\Lambda) \) with some \( r > 0 \), and \( \lambda \in (0, 1) \), we define
\[
\varphi_\lambda^{s,y}(t, x) \defeq \lambda^{-(d+2)} \varphi(\lambda^{-2}(t-s), \lambda^{-1}(x-y)) \tag{1.11}
\]

namely \( \varphi \) is re-centered to \((s, y)\) and parabolically rescaled by \( \lambda \).

For \( \alpha \in \mathbb{R}_+ \setminus \mathbb{Z}_+ \), let \( C^\alpha_r \) be the completion of \( C^\infty_r \) under
\[
\| f \|_{C^\alpha_r} = \sum_{|k| < |\alpha|} \sup_{z \in \Lambda} |D^k f(z)| + \sup_{z, \tilde{z} \in \Lambda} \frac{|D^k f(z) - D^k f(\tilde{z})|}{\|z - \tilde{z}\|^{\alpha - |\alpha|}}
\]
where \( |k| = 2k_0 + k_1 \) for \((k_0, k_1) \in \mathbb{N} \times \mathbb{N}^d \). When \( \alpha < 0 \), the space \( C^\alpha_r \) is defined as the completion of \( C^\infty_r \) with respect to
\[
\| f \|_{C^\alpha_r} = \sup_{\lambda \in (0, 1)} \sup_{z \in \Lambda} \lambda^{-|\alpha|} |f(\varphi_\lambda^{s,y}(t, x))| \tag{1.12}
\]
where \( \sup_\varphi \) is over all functions \( \varphi \) which have \( \| \varphi \|_{C^\alpha_r} \leq 1 \) for \( r_0 = -|\alpha| \) and supported in a unit ball. For \( \alpha < 0 \) it can be shown that \( C^\alpha_r \) is essentially equivalent to the Besov space \( B^\alpha_{\infty, \infty} \), but with respect to the parabolic distance over space-time.

For the space-time white noise \( \xi \), a simple second moment calculation using (1.2) shows that
\[
E[\xi(\varphi_\lambda^{s,y})^2] = \int_\Lambda (\varphi_\lambda^{s,y}(w))^2 dw \lesssim \lambda^{-2(d+2)} \int_{\|w - z\| \leq \lambda} dw = \lambda^{-2(d+2)} \lambda^{d+2} = \lambda^{-(d+2)} \tag{1.13}
\]
where we applied a brutal bound on \( \varphi_\lambda^{s,y}(w) \) using (1.11), and \( \lesssim \) stands for \( \leq \) up to a proportional constant that is uniform in \( \lambda, z, \varphi \). This calculation is consistent with the scaling dimension discussed around (1.4).
Our goal here is to prove that \( \xi \in C^\alpha_s \) for any \( \alpha < -\frac{d+2}{2} \), by showing \( E\|\xi\|_{C^\alpha_s}^p < \infty \), similarly as we’ve done for Besov space in (1.10). The challenge is to take expectation of the supremum over infinitely (uncountably) many \( \lambda, z, \varphi \) on RHS of (1.12). The theory of wavelets allows us to simply deal with countably many of them, and thereby replace the supremum by sum, making it easier to take expectation. Here’s a quick tour to wavelets.

Wavelets. Let \( \Lambda_n \overset{df}{=} ((2^{-2n}Z) \times (2^{-n}Z)^d) \cap \Lambda \). Given \( \varphi \) and \( n \in Z_+ \) we write
\[
\varphi^n_{(s,y)}(t,x) \overset{df}{=} 2^{(d+2)n/2} \varphi(2^n(t-s), 2^n(x-y)).
\]
Note that the difference between the notation \( \varphi^\lambda \) and \( \varphi^n \) is that \( \|\varphi^\lambda\|_{L^1} \) stays constant as \( \lambda \to 0 \) whereas \( \|\varphi^n\|_{L^2} \) stays constant as \( n \to \infty \). Here is an important theorem in wavelets.

**Theorem 1.3** Fix \( r > 0 \). There exist \( \varphi \in C^r_c \) and a finite collection \( \Psi = \{ \psi \} \) of \( C^r_c \) functions on \( \Lambda \), such that \( \{ \varphi^n_x \}_{z \in \Lambda_0} \cup \{ \psi^n_x \}_{n \in Z_+, z \in \Lambda_n} \) form an orthonormal basis of \( L^2(\Lambda) \).

We skip the proof of this theorem, but only briefly explain the idea behind it, for \( L^2(\mathbb{R}) \) instead of \( L^2(\Lambda) \) for simplicity. Daubechies proved that given \( r > 0 \), there exists \( \varphi \in C^r_c(\mathbb{R}) \) such that
\begin{enumerate}
  \item For each \( k \in Z \), \( \int_\mathbb{R} \varphi(x)\varphi(x+k)dx = 1_{k=0} \);
  \item there exist “structure constants” \( a_k \) such that \( \varphi(x) = \sum_{k \in Z} a_k \varphi(2x + k) \).
\end{enumerate}

In view of property 1. we define \( V_n = \text{span}\{ \varphi^n_x : x \in \Lambda_n \} \subset L^2(\mathbb{R}) \). Property 2. then shows that \( V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_n \subset V_{n+1} \subset \cdots \). Writing \( V_{n+1} = V_n \oplus V_n \), it turns out that there exists \( \psi(x) = \sum_{k \in Z} b_k \varphi(x + k) \in C^r_c \) for some constants \( b_k \) such that \( V_n = \text{span}\{ \varphi^n_x : x \in \Lambda_n \} \). Therefore, we have an \( L^2 \) decomposition \( L^2(\mathbb{R}) = V_0 \oplus V_0 \oplus V_1 \oplus V_2 \oplus \cdots \), with an orthonormal basis.

The above construction can be extended to multi-dimensions, and in particular we have \( \varphi, \psi \) defined on space-time \( \mathbb{R}^{d+1} \); by rescaling as in (1.14) we get the orthonormal basis \( \{ \varphi^n_x \}_{z \in \Lambda_0} \cup \{ \psi^n_x \}_{n \in Z_+, z \in \Lambda_n} \) as in Theorem 1.3.

We can characterize \( C^\alpha_s \) for \( \alpha < 0 \) by the above wavelet basis, in the same spirit as we define distribution spaces using Fourier coefficients once we have the \( L^2 \) Fourier basis.

**Theorem 1.4** Let \( \alpha < 0 \). \( f \in C^\alpha_s \) if and only if
\[
|f(\varphi^n_x)| \lesssim 2^{-\frac{d+2n}{2} - \alpha} \quad |f(\varphi^n_y)| \lesssim 1
\]
where the proportional constants in \( \lesssim \) are uniform in \( n \in Z_+, z \in \Lambda_n, y \in \Lambda_0, \psi \in \Psi \).

**Proof.** We only prove necessity (sufficiency is harder). Since \( |f(\varphi^n_x)| \lesssim \lambda^n \) for any \( \lambda \in (0,1), z \in \Lambda \) and \( \varphi \in C^r_c \) with unit support, in particular it holds for the \( \varphi \) and \( \psi \) in Theorem 1.3, for \( \lambda = 2^{-n} \), and \( z \in \Lambda_n \). Noting the difference between the definitions of \( \varphi^\lambda \) and \( \varphi^n \), with \( \lambda = 2^{-n} \) one has the following identity to translate between the two notation
\[
f(\varphi^\lambda_x) = 2^{-\frac{d+2n}{2}} f(\varphi^n_x)
\]
from which (1.15) follows. We refer to [Hat14, Proposition 3.20] for a complete proof.

---

\footnote{It is easy to find a discontinuous one, e.g. the Haar wavelet, but finding one in \( C^r_c \) is very nontrivial.}
Thanks to Theorem \[1.4\] we can now bound \( E\|\xi\|_{C^2_p}^p \):
\[
E\|\xi\|_{C^2_p}^p = E \left[ \sup_{\psi} \sup_{n \geq 0} \sup_{z \in \Lambda_n} \left( 2^{(d+2)n} + n \alpha \right) E(\psi^p) \right] \lesssim \sum_{n \geq 0} \sum_{z \in \Lambda_n} 2^{(d+2)n} + n \alpha E(\psi^p) 
\]
(1.17)
where \( \psi \) ranges in the finite collection \( \Psi \cup \{ \varphi \} \) given in Theorem \[1.3\]. This is finite if and only if \( \alpha < -\frac{d+2}{2} - \frac{d+2}{2} \). Therefore, given \( \alpha = -\frac{d+2}{2} - \epsilon \) with \( \epsilon > 0 \), one can choose \( p \) large enough so that \( -\epsilon < -\frac{d+2}{p} \), which ensures \( E\|\xi\|_{C^2_p}^p < \infty \). Here, an important input of the bound is \( E(\xi(\psi^p) \psi^2) \lesssim E(\xi(\psi^p)^2) \frac{2}{\lambda} = \|\psi\|_{L^2} \lesssim 1 \), which is essentially the same calculation as (1.13). In future, it will be more convenient to do this second moment calculation in the way as in (1.13), because one can just read off the regularity from the exponent of \( \lambda \) from the right hand side (and it is equivalent anyway, in view of (1.16)). The above estimate holds with more generality and let’s write it as a lemma:

**Lemma 1.5** If \( \zeta \) is a Gaussian random distribution over space-time, such that \( E(\zeta(\varphi^h)^2) \lesssim \lambda \) for some \( \alpha < 0 \) then \( \zeta \in C^a_p \) for any \( \alpha < \alpha \).

**Proof.** With \( \lambda = 2^{-n} \) we have \( E(\xi(\psi^p)) \lesssim E(\xi(\psi^p)^2)^{n/2} \lesssim (2^{-(d+2)n} - 2^{-2n})^{p/2} \), so in (1.17) the summability condition reads \( n(d+2) + \frac{(d+2)np}{2} + n \alpha - \frac{(d+2)np}{2} - n \alpha p < 0 \), which is \( \alpha < \alpha - (d+2)/p \). We then choose \( p \) large as above.

---

**Gaussian free field as invariant measure.** For each \( k \neq 0 \), (1.7) is an Ornstein-Uhlenbeck process, which has the (1-dimensional complex) Gaussian measure \( \mathcal{N}(0, 1|k|) \) as an invariant measure, namely \( e^{-\frac{1}{2}|k|^2|\hat{a}(k)|^2}/Z \) (where \( Z \) is the suitable normalization).

The Gaussian free field \( \Phi \) on \( T^d \) is a random distribution, with \( \hat{\Phi}(0) = 0 \) (namely \( \Phi(1) = \int_{\mathcal{T}^d} \Phi = 0 \)) which is given by
\[
\Phi = \sum_{k \in Z^d \setminus \{0\}} \frac{a_k}{|k|} e^{ik \cdot x}
\]
(1.18)
where \( \{a_k\} \) are independent complex standard Gaussians, except for a constraint \( a_k = \bar{a}_{-k} \), namely Re(\( a_k \)), Im(\( a_k \)) \( \sim \mathcal{N}(0, \frac{1}{2}) \) independent s.t. \( E|a_k|^2 = 1 \). Since
\[
E|\hat{\Phi}(k)|^2 = E|a_k|^2 / |k|^2 = 1 / |k|^2
\]

---

8 A student in the class asked that given \( \alpha = -\frac{d+2}{2} - \epsilon \) with a small \( \epsilon > 0 \), why it happens that \( E\|\xi\|_{C^2_p}^p < \infty \) for a sufficiently large \( p \) while the bound (1.17) would be \( \infty \) for a small \( p' \) (say \( p' = 1 \)), which seems to contradict with the general fact \( \|X\|_{L^{p'}} \leq \|X\|_{L^p} \) for \( p < p' \). The explanation is that when \( E\|\xi\|_{C^2_p}^p < \infty \) for a sufficiently large \( p \) we do have \( E\|\xi\|_{C^2_p}^p < \infty \) for any \( p' \leq p \), but this can’t be seen from (1.17). The bound (1.17) is only an upper bound, where we replaced \( \sup_{z \in \Lambda_n} \) by \( \sum_{z \in \Lambda_n} \), causing a factor \( 2^{(d+2)n} \); if we did not have this factor \( 2^{(d+2)n} \), we would be able to conclude that \( \sum_{n \geq 0} 2^{(d+2)n} + n \alpha p < \infty \) if only if \( \alpha < -\frac{d+2}{2} \) (for any \( p \)). This “loss of sharpness” of course does not matter at all, since we only need to find some \( p \) so that \( E\|\xi\|_{C^2_p}^p < \infty \).

9 The invariance of this Gaussian measure for the SDE \( dX = -|k|^2 X dt + dB \) can be checked, for instance by Kolmogorov forward equation which states that the probability density \( p(t, x) \) for \( X_t \) satisfies \( \partial_t p = \frac{d}{2} p - \partial_x (-|k|^2 x p) \), and \( p(x) = e^{-\frac{1}{2}|k|^2|x|^2}/Z \) indeed satisfies this PDE (both sides vanish), namely \( X_t \) can have density \( p(x) \) for all \( t \).
the Gaussian free field is invariant for the dynamic (1.7). Alternatively, one can view the Gaussian free field as a centered Gaussian random distribution with “covariance” \((-\Delta)^{-1}\) in the following sense: for any \(f, g \in C_c^\infty(T^d)\) with \(\int_{T^d} f = \int_{T^d} g = 0\),

\[
\mathbb{E}[\Phi(f)\Phi(g)] = \mathbb{E} \sum_{k \neq 0} \hat{\Phi}(k) \hat{f}(k) \sum_{\ell \neq 0} \hat{\Phi}(\ell) \hat{f}(\ell) = \sum_{k \neq 0} \frac{\hat{f}(k) \hat{g}(-k)}{|k|^2} = \langle f, (-\Delta)^{-1} g \rangle_{L^2(T^d)}.
\]

This leads to the formal notation of Gaussian free field in some literature as a centered Gaussian random distribution with “covariance” \((-\Delta)^{-1}\) in the following sense: for any \(f, g \in C_c^\infty(T^d)\) with \(\int_{T^d} f = \int_{T^d} g = 0\),

\[
\mathbb{E}[\Phi(f)\Phi(g)] = \mathbb{E} \sum_{k \neq 0} \hat{\Phi}(k) \hat{f}(k) \sum_{\ell \neq 0} \hat{\Phi}(\ell) \hat{f}(\ell) = \sum_{k \neq 0} \frac{\hat{f}(k) \hat{g}(-k)}{|k|^2} = \langle f, (-\Delta)^{-1} g \rangle_{L^2(T^d)}.
\]

This is independent of \(z\). To show this, recall the formula (1.6) for \(P(t, x); \) when \(\sqrt{t} \geq |x|\), one has

\[
P(t, x) = (4\pi t)^{-\frac{d}{2}} e^{-\frac{|x|^2}{4t}} \lesssim |t|^{-\frac{d}{2}} e^{-\frac{|x|^2}{4t}} \lesssim |z|^{-d}
\]

and when \(\sqrt{t} \leq |x|\), one has \(P(t/x^2, 1) = \frac{|x|^d}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}} \lesssim \frac{|x|^d}{(4\pi)^{\frac{d}{2}}} \left(\frac{|x|^2}{4}\right)^{-d/2} \lesssim 1\), so

\[
P(t, x) = |x|^{-d} P(t/x^2, 1) \lesssim |x|^{-d} \lesssim \|z\|^{-d}.
\]

Another useful result is:

**Lemma 1.6** If \(f\) and \(g\) are compactly supported functions on space-time, smooth except for the origin, and \(|f(z)| \lesssim \|z\|^{\alpha}, |g(z)| \lesssim \|z\|^{\beta}\) where \(\alpha, \beta \in (-d-2, 0)\) and \(\alpha + \beta + (d+2) < 0\) then \(|(f \ast g)(z)| \lesssim \|z\|^{\alpha + \beta + (d+2)}\).

**Proof.** \((f \ast g)(z) = \int f(z-w)g(w)dw\). Let \(r \triangleq \|z\|/2\). We consider three regimes in space-time. First, \(\|w\| < r\), in which case \(\|z-w\| > C\|z\|\) for some \(C > 0\), so that

\[
\int_{\|w\| < r} |f(z-w)| |g(w)| dw \lesssim \int_{\|w\| < r} \|z\|^\alpha \|w\|^\beta dw \lesssim \|z\|^{\alpha + \beta + (d+2)}.
\]

The second regime is \(\|z-w\| < r\) and it is treated the same way as the first regime by symmetry. The third regime is the remaining regions, where one must have \(\|z-w\| \geq C\|w\|\) for some \(C > 0\) (Exercise!) so the contribution of this part to the convolution is bounded by

\[
\int_{\|w\| \in \Sigma(r, C')} \|w\|^\alpha \|\beta dw \lesssim \|z\|^{\alpha + \beta + (d+2)}.
\]

\(\square\)
We briefly review a number of physically important nonlinear equations, and discuss the well-posedness of the KPZ equation in one spatial dimension which was first solved in [Hai13]. The effect of $\partial_x^2$ becomes higher.

Kardar-Parisi-Zhang (KPZ) equation. The equation, proposed by [KPZ86], models interface growth, which is ubiquitous in nature, where each point of the interface randomly grows up or drops down over time, with a trend to locally smooth the interface out (the effect of $\partial_x^2 H$), and the growth depends in a nonlinear way in the slope (the effect of $(\partial_x H)^2$):  

$$\partial_t H = \partial_x^2 H + (\partial_x H)^2 + \xi.$$  

The equation solves KPZ until [Hai13]. The common difficulty in defining their solutions is lack of regularity; as we will see, this is related to the so-called “ultraviolet divergence” in physics. We restrict ourselves to examples that are built on top of the stochastic heat equation, and recall that its solution $u \in C^\alpha_s$ for any $\alpha < -\frac{d-2}{2}$, that is, more singular as $d$ becomes higher.

1.3 Examples and challenges of nonlinear SPDEs

We briefly review a number of physically important nonlinear equations, and discuss the challenge to define the meaning of a solution to nonlinear SPDEs driven by very singular noises. The common difficulty in defining their solutions is lack of regularity; as we will see, this is related to the so-called “ultraviolet divergence” in physics. We restrict ourselves to examples that are built on top of the stochastic heat equation, and recall that its solution $u \in C^\alpha_s$ for any $\alpha < -\frac{d-2}{2}$, that is, more singular as $d$ becomes higher.

Kardar-Parisi-Zhang (KPZ) equation. The equation, proposed by [KPZ86], models interface growth, which is ubiquitous in nature, where each point of the interface randomly grows up or drops down over time, with a trend to locally smooth the interface out (the effect of $\partial_x^2 H$), and the growth depends in a nonlinear way in the slope (the effect of $(\partial_x H)^2$):  

$$\partial_t H = \partial_x^2 H + (\partial_x H)^2 + \xi.$$  

In $d = 1$, the solution to the linear part is below Hölder $\frac{1}{2}$, and we can not expect the nonlinearity to improve regularity; thus $(\partial_x H)^2$ does not have any classical meaning. The well-posedness of the KPZ equation in one spatial dimension was first solved in [Hai13]. The problem is more severe when $d = 2$.

Stochastic heat equation with multiplicative noise (mSHE). Let $f$ be a continuous function, consider  

$$\partial_t u = \Delta u + f(u)\xi.$$  

(1.20)

The specialization $f(u) = u$, i.e.  

$$\partial_t u = \Delta u + u\xi.$$  

(1.21)

has a significant connection to the KPZ equation: one can formally check that if $H$ solves KPZ, then the Hopf-Cole transform $u := e^H$ solves (1.21). Other choices of $f$ such as $f(u) = \alpha \sqrt{u(1 - u)}$ arise in modeling population dynamics and genetics.

A classical result known as Young’s theorem states that the multiplication $f \cdot g$ can be classically defined if the sum of their Hölder regularities is positive (thinking of $\int BdB$ as a counter-example). In $d = 1$, since $u$ is expected to have Hölder regularity below $\frac{1}{2}$, and $\xi$ below $-\frac{3}{2}$, the multiplication $u\xi$ lacks classical meaning. It turns out that the Itô solution theory successful for stochastic ordinary differential equations can extend to this SPDE\footnote{so this gives a roundabout meaning to the KPZ equation, via the Hopf-Cole transform, but it was not clear in what sense $\log u$ solves KPZ until [Hai13].}.
this was summarized in for instance the lecture notes [Wal86], and we will discuss this in Section 2.

**Nonlinear parabolic Anderson model (PAM).** The equation reads

\[ \partial_t u = \Delta u + f(u)\zeta \]

where \( f \) is a continuous function and \( \zeta \) is a noise which typically is assumed to be white in space, but constant in time (i.e. "spatial white noise"). This models the motion of mass through a random media. One can prove (say, using Lemma 1.5), that \( \zeta \) has Hölder regularity below \(-\frac{d}{2}\), so when \( d \geq 2 \), one encounters the same difficulty to define its solution as for mSHE. We refer to [GIP15, HL18] and references therein for well-posedness results in \( d = 2, 3 \).

PAM (especially the linear case \( f(u) = u \)) is a simple model which exhibits intermittency over long time; for the study of long time behavior, one often considers the spatial-discrete equation with \( \zeta \) being independent noises on lattice sites, see for instance the reviews [CM94] and [Kü16] for further discussion and references regarding long time behaviors of PAM.

**Stochastic Navier-Stokes equation.** This is a vector valued equation:

\[ \partial_t \vec{u} + \vec{u} \cdot \nabla \vec{u} = \Delta \vec{u} - \nabla p + \vec{\zeta}, \quad \text{div} \vec{u} = 0 \]

where \( p \) is the pressure, \( \vec{\zeta} \) is a \( d \)-vector valued noise. For instance, when each component of \( \vec{\zeta} \) is taken as an independent space-time white noise, it models motion of fluid with randomness arising from microscopic scales, and in this case one has the same difficulty as above for \( d \geq 2 \); we refer to [DPD02, ZZ15] for well-posedness results.

**Remark 1.7** We remark that while this course focuses on singular noises, when modeling large scale random stirring of the fluid, the noise \( \vec{\zeta} \) is often assumed to be smooth (called "colored noise" in contrast with white noise), and in fact the most important case is that the equation is driven by only a few number of random Fourier modes. In these situations the long-time behavior is of primary interest, and various dynamical system questions such as ergodicity and mixing are studied. There is a vast literature on this topic, and we only refer to the book [KS12] and the survey articles [Mat03, Fla08, Kup10].

**Parisi-Wu stochastic quantization.** This refers to a large class of singular SPDEs arising from Euclidean quantum field theories defined via Hamiltonians (or actions, energy etc.). They were introduced by Parisi and Wu in [PW81]. Given a Hamiltonian \( \mathcal{H}(\Phi) \) which is a functional of \( \Phi \), one considers a gradient flow of \( \mathcal{H}(\Phi) \) perturbed by space-time white noise \( \xi \):

\[ \partial_t \Phi = -\frac{\delta \mathcal{H}(\Phi)}{\delta \Phi} + \xi. \]

Here \( \frac{\delta \mathcal{H}(\Phi)}{\delta \Phi} \) is the variational derivative of the functional \( \mathcal{H}(\Phi) \); for instance, when \( \mathcal{H}(\Phi) = \frac{1}{2} \int (\nabla \Phi)^2 dx \) is the Dirichlet form, \( \frac{\delta \mathcal{H}(\Phi)}{\delta \Phi} = -\Delta \Phi \) and (1.24) boils down to the stochastic heat equation (1.1). Note that \( \Phi \) can be also multi-component fields, with \( \xi \) being likewise multi-component. The famous \( \Phi^4 \) equation

\[ \partial_t u = \Delta u - u^3 + \xi \]
also arises from this procedure with \( \mathcal{H}(\Phi) = \int \frac{1}{2}(\nabla \Phi)^2 + \frac{1}{4}\Phi^4 dx \).

The significance of these “stochastic quantization equations” (1.24) is that given a Hamiltonian \( \mathcal{H}(\Phi) \), the formal measure

\[
\frac{1}{Z} e^{-\mathcal{H}(\Phi)} D\Phi
\]  

is formally an invariant measure\(^{11}\) for Eq. (1.24). Here \( D\Phi \) is the formal Lebesgue measure and \( Z \) is a “normalization constant”. We emphasize that (1.25) are only formal measures because, among several other reasons, there is no “Lebesgue measure” \( D\Phi \) on an infinite dimensional space and it is a priori not clear at all if the measure can be normalized. These measures arise from Euclidean quantum field theories. In their path integral formulations quantities of physical interest are defined by expectations with respect to these measures. The task of constructive quantum field theory is to give precise meaning or constructions to these formal measures, see the book \([Jaf00]\).

**Exercises**

1. Provide a complete proof of Lemma 1.2 (rather than its simplified version (1.10)).

2. Provide a complete proof to Theorem 1.4.

3. Let \( H^\alpha \) be the Sobolev (Hilbert) space. With similar arguments as in the proof for (1.9), show that \( u \in C([0, T], H^{-\frac{d+2}{2}-\varepsilon}) \) a.s. for any \( \varepsilon > 0 \).

4. Prove that the Gaussian free field given by (1.18) a.s. belongs to the Sobolev (Hilbert) space \( H^{-\frac{d+2}{2} - \varepsilon}, \) or \( B_{\infty, \infty}^{\frac{d+2}{2} - \varepsilon} \), for any \( \varepsilon > 0 \).

5. Let \( \zeta \) be the spatial white noise in (1.22), that is, \( \zeta \) is constant in time, and \( E[\zeta(x)\zeta(y)] = \delta(x - y) \). Prove using similar criteria as in Lemma 1.5 that \( \zeta \) has Hölder regularity below \(-\frac{d}{2}\).

**2 Stochastic heat equation with multiplicative noise**

\[
\partial_t Z = \Delta Z + Z \xi .
\]  

(2.1)

We call \( Z \) a mild solution if

\[
Z(t, x) = \int_{\mathbb{R}^d} \int_0^t P(t - s, x - y)Z(s, y)\xi(s, y)dsdy + \int_{\mathbb{R}^d} P(t, x - y)u_0(y)dy
\]  

(2.2)

In this, we must have that \( \int_0^t \int_{\mathbb{R}^d} P^2(t - s, x - y)E[Z^2(s, y)]dyds < \infty \) for the Itô integrals to make sense and be finite.

---

\(^ {11}\) Being invariant means that if the initial condition of (1.24) is random with “probability law” given by (1.25), then the solution at any \( t > 0 \) will likewise be distributed according to this same “probability law”. For readers familiar with stochastic ordinary differential equations, one simple example is given by the Ornstein-Uhlenbeck process \( dX_t = -\frac{1}{2} X_t dt + dB_t \) where \( B_t \) is the Brownian motion, and its invariant measure is the (one-dimensional) Gaussian measure \( \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \).
2.1 Itô solution

Itô integral. Recall from Section 1 that the space-time white noise is the time derivative of the cylindrical Wiener process, which is a Brownian motion in each 1-dimensional projection. The Itô integral w.r.t. white noise (or rather w.r.t. the cylindrical Wiener process) is defined analogously to the 1-dimensional case.

Given $\varphi \in C^\infty_c(\mathbb{R})$ we can define $\int_{\mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{[0,t]}(s)\varphi(x)\xi(x, s)dxds$, as in Section 1. This is a Brownian motion in $t$ with variance $\int \varphi^2(x)dx$, since by a similar calculation as in below (1.3), the correlation of this integral at time $t = t_1$ and $t = t_2$ equals $(t_1 \wedge t_2)^2 \int \varphi^2(x)dx$.

Let $F_0 = \emptyset$ and for each $t > 0$ define $F_t$ to be the $\sigma$–field generated by $\left\{ \int_{\mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{[0,s]}(u)\varphi(u)\xi(u, x)dxdu : 0 \leq s \leq t, \varphi \in C^\infty_c(\mathbb{R}) \right\}$. It is clear that $F_t$ is a filtration. As the next step, we consider “piece-wise constant” processes. Let $S$ be the set of functions of the form $f(t, x, \omega) = \sum_{i=1}^{n} X_i(\omega)\mathbf{1}_{(a_i, b_i]}(t)\varphi_i(x)$, where $X_i$ is a bounded $\mathcal{F}_{a_i}$–measurable random variable and $\varphi_i \in C^\infty_c(\mathbb{R})$. For $f \in S$, define

$$\int_{\mathbb{R}_+ \times \mathbb{R}} f(t, x)\xi(t, x)dxdt = \sum_{i=1}^{n} X_i \int_{\mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{(a_i, b_i]}(t)\varphi_i(x)\xi(t, x)dxdt.$$

It is easy to check that the integral is linear and satisfy Itô isometry from $L^2(\mathbb{R}_+ \times \mathbb{R} \times \Omega, \mathcal{F}, \mathbb{P})$ to $L^2(\Omega)$, that is

$$\mathbb{E} \left[ \left( \int_{\mathbb{R}_+ \times \mathbb{R}} f(t, x)\xi(t, x)dxdt \right)^2 \right] = \int_{\mathbb{R}_+ \times \mathbb{R}} \mathbb{E}[f^2(t, x)]dxdt.$$

Let $\mathcal{P}$ be the sub–$\sigma$–field of $\mathcal{B}(\mathbb{R}_+ \times \mathbb{R}) \times \mathcal{F}$ generated by $S$. Let $L^2(\mathbb{R}_+ \times \mathbb{R} \times \Omega, \mathcal{F}, \mathbb{P})$ be the space of square integrable $\mathbb{P}$–measurable random variables $f(t, x, \omega)$. These will be the integrators. It is important to note that these are non–anticipating in the sense that $f(t, x, \omega)$ only depends on the information $\mathcal{F}_t$ up to time $t$. This is analogous to the distinction between Itô and Stratonovich integrals in the one–dimensional case. The construction of the stochastic integral will be defined through the isometry and approximation.

**Lemma 2.1** $S$ is dense in $L^2(\mathbb{R}_+ \times \mathbb{R} \times \Omega, \mathcal{F}, \mathbb{P})$

*Proof.* Same as one–dimensional case. \qed

Thus, if $f \in L^2(\mathbb{R}_+ \times \mathbb{R} \times \Omega, \mathcal{F}, \mathbb{P})$ there exist $f_n \in S$ wuch that $f_n$ converges to $f$ in $f \in L^2(\mathbb{R}_+ \times \mathbb{R} \times \Omega, \mathcal{F}, \mathbb{P})$. By the isometry,

$$I_n(\omega) := \int_{\mathbb{R}_+ \times \mathbb{R}} f_n(t, x, \omega)\xi(t, x)dxdt$$
is a Cauchy sequence in $L^2(\Omega, F, \mathbb{P})$. Hence there is a limit point $I \in L^2(\Omega, F, \mathbb{P})$ which is defined to be the stochastic integral $\int_{\mathbb{R}^+ \times \mathbb{R}} f(t, x)\xi(t, x)dxdt$. This is linear in $f$ and the Itô isometry holds.

**Existence and uniqueness of mild solution.** We will work with $L^2(\Omega, F, \mathbb{P})$ bounded initial data and solutions:

**Definition 2.2** A function $z_0(\cdot)$ is $L^2(\Omega, F, \mathbb{P})$ bounded if $\sup_{x \in \mathbb{R}} E[z_0(x)^2] < \infty$ and a space-time function $Z(\cdot, \cdot)$ is $L^2(\Omega, F, \mathbb{P})$ bounded if $\sup_{x \in [0, T]} E[Z(t, x)^2] < \infty$.

**Theorem 2.3** For $L^2(\Omega, F, \mathbb{P})$ bounded initial data $z_0$ there exists a unique $L^2(\Omega, F, \mathbb{P})$ bounded mild solution to mSHE.

**Proof of uniqueness.** Assume $Z$ and $Z'$ solve SHE with the same initial data. Let $u = Z - Z'$, hence $u(0, x) \equiv 0$. Define

$$f(t, x) = E[u(t, x)^2] \quad f(t) = \sup_x f(t, x)$$

and note $f(t) < \infty$ by $L^2(\Omega, F, \mathbb{P})$ boundedness. Thus, by Itô isometry,

$$f(t) = \sup_{x \in \mathbb{R}} f(t, x) = \sup_{x \in \mathbb{R}} \int_0^t \int_{\mathbb{R}} E\left[\left(z(y, s) - z'(y, s)\right)^2\right] \cdot p^2(t - s, x - y)dyds$$

$$\leq \sup_{x \in \mathbb{R}} \int_0^t \int_{\mathbb{R}} f(s, y)p^2(t - s, x - y)dyds$$

$$\leq c \int_0^t f(s) \frac{ds}{\sqrt{t - s}}$$

for some constant $c > 0$ whose value is not important for our purposes. Note that between the second and third line we have evaluated the $x$-independent integral in $y$, thus explaining why we have dropped the supremum. Hence,

$$f(t) \leq c \int_0^t f(s) \frac{ds}{\sqrt{t - s}}.$$

Iterate to get

$$f(t) \leq c^2 \int_0^t \int_0^s f(u) \frac{duds}{\sqrt{(s - u)(t - s)}} = c^2 \int_0^t f(u) \int_u^t \frac{ds}{\sqrt{(s - u)(t - s)}} du. \quad (2.3)$$

By Exercise we can show that $f(t) \equiv 0$ hence $Z = Z'$.

**Proof of existence.** We will use Picard iteration. Let $z^0(t, x) \equiv 0$ and define progressively measurable approximations

$$z^{n+1}(t, x) = \int_{\mathbb{R}} p(t, x - y)z_0(y)dy + \int_0^t \int_{\mathbb{R}} p(t - s, x - y)z^n(s, y)\xi(s, y)dyds.$$
Then \( z^n(t, x) := z^{n+1}(t, x) - z^n(t, x) \) satisfies
\[
z^{n+1}(t, x) = \int_0^t \int_{\mathbb{R}} p(t - s, x - y)z^n(s, y)dyds.
\]
Hence, as before, via the Itô isometry
\[
E\left( (z^{n+1}(t, x))^2 \right) = \int_0^t \int_{\mathbb{R}} p^2(t - s, x - y)E\left( (z^n(s, y))^2 \right)dyds.
\]
Define
\[
f^n(t) = \sup_{s \in [0, t]} \sup_{x \in \mathbb{R}} E\left( (\bar{z}(s, x))^2 \right)
\]
and note that \( f^0(t) < \infty \) by hypothesis. Then,
\[
f^{n+1}(t) = \sup_{u \in [0, t]} \sup_{x \in \mathbb{R}} E\left( (\bar{z}^{n+1}(t, x))^2 \right) = \sup_{u \in [0, t]} \sup_{x \in \mathbb{R}} \int_0^u \int_{\mathbb{R}} p^2(u - s, x - y)f^n(s)dyds
\]
\[
\leq \sup_{u \in [0, t]} \sup_{x \in \mathbb{R}} \int_0^u \int_{\mathbb{R}} p^2(u - s, x - y)f^n(s)dyds
\]
\[
\leq \sup_{u \in [0, t]} c \int_0^u f^{n-1}(s) ds \frac{ds}{\sqrt{u - s}} \leq c \int_0^t f^{n-1}(s) ds \frac{ds}{\sqrt{t - s}}.
\]
Between the second and third lines we evaluated the \( x \)-independent integral in \( y \). The inequality in the last line can be seen by applying a change of variables so the integral is from 0 to 1, and then using the fact that \( f^{n-1}(s) \) is an increasing function.

As before we may iterate this once more and then change the order of integration. We thus find that \( f^{n+1}(t) \leq c \int_0^t f^{n-1}(u)du \). By Exercise 2 this goes to zero as \( n \to \infty \) hence proving that the \( z^n(\cdot, \cdot) \) form a Cauchy sequence in \( L^2(\Omega, \mathcal{F}, P) \). The limit point is \( z(\cdot, \cdot) \). It is clear, by convergence of stochastic integrals, that \( z \) solves the mild form of SHE.

\[\square\]

### 2.2 Weak solution

The solution as discussed in Section 1, such as the mild solution given by (1.5), is called strong solution, in the sense that we are given a probability space \( \Omega \), each realization of \( \xi(\omega) \) (\( \omega \in \Omega \)) is mapped to a solution \( u(\omega) \). Strong solutions could sometimes be inconvenient for instance when proving certain results on convergence in law to the solutions from discrete systems in probability theory or statistical physics. A weak solution only seeks for a process (which could live on a different probability space) which has the “right law”.**13**

Recall that if \( M \) is a continuous square-integrable \((E(M_t^2) < \infty \) for all \( t )) martingale then the quadratic variation \((M)_t \) is the unique non-decreasing stochastic process such that \( M_t^2 - (M)_t \) is a martingale**14**. As an example, if \( B \) is the standard Brownian motion, \( B^2 - t \) is a martingale and thus \( \langle B \rangle_t = t \). Lévy’s theorem states that if \( X \) is continuous local martingale, \( X_0 = 0, \langle X \rangle_t = t \), then \( X \) is a standard Brownian motion. As another example, given \( \varphi \in C_{c,0}(\mathbb{R}), \)
\[
\int_{\mathbb{R}^2} 1_{[0,t]}(s)\varphi(x)\xi(s, x)dxds \quad (2.4)
\]

---

**13**The notion of “weak solution” in SPDE is totally different from “weak solution” in classical PDE.

**14**There is another notion of quadratic variation usually denoted by \([M]_t\), which is in general different from \(\langle M \rangle_t\), but for continuous \( M \) they coincide.
considered in Section 2.1 is a Brownian motion with variance \( \| \varphi \|_{L^2}^2 \) (its quadratic variation is \( t \| \varphi \|_{L^2}^2 \)).

Just as the martingale characterization of the law of Brownian motion in Lévy’s theorem, a weak notion of solution to the aSHE on \( T^d \)

\[
\frac{\partial_t u}{(\Delta u + \xi)} u(0) = u_0
\]

(2.5)
can be also defined in a similar way. To this end, given \( \varphi \in C^\infty(T^d) \) writing \((u, \varphi)\) for the evaluation of the distribution \( u \in S'(T^2) \) against \( \varphi \), we have, by (2.5)

\[
(u(t), \varphi) = (u_0, \varphi) + \int_0^t (u(s), \Delta \varphi) ds + \int_0^t (\varphi, \xi(s)) ds
\]

and the last term is precisely (2.4), that is, a Brownian motion with quadratic variation \( t \| \varphi \|_{L^2}^2 \).

This motivates the following definition.

**Definition 2.4** \( v \in C(R_+, S'(T^2)) \) is called a weak solution to (2.5) if

\[
M_\varphi(t) \overset{\text{def}}{=} (v(t), \varphi) - (u_0, \varphi) - \int_0^t (v(s), \Delta \varphi) ds
\]

\[
\Gamma_\varphi(t) \overset{\text{def}}{=} M_\varphi(t)^2 - t \| \varphi \|_{L^2}^2
\]

are local martingales for any \( \varphi \in C^\infty(T^d) \).

Note that \( v \) could be defined on a different probability space. We also say that \( v \) “solves a martingale problem (2.6)”. Weak solutions can also be defined for mSHE

\[
\frac{\partial_t u}{(\Delta u + Z \xi)} u(0) = u_0
\]

(2.7)

The main modification is due to the fact that the last term integrating test functions has different quadratic variation.

**Definition 2.5** \( v \in C(R_+, S'(T^2)) \) is called a weak solution to (2.7) if \( M_\varphi(t) \) defined as in (2.6) and

\[
\Gamma_\varphi(t) \overset{\text{def}}{=} M_\varphi(t)^2 - \int_0^t (v(s)^2, \varphi^2) ds
\]

are local martingales for any \( \varphi \in C^\infty(T^d) \).

**Theorem 2.6** If \( v \) is a weak solution to aSHE (resp. mSHE), then \( v \) has the same law as the mild solution given by (1.5) (resp. Theorem 2.3).

**Exercises**

1. Show that the integral over \( s \) is equal to \( \pi \) so that \( f(t) \leq \pi c^2 \int_0^t f(u) du \). Iterate to show

\[
f(t) \leq \frac{(\pi c^2)^{n+1}}{n!} \int_0^t f(u)(t - u)^n du.
\]

Since \( f(u) < \infty \) this shows that \( f(t) \equiv 0 \) hence \( Z = Z' \).

2. Iterating and using our a priori knowledge that \( f^0, f' < \infty \) show that

\[
f^n(t) \leq \frac{(ct)^{n/2}}{(n/2)!}.
\]
3 \( \Phi^4 \) model

3.1 \( \Phi^4 \) Euclidean quantum field theory

The \( \Phi^4 \) Euclidean quantum field theory is formally given by a “measure” on the space of “functions” (or “fields”) \( \Phi \):

\[
e^{-S(\Phi)} \mathcal{D} \Phi, \quad S(\Phi) = \int \left( \frac{1}{2} |\nabla \Phi|^2 + \frac{m^2}{2} \Phi^2 + \frac{\lambda}{4!} \Phi^4 \right) d^4 x.
\]

(3.1)

For simplicity of our exposition we will assume that \( m = 0 \). If \( \lambda = 0 \), this is a Gaussian measure with covariance \((-\Delta)^{-1}\) which can be rigorously defined by classical theorems (this is of course called Gaussian free field). We thus denote by \( \mu \sim \mathcal{N}(0, (-\Delta)^{-1}) \) this Gaussian measure, and interpret (still formally) the above measure as

\[
\nu(d\Phi) = e^{-\frac{\lambda}{2} \int \Phi(x)^4 \, dx} \mu(d\Phi)
\]

**Feynman diagrams.** Before any mathematical discussion, we explain how physics would calculate physical quantities for this measure. Given such a (formal) measure, one physically interesting quantity is of course its 2-point correlation

\[
\langle \Phi(x_1)\Phi(x_2) \rangle = \mathbb{E} \left[ \Phi(x_1)\Phi(x_2) e^{\frac{\lambda}{2} \int \Phi(y)^4 \, dy} \right] / Z, \quad \text{where} \quad Z = \mathbb{E} \left[ e^{\frac{\lambda}{2} \int \Phi(y)^4 \, dy} \right]
\]

where \( \mathbb{E} \) is the expectation w.r.t. \( \mu \) and \( \langle \cdots \rangle \) is the expectation w.r.t. \( \nu \).

We know that when \( \lambda = 0 \), \( \langle \Phi(x_1)\Phi(x_2) \rangle = C(x_1 - x_2) \equiv (-\Delta)^{-1}(x_1 - x_2) \) where \( C \) is the Green’s function of \( \Delta \); in QFT \( C \) is called the free propagator. So physicists would just Taylor expand \( e^{\frac{\lambda}{2} \int \Phi(y)^4 \, dy} \) in \( \lambda \) to compute the correlation, without justifying the convergence of this expansion (everything is formal now):

\[
\langle \Phi(x_1)\Phi(x_2) \rangle = \mathbb{E} \left[ \Phi(x_1)\Phi(x_2) \left( 1 - \frac{\lambda}{4!} \int \Phi(y)^4 \, dy + \cdots \right) \right] / Z,
\]

where

\[
Z = \mathbb{E} \left[ 1 - \frac{\lambda}{4!} \int \Phi(y)^4 \, dy + \cdots \right].
\]

The advantage is that now we know how to compute the expectation \( \mathbb{E} \) of a product of Gaussian random variables - this is called the Wick theorem, which states that

\[
\mathbb{E} \left[ \Phi(x_1)\Phi(x_2)\Phi(y)^4 \right] = \mathbb{E} \left[ \Phi(x_1)\Phi(x_2)\Phi(y)\Phi(y)\Phi(y)\Phi(y) \right]
\]

is equal to a sum over all the parings of the 6 Gaussian random variables

\[
3\mathbb{E} \left[ \Phi(x_1)\Phi(x_2) \right] \mathbb{E} \left[ \Phi(y)\Phi(y) \right] \mathbb{E} \left[ \Phi(y)\Phi(y) \right] + 12\mathbb{E} \left[ \Phi(x_1)\Phi(x_2) \right] \mathbb{E} \left[ \Phi(y)\Phi(x_2) \right] \mathbb{E} \left[ \Phi(y)\Phi(y) \right] = 3C(x_1 - x_2)C(y - y)C(y - y) + 12C(x_1 - y)C(x_2 - y)C(y - y).
\]

Feynman diagram is a graphic tool to represent such terms: for each term we draw the points \( x_1, x_2, y, \) and simply draw a line for each \( C \). One then has

\[
\mathbb{E} \left[ \Phi(x_1)\Phi(x_2) \left( 1 - \frac{\lambda}{4!} \int \Phi(y)^4 \, dy + \cdots \right) \right]
\]

(3.2)
\[ \Phi^4 \text{ MODEL} \]

\[
= x_1 \frac{3\lambda}{4!} x_2 - \frac{12\lambda}{4!} x_1 y x_2 + \ldots
\]

(The reason that in physics literature one starts with a $4!$ in the definition of the model is just for convenience and to cancel various factors such as $12$ appearing here.)

Such calculation can be carried out at higher orders, e.g. at order $\lambda^2$ we need to compute

\[
E\left[ \Phi(x_1)\Phi(x_2) \left( \frac{\lambda}{4!} \int \Phi(y)^4 \, d^d y \right)^2 \right] = \left( \frac{\lambda}{4!} \right)^2 \int \int E\left[ \Phi(x_1)\Phi(x_2)\Phi(y_1)^4\Phi(y_2)^4 \right] d^d y_1 d^d y_2
\]

The reason that Feynman diagram is a powerful calculation tool in QFT is that in fact one can forget about Wick theorem and directly compute with diagrams. This is because, when computing $\left\langle \Phi(x_1)\Phi(x_2) \right\rangle$, all the diagrams we get in the end are the graphs in which every vertex is degree 4, except that the degree of $x_1$ and $x_2$ is 1. (Here the degree 4 is due to the quartic $\Phi^4$ action.) Therefore, one can just directly draw these diagrams without going through the Wick theorem. It is also not very hard to convince oneself that the “symmetry factors” such as $12$ above can be read off from the shape of the graph.

One still needs to compute the denominator $Z$. Using Feynman diagrams,

\[
Z = E\left[ 1 - \frac{\lambda}{4!} \int \Phi(y)^4 \, d^d y + \ldots \right] = 1 - \frac{3\lambda}{4!} y + \ldots
\]

So, by \(\frac{1}{1-x} = 1 + x + \cdots\),

\[
\frac{1}{Z} = 1 + \frac{3\lambda}{4!} y + \ldots
\]

Interestingly, when we multiply this with (3.2), the graph \(y\) times \(x_1 x_2\) precisely cancels out the second diagram on the r.h.s. of (3.2). So we get

\[
\left\langle \Phi(x_1)\Phi(x_2) \right\rangle = x_1 \frac{\lambda}{2} x_2 - \frac{\lambda}{2} x_1 y x_2 + \ldots
\]

A combinatoric “miracle” is that this cancellation occurs at every order in $\lambda$, and in the end only the diagrams in which every vertex is connected with one of the $(x_i)$ (i.e. connected via a sequence of propagators $C$) are left. This is also true for multi-point correlations.

**Renormalization.** In the above calculation, if $d \geq 2$, $C(y - y) = C(0)$ is infinite, so in fact the diagram

\[
x_1 y x_2
\]

does not really have a well-defined value. This is called ultra-violet divergence. The cure is to introduce renormalization constants in the $\Phi^4$ model

\[
S(\Phi) = \int \left( \frac{1}{2} |\nabla \Phi|^2 + \frac{\lambda}{4!} \Phi^4 \right) \, d^d x \quad \rightarrow \quad S(\Phi) = \int \left( \frac{1}{2} |\nabla \Phi|^2 - \frac{C}{2} \Phi^2 + \frac{\lambda}{4!} \Phi^4 \right) \, d^d x
\]
where $C$ is an “infinite constant” (i.e. renormalization constant). More precisely, one should first introduce a small scale cutoff $\varepsilon$, such as a lattice cutoff or Fourier truncation, and introduce a constant $C_\varepsilon$ depending on $\varepsilon$ which diverges as $\varepsilon \to 0$. With this modification of the model, we re-calculate the 2-pt correlation

$$\langle \Phi(x_1)\Phi(x_2) \rangle = \frac{x_1}{x_2} - \frac{\lambda}{2} \Phi(x_1) \langle \Phi(x_2) \rangle + C \frac{y}{x_1} \langle \Phi(x_2) \rangle + \ldots$$

so if we choose $C_\varepsilon = \frac{\lambda}{2} (-\Delta)^{-1}(y - y) = \frac{\lambda}{2} (-\Delta)^{-1}(0)$ which diverges at $d \geq 2$, then the second and the third diagrams on the r.h.s. cancel. This particular renormalization is called the “Wick renormalization”.

Physicists took much effort in the 60s in answering the following question known as perturbative renormalizability: can one introduce a (finite or infinite) collection of renormalization constants in the action $S(\Phi)$ such that every Feynman diagram in the perturbation theory is finite. Physicists know that in $d = 2$, the above constant $C_\varepsilon$ alone is enough to render all diagrams in any order of the perturbation theory finite. In $d = 3$, one needs two constants i.e. $C_\varepsilon = \lambda C^{(1)}_\varepsilon + \lambda^2 C^{(2)}_\varepsilon$. $d = 4$ is much more complicated: one needs to renormalize as

$$S(\Phi) = \int \left( \frac{1}{2} \sum_{i=1}^{N} |\nabla \Phi_i|^2 - \frac{C}{2} \Phi^2 + \frac{\lambda + \tilde{\lambda}}{4!} \Phi^4 \right) d^d x$$

where each of $Z, C, \tilde{\lambda}$ is an infinite series of divergent constants, so that every Feynman diagram in the whole perturbation theory is finite. This is called the Bogoliubov-Parasiuk-Hepp-Zimmermann (BPHZ) theorem. In $d \geq 5$ the theory is not renormalizable.

We caution the reader that each order of perturbation theory being finite does not imply that the measure $e^{-S(\Phi)}D\Phi$ is mathematically well-defined. In $d \geq 4$, mathematicians have proved that no matter how one renormalizes, the only well-defined limit one can get is the Gaussian field (“triviality theorem”).

**Exercise 3.1** Consider the $N$ component generalization of $\Phi^4$ model, where $\Phi = (\Phi_1, \ldots, \Phi_N)$

$$e^{-S(\Phi)}D\Phi, \quad S(\Phi) = \int \left( \frac{1}{2} \sum_{i=1}^{N} |\nabla \Phi_i|^2 + \frac{\lambda}{N} (\Phi_1^2 + \cdots + \Phi_N^2)^2 \right) d^d x.$$ Fix $i$ and find the leading order Feynman diagrams for $\langle \Phi_i(x_1)\Phi_i(x_2) \rangle$. You should find that every vertex is still of degree 4, but should have the form

```
\begin{array}{c}
  \text{\textbullet} \\
  j \quad k \quad j
\end{array}
```

corresponding to the interaction $\frac{\lambda}{N} \sum_{k,j=1}^{N} \Phi_k^2 \Phi_j^2$. Apparently the more vertices we have in a Feynman diagram, the more factors of $\frac{1}{N}$ we have; but on the other hand the indices $j, k$ are also being summed which essentially produces positive powers of $N$.

Count the power of $N$ in each diagram you get for $\langle \Phi_i(x_1)\Phi_i(x_2) \rangle$, and show that all the diagrams come with negative powers of $N$, except that $\frac{x_1}{x_2}$ and the diagrams which will be cancelled by Wick renormalization such as $\frac{x_1}{x_2}$ are of order $O(1)$ i.e. no power of $N$.  


We suitably choose $N$ such that

$$
\Psi
$$

we will show that $\Phi$ where $\mu$ is enough to imply $\Phi = \Phi$ has covariance $\Phi - \mu$ is deterministically larger than $e$ might get very negative and the density diverges logarithmically as $N$.

This argument isn’t really true due to the Wick renormalization - let’s recall that $G$ is a a cut-off function which is smooth, positive, compactly supported, and such $|k/N| = 1$. Below we prove that $e^{-\int_{T^2} :\Phi(x)^4: \, d^2x}$ is a Fourier mollification of $N$.

The key point is to argue that $\Phi$ and from this one might worry that as $N \to \infty$, the energy

$$
H \overset{\text{def}}{=} \int_{T^2} :\Phi(x)^4: \, d^2x
$$

might get very negative and the density $e^{-H}$ might be very large.

So the key point is to argue that $P(H < -K)$ will be very small when $K$ is large. In fact we will show that

$$
P(H < -K) \leq C \exp \left( -c e^{\sqrt{K}} \right)
$$

which is enough to imply $e^{-H} \in L^1(\mu)$.

The basic idea is to decompose $H = H_N + \tilde{H}_N$ with a suitably chosen $N$ such that $H_N$ is deterministically larger than $-K$ and $\tilde{H}_N$ is around 0 with large probability. Define

$$
H_N \overset{\text{def}}{=} \int_{T^2} :\Phi_N(x)^4: \, d^2x \geq -c(\log N)^2 .
$$

We suitably choose $N$ such that $K \approx c(\log N)^2$, so that

$$
P(H < -K) \leq P(|\tilde{H}_N| \geq 3)
$$

and we aim to prove that this is small for $K$ large.
Recalling that $\Phi = \Phi_N + \Psi_N$, $\tilde{H}_N$ then involves several terms by binomial identity, and we only consider one of them called $Y_N$

$$\tilde{H}_N = 4 \int_{\mathbb{T}^2} :\Phi_N(x)^3 : \Psi_N(x) d^2 x + \cdots \overset{\text{def}}{=} Y_N + \cdots$$

and the other terms can be estimated in an analogous way. We will estimate the probability of $Y_N$ by Markov inequality and thus we consider its moments. By Wick theorem,

$$\mathbb{E}|Y_N|^2 = 96 \int \int G_N(x-y)^3 (G - G_N)(x-y) d^2 x d^2 y \lesssim (\log N)^4/N^2 \lesssim 1/N . \quad (3.3)$$

The last inequality can be shown using the following bounds

$$|G_N(x)| \lesssim \log |x| \wedge \log N$$
$$|G(x) - G_N(x)| \lesssim \frac{1}{N^2|x|^2} \wedge \frac{N}{|x|}$$

**Exercise: prove these two inequalities.** (Hint: when $|x| \ll N^{-1}$, one estimates $|G(x) - G_N(x)|$ by bounding $G$ and $G_N$ separately which yields $\log \frac{1}{|x|} + \log N$).

Now we recall the hypercontractivity estimate in Malliavin calculus which gives us that for any random variable $X$ in the $n$-th Wiener chaos,

$$\mathbb{E}[X^{2p}] \leq (2p - 1)^p \mathbb{E}[X^2]^p .$$

Now $Y_N$ is in the 4-th Wiener chaos, so by (3.3) there is a constant $C > 0$ so that for all $N > C$ and $p \geq 1$,

$$\mathbb{E}|Y_N|^{2p} \leq C p^4/N^p .$$

So,

$$\mathbb{P}(|Y_N| \geq 3) \leq \mathbb{E}|Y_N|^{2p}/3^p \lesssim p^p/N^p \lesssim (pe^{-c\sqrt{K}})^4p$$

for some constant $c > 0$. We then choose $p = e^{c\sqrt{K}}$ for some $\tilde{c} < c$ so that $pe^{-c\sqrt{K}} \in (0, 1)$.

We then have

$$\mathbb{P}(|Y_N| \geq 3) \lesssim \exp(-\tilde{c}e^{c\sqrt{K}})$$

for some $\tilde{c} > 0$.

### 3.2 Stochastic quantization and local mild solution

**Stochastic Quantization.** Stochastic Quantization is a procedure to associate a dynamic to a (formal) measure so that the measure is (formally) invariant under the dynamic. Recall that the OU process $\frac{d}{dt} X = -X + \frac{d}{dt} W$ leaves the Gaussian measure $e^{-x^2} dx$ invariant. In general given a measure $e^{-V(x)} d^d x$ on $\mathbb{R}^d$, the dynamic $\frac{d}{dt} X = -\frac{1}{2} \nabla V(X) + \frac{d}{dt} W$ leaves it invariant, where $W$ is the $d$-dimensional Wiener process.

Given a (formal) quantum field theory measure $e^{-S(\Phi)} D\Phi$, its Stochastic Quantization is the following (formal) dynamic, where $\xi$ is the space-time white noise:

$$\partial_t u = -\frac{1}{2} \nabla S(u) + \xi .$$
For instance for the Gaussian free field, the dynamic is
\[ \partial_t u = \Delta u + \xi, \]  
which is just the stochastic heat equation. For \( \Phi^4 \), the associated SPDE is
\[ \partial_t u = \Delta u - u^3 + \xi. \]

**Scaling (dimension analysis).** Recall that if \( u \) satisfies (3.4),
\[ \tilde{u}(t, x) \overset{\text{def}}{=} \lambda^{d+2} u(\lambda^2 t, \lambda x) \]
then it satisfies \( \partial_t \tilde{u}(t, x) = \Delta \tilde{u}(t, x) + \tilde{\xi}(t, x) \) where \( \tilde{\xi}(t, x) := \lambda^{d+2} \xi(\lambda^2 t, \lambda x) \overset{\text{law}}{=} \xi(t, x) \). In other words, (3.4) is scaling invariant.

This can be formalized by saying that “\( t \) has scaling dimension \( x^2 \)” or simply “\( t \) has scaling dimension \( 2 \)”; and, “\( \xi \) has scaling dimension \(- (d + 2)/2 \)” “\( u \) has scaling dimension \(- (d - 2)/2 \)”. Eq. 3.4 then has balanced scalings for all the terms
\[ \frac{\partial_t u}{-\frac{d-2}{2}-2} = \frac{\Delta u}{-\frac{d-2}{2}-2} + \frac{\xi}{-\frac{d+2}{4}}, \]
which is another way of saying ‘scaling invariance’.

(3.5) is not scaling invariance in general. With the above scaling dimensions for \( \xi, u \), we have
\[ \frac{\partial_t u}{-\frac{d+2}{2}} = \frac{\Delta u}{-\frac{d+2}{2}} - \frac{u^3}{-\frac{3(d-2)}{2}} + \frac{\xi}{-\frac{d+2}{4}}, \]
so it’s scaling invariance only if \( d = 4 \) in which case all the terms have dimension \(-3 \). We say that (3.5) is critical at \( d = 4 \), subcritical at \( d < 4 \), and supercritical at \( d > 4 \).

**Exercise 3.2** Prove that the KPZ equation
\[ \partial_t h = \Delta h + (\nabla h)^2 + \xi \]
where \( \xi \) is space-time white noise is critical at \( d = 2 \). Prove that the Navier-Stokes equation driven by space-time white noise
\[ \partial_t u + u \cdot \nabla u = \Delta u - \nabla p + \xi, \quad \text{div } u = 0 \]
is critical at \( d = 4 \).

The \( \Phi^4 \) QFT has the same dimension analysis (of course):
\[ S(\Phi) = \int \left( \frac{\| \nabla \Phi \|^2}{(-\frac{d-2}{2}-1) \times 2} + \frac{\lambda}{2} \frac{\Phi^4}{-\frac{d-2}{2} \times 4} \right) d^d x \cdot d. \]
Here \( (-\frac{d-2}{2}-1) \times 2 + d = 0 \), meaning that the Gaussian free field is scaling invariant; \(-\frac{d-2}{2} \times 4 + d = -d + 4 \), so the \( \Phi^4 \) QFT model is scaling invariant only if \( d = 4 \). Instead of
saying critical or subcritical, physicists say that \( \Phi^4 \) QFT is (marginally) renormalizable at \( d = 4 \), super-renormalizable at \( d < 4 \), and non-renormalizable at \( d > 4 \).

**Local mild solution.** A is called a mild solution with initial condition \( u(0) = u_0 \) if

\[
u(t, x) = \int_{T_0}^t \int_0^T P_{t-s}(x-y)\xi(ds, dy) + P_t u_0(x) - \int_{T_0}^t \int_0^T P_{t-s}(x-y)u(s, y)^3 ds dy \tag{3.6}\]

where \( P_t u_0(x) = \int_{T_0}^t P_t(x-y)u_0(y)dy \) is the solution to heat equation from initial condition \( u_0 \). We call \( \nu \) a local mild solution if there exists a stopping time \( \tau \) s.t. \( \tau > 0 \) a.s. and \( \nu(t, x) \) is a contraction in a ball centered at \( \tau \leq \tau \) a.s.

**Theorem 3.3** Let \( d = 1, \alpha \in (0, \frac{1}{2}) \), and \( u_0 \in C^\alpha \). There exists a unique maximal local mild solution \( (u, \tau) \) and \( u \in C([0, \tau), C^\alpha) \). Moreover \( \lim_{t \to \tau} \|u(t)\|_{C^\alpha} = \infty \) a.s. on \( \{\tau < \infty\} \).

The proof is based on very standard fixed point argument in classical PDE. We will often use this kind of fixed point arguments later.

**Proof.** First, we have the following fact: since \( \alpha > 0 \), \( \|P_t u_0\|_{C^\alpha} \leq \|u_0\|_{C^\alpha} \) (“maximal principle” for heat equation). Given \( T > 0 \), let \( g \in C(\mathbf{R}_+, C^\alpha) \) be given by

\[
g(t, x) \overset{\text{def}}{=} \int_{T_0}^t \int_0^T P_{t-s}(x-y)\xi(ds, dy) + P_t u_0(x) \cdot
\]

Define the fixed point map \( M_{g,T} : C([0, T], C^\alpha) \to C([0, T], C^\alpha) \) by

\[
(M_{g,T} u)(t) = - \int_{T_0}^t \int_0^T P_{t-s}(x-y)u(s, y)^3 ds dy + g(t) .
\]

A fixed point of \( M_{g,T} \) will be a mild solution. **Exercise:** Find such a (unique) fixed point by showing that \( M_{g,T} \) is a contraction in a ball centered at \( g \) in \( C([0, T], C^\alpha) \), provided \( T > 0 \) is small.

\[\Box\]

### 3.3 \( \Phi^4 \) equation in \( d = 2 \)

Consider the renormalized \( \Phi^4 \) equation

\[
\partial_t u_\varepsilon = \Delta u_\varepsilon - (u_\varepsilon^3 - 3C_\varepsilon u_\varepsilon) + \xi_\varepsilon . \tag{3.7}
\]

**Da Prato - Debussche argument.** Write \( u_\varepsilon = \tilde{\Phi}_\varepsilon + v_\varepsilon \), where

\[
\partial_t \Phi_\varepsilon = \Delta \Phi_\varepsilon + \xi_\varepsilon .
\]

The function \( v_\varepsilon \) can be thought of as the “remainder” after the zero-th order perturbation, and it satisfies

\[
\partial_t v_\varepsilon = \Delta v_\varepsilon - \left(v_\varepsilon^3 + 3\Phi_\varepsilon v_\varepsilon^2 + 3(\Phi_\varepsilon^2 - C_\varepsilon) v_\varepsilon + \Phi_\varepsilon^3 - 3C_\varepsilon \Phi_\varepsilon \right) . \tag{3.8}
\]

We can show:
(1) $\Phi^2_\varepsilon - C_\varepsilon$ and $\Phi^3_\varepsilon - 3C_\varepsilon \Phi_\varepsilon$ converge in probability in $C^\alpha_\varepsilon$ for any $\alpha < 0$. We will call the limits $\Phi^2_\varepsilon$ and $\Phi^3_\varepsilon$ respectively. (The fact that their integrations against smooth functions have bounded second moment was essentially already seen in Section ??). Here we identify the distribution space $C^\alpha_\varepsilon$ in which they converge.

(2) When passing to the limit $\varepsilon \to 0$, we can find a local mild solution to the equation

$$\partial_t v = \Delta v - \left( v^3 + 3\Phi v^2 + 3 :\Phi^2 : v + :\Phi^3 : \right)$$

(3.9)

with classical PDE fixed point argument (similarly as in $d = 1$). We could then conclude that $u = \Phi + v$ is the local solution to the renormalized $\Phi^4$ equation in two spatial dimensions – note that $\Phi$ is singular but is Gaussian and explicit, while $v$ is more regular but implicit.

The above argument was first used by Da Prato and Debussche in [DPD03].

**Wiener chaos.** Before proving the above two results, one might ask: why has the constant $C_\varepsilon$ in (3.7) been precisely distributed to the “right places” in (3.8) to make the powers of $\Phi_\varepsilon$ converge? We first give a better interpretation to the renormalization in this context.

The space-time white noise $\xi$ as defined by (1.2) can be viewed as an imbedding from $L^2$ functions to $L^2$ random variables:

$$L^2(\mathbb{R}^{d+1}) \hookrightarrow L^2(\Omega) \quad f \mapsto \xi(f).$$

It preserves the $L^2$ norm: $E[\xi(f)^2] = \|f\|^2_{L^2}$. Write $H \subset L^2(\Omega)$ for the image of this imbedding; elements of $H$ are centered Gaussian random variables of the form $\xi(f)$. Let

$$H^n = \{ p(Z_1, \ldots, Z_k) \mid p \in \mathbb{R}[z_1, \ldots, z_k], \deg p \leq n, Z_1, \ldots, Z_k \in H \},$$

where $\mathbb{R}[z_1, \ldots, z_k]$ is the ring of real-coefficient polynomials of $k$ unknowns. In plain words, $H^n$ contains random variables which are $n$-th order polynomials of Gaussians. Its closure $\bar{H}^n$ in $L^2(\Omega)$ is called the $n$-th Wiener chaos. The orthogonal complement of $\bar{H}^{n-1}$ in $\bar{H}^n$, denoted by $\bar{H}^{n-1}$, is called the homogeneous chaos of order $n$. A useful fact is that if $Z \in H$, then $h_n(Z) \in \bar{H}^n$: where $h_n$ is the $n$-th Hermite polynomial. (This is due to the fact that the Hermite polynomials form orthonormal basis of $L^2(\Omega, \sqrt{2\pi} e^{-x^2/2} dx)$.)

### 3.4 $\Phi^4$ equation in $d = 3$

The above argument for solving the fixed point problem

$$\partial_t v = \Delta v - \left( v^3 + 3\Phi v^2 + 3 :\Phi^2 : v + :\Phi^3 : \right)$$

(3.10)

does not work in $d = 3$.

Local well-posedness in $d = 3$ was proved using regularity structures, or paracontrolled distributions. For simplicity of our exposition, we consider a simpler equation (without loss of any key ideas)

$$\partial_t w = \Delta w - 3 :\Phi^2 : w.$$  

(3.11)

Roughly speaking, the idea is that we need “higher order” approximations. We denote $\Phi$ (which is equal to the heat kernel convolving the white noise) by $\Phi$, with a line representing the heat kernel, and the circle representing the noise. Similarly, $:\Phi^2 :$ is denoted by $\Phi$. 

Since we expect that \( w \) should be more regular than \( \Phi^2 : \), the small scale roughness of the product \( \Phi^2 : w \) should be dominated by \( \Phi^2 : \), and the solution \( w \) to (3.11) can be "locally approximated" by the solution to the following equation

\[
\partial_t \Phi = \Delta \Phi + \Phi^2 : \Phi^2 : 
\]

namely, \( \Phi \) is the heat kernel convolving \( \Phi^2 : \), as the graphic notation indicates. One can prove that \( \Phi \in C^\alpha \) for any \( \alpha < 1 \).

The above saying that \( w \) can be "locally approximated" by \( \Phi \) can be formalized in the following way. Letting \( x_0 \) be a fixed space-time point, in a neighborhood of \( x_0 \) one should have

\[
w(x) - w(x_0) = -3w(x_0)(\Phi(x) - \Phi(x_0)) + R(x_0, x) \tag{3.12}
\]

where \( R(x_0, x) \) is something smoother, in the sense that as \( x \to x_0 \), \( R(x_0, x) \) vanishes faster than \( |\Phi(x) - \Phi(x_0)| \approx |x - x_0|^{\alpha - 1} \). Note that this is similar as Taylor expansion where a \( k \)-th differentiable function can be locally approximated by a \( k \)-th order polynomial with a "Taylor remainder" \( R \).

With the above expansion (3.12) for \( w \), we can derive an expansion for \( \Phi^2 : w = w \Phi \)

\[
(w \Phi)(x) - (w \Phi)(x_0) \\
= \left[ w(x_0) - 3w(x_0)(\Phi(x) - \Phi(x_0)) + R(x_0, x) \right] \left( \Phi(x_0) + (\Phi(x) - \Phi(x_0)) \right) - (w \Phi)(x_0) \\
= w(x_0)(\Phi(x) - \Phi(x_0)) - 3w(x_0)(\Phi(x) - \Phi(x_0))\Phi(x) + R(x_0, x)\Phi(x)
\]

Now we have a product of \( \Phi \) and \( \Phi \), which have regularities \( C^{1-} \) and \( C^{1-} \) respectively. It can’t be defined classically (just like the product of \( w \) and \( \Phi \)), but, it can be defined with probabilistic method. We call this product \( \Phi \Phi \). This “tree” requires another renormalization.
4  (Controlled) rough paths

In Itô calculus, we define Itô integral by approximating measurable, adapted processes by simple processes, and apply Itô isometry. It turns out that there is a different approach: roughly speaking, by postulating the simple integral \( \int X \, dX \) one constructs general integrals \( \int Y \, dX \) for \( Y \) which locally “look like” \( X \). This is called “rough path theory”. The true power of “rough path theory” is that it gives an almost sure \textit{pathwise} definition of stochastic differential equation; and the solution to the equation, when lifted onto the level of rough paths, is \textit{continuous} in the driven noise.

We denote by \( C = C([0, T] \times \mathbb{R}^n) \) the space of continuous functions, and by \( C_2([0, T], \mathbb{R}^n) \) the space of continuous functions \( f: \mathbb{R}^n \to \mathbb{R}^n \) that vanish on the diagonal. We will often omit \([0, T]\) and \( \mathbb{R}^n \) in our notations. We define a difference operator \( \delta: C \to C_2 \) by
\[
\delta X_{s,t} = X_t - X_s. \quad \text{For an element } X \in C_2 \text{ we will mainly be interested in the rate at which it vanishes at diagonal } \|X\|_\beta = \sup_{s \neq t} \frac{|X_{s,t}|}{|s-t|^\beta}, \text{ rather than its regularity in each of its variables.}
\]

**Rough paths.** A rough path \((X, X)\) consists of two parts: a continuous function \( X \in C([0, T], \mathbb{R}^n) \), as well as a continuous “area process” \( X \in C_2([0, T], \mathbb{R}^{n \times n}) \) such that the algebraic relations
\[
X^{ij}(r, t) - X^{ij}(r, s) - X^{ij}(s, t) = \delta X^i(r, s) \delta X^j(s, t), \quad (4.1)
\]
hold for every triple of points \((r, s, t)\) and every pair of indices \((i, j)\). One should think of \( X \) as an “auxiliary piece of information” which postulates the value of the quantity
\[
\int_s^t \delta X^i(s, r) \, dX^j(r) \overset{\text{def}}{=} X^{ij}(s, t), \quad (4.2)
\]
where we take the right hand side as a \textit{definition} for the left hand side. The aim of imposing \((4.1)\) is to ensure that \( X^{ij} \) does indeed behave like an integral when considering it over two adjacent intervals. We sometimes also call a rough path \( X \) instead of \((X, X)\) for simplicity. For \( \alpha \in (0, 1) \) we say \((X, X) \in \mathcal{G}^\alpha\) if \((4.1)\) holds and
\[
\|X\|_\alpha = \sup_{s \neq t} \frac{\|\delta X_{s,t}\|}{|s-t|^\alpha} < \infty, \quad \|X\|_{2\alpha} = \sup_{s \neq t} \frac{|X_{s,t}|}{|s-t|^{2\alpha}} < \infty.
\]

**Example 4.1** Let \( n = 1 \), \( X \) be the Brownian motion\(^{15}\) and \( X_{s,t} = \frac{1}{2} (B_t - B_s)^2 - ct(t-s) \) where \( c \in \mathbb{R} \). Then \((X, X)\) is a rough path, and \((X, X) \in \mathcal{G}^\alpha\) for \( \alpha < \frac{1}{2} \). In fact, we are here just postulating that
\[
\int_0^t B_s \, dB_s = \frac{1}{2} B_t^2 - ct \quad (4.3)
\]
because \((4.3)\) is equivalent (Exercise) with \( \int_s^t \delta B(s, r) \, dB(r) = \frac{1}{2} (B_t - B_s)^2 - c(t-s) \). It can be easily checked that \((4.1)\) holds (Exercise).

**Controlled rough paths.** Given a rough path \( X \) taking values in \( \mathbb{R}^n \), we say that a pair of functions \((Y, Y')\) is a rough path \textit{controlled by} \( X \) if the “remainder term” \( R \) given by\(^{16}\)
\[
R(s, t) \overset{\text{def}}{=} \delta Y(s, t) - Y'(s) \delta X(s, t), \quad (4.4)
\]
\(^{15}\)a typical sample path of Brownian motion  
\(^{16}\)Here, \( Y(s, t), R(s, t) \in \mathbb{R}^m \) and the second term is a matrix-vector multiplication. Note that this definition does not depend on \( X \) thus we did not explicitly mention it.
has better regularity properties than $Y$. Typically, we will assume that $\|Y\|_\alpha < \infty$ and $\|Y'\|_\alpha < \infty$ for some Hölder exponent $\alpha \in (0, 1)$, but $\|R\|_{2\alpha} < \infty$, where $\|R\|_{2\alpha}$ is the norm for functions of two variables defined as above, namely $|R(s, t)| \lesssim |s - t|^{2\alpha}$.

One could heuristically think of (4.4) as similar with Taylor expansion for a smooth function

$$f(t) - f(s) = f'(s)(t - s) + O(|t - s|^2)$$

and just like smooth functions can be approximated by polynomials in $t - s$ in a neighborhood of $s$, (4.4) says that $Y$ can be approximated by $X$ locally. Thus $Y'$ is often called the Gubinelli derivative.

The space of controlled rough paths controlled by $X$ is denoted by $\mathcal{D}_X^{2\alpha}$ endowed with the semi-norm

$$\|Y, Y'\|_{X, 2\alpha} = \|Y'\|_\alpha + \|R\|_{2\alpha}.$$ 

$\mathcal{D}_X^{2\alpha}$ is a Banach space under the norm $|Y_0| + |Y_0'| + \|Y, Y'\|_{X, 2\alpha}$. On a fixed interval $[0, T]$ the above norm on $\mathcal{D}_X^{2\alpha}$ being finite implies that $Y$ is $\alpha$-Hölder: indeed,

$$\|Y\|_\alpha \leq \|Y'\|_\infty \|X\|_\alpha + \|R\|_\alpha \leq C(1 + \|X\|_\alpha)(|Y_0'| + \|Y, Y'\|_{X, 2\alpha}).$$

A priori there could be many distinct “derivative processes” $Y'$ associated to a path $Y$. However, under certain conditions (so called “truly rough” [FH14 Section 6.2]) $Y'$ is unique. For instance this is the case if $X$ is a typical sample path of Brownian motion and if we impose the bound $\|R\|_\beta < \infty$ for some $\beta > \frac{1}{2}$.

Example 4.2 If $f$ is a smooth function, $X = B$, and consider $Y := f(B)$, then

$$\delta Y(s, t) = f(B_t) - f(B_s) = f'(B_s)(B_t - B_s) + f''(B_s)(B_t - B_s)^2 + \cdots$$

So we should choose $Y'(t) = f'(B_t)$ then $(Y, Y')$ is a rough path controlled by $B$.

Here is an interesting lemma, which says that composition of a controlled path with a smooth function is still a controlled path.

Lemma 4.3 If $f$ is a smooth function, $(Y, Y') \in \mathcal{D}_X^{2\alpha}$, then $(f(Y), f(Y')) \in \mathcal{D}_X^{2\alpha}$ where $f(Y') \triangleq DF(Y)Y'$. Furthermore

$$\|f(Y), f(Y')\|_{X, 2\alpha} \leq C_{\alpha, T, f}(1 + \|X\|_\alpha)^2(1 + |Y_0'| + \|Y, Y'\|_{X, 2\alpha}).$$

Proof. (Sketch.) Using $(Y, Y') \in \mathcal{D}_X^{2\alpha}$ it is clear that

$$f(Y_t) = f(Y_s) + Df(Y_s)\delta Y_{s,t} + O(\delta Y_{s,t}^2) = f(Y_s) + Df(Y_s)(Y'_s \delta X_{s,t} + R^Y_{s,t}) + O(\delta Y_{s,t}^2)$$

so

$$f(Y_t) = Df(Y)Y', \quad R^f(Y)_{s,t} = Df(Y_s)R^Y_{s,t} + O(\delta Y_{s,t}^2).$$

From this and the assumed bounds on $Y, Y', R^Y$ it’s then easy to prove the lemma. See [FH14 Lemma 7.3] for details.

Integrating controlled paths. It turns out that if $(X, X)$ is a rough path taking values in $\mathbb{R}^n$ and $Y$ is a path controlled by $X$ that also takes values in $\mathbb{R}^n$, then one can give a natural meaning to the integral $\int_{t}^{b}(Y_t, dX_t)$, provided that $X$ and $Y$ are sufficiently regular. The
approximation $Y_t \approx Y_s + Y'_s \delta X_{s,t}$ suggested by (4.4) shows that it is reasonable to define the integral as the following limit of “second-order Riemann sums”:

$$\int_a^b \langle Y(t), dX(t) \rangle = \lim_{||P|| \to 0} \sum_{[s,r] \in P} \left( \langle Y(s), \delta X(s, r) \rangle + \text{tr} Y'(s) X(s, r) \right), \quad (4.5)$$

where $P$ denotes a partition of the integration interval.

**Theorem 4.4** Let $(X, X) \in \mathcal{C}^\alpha$ and $(Y, Y') \in \mathcal{D}^{2\alpha}_X$ with a remainder $R$ given by (4.4). Then, provided that $\alpha > 1/3$, (4.5) converges, and

$$\left| \int_s^t \langle \delta Y(s, r), dX(r) \rangle - \text{tr} Y'(s) X(s, t) \right| \leq C_{n,\alpha} |t - s|^3 \alpha^3 (\|X\|_\alpha \|R\|_{2\alpha} + \|X\|_{2\alpha} \|Y'\|_\alpha). \quad (4.6)$$

The basic idea of proof of Theorem 4.4 goes as follows. Recall that

$$Y(t) - Y(s) = Y'(s)(X(t) - X(s)) + R(s, t)$$

When integrate the left hand side w.r.t. $X$, one only needs to define the integrations of the two terms on the right hand side w.r.t. $X$:

1. The integration of $X(t) - X(s)$ w.r.t. $X$ is postulated by $X$, so as long as $\|X\|_{2\alpha} < \infty$ and $\|Y'\|_\alpha < \infty$ and $3\alpha > 1$ (i.e. the Young’s condition holds) one can integrate $Y'(s)(X(t) - X(s))$ w.r.t. $X$.

2. Regarding $R$, since $\|X\|_\alpha < \infty$ and $\|R\|_{2\alpha} < \infty$, and $\alpha > 1/3$, its integration against $X$ is negligible.

For detailed proofs, we refer to [FH14, Section 4.1-4.3].

**Example 4.5** Take the above examples of Brownian motion, $f$ a smooth function, we can define the integral against $B$ as

$$\int_a^b f(B_t) dB_t = \lim_{||P|| \to 0} \sum_{[s,r] \in P} \left( f(B_s)(B_r - B_s) + f'(B_s) X(s, r) \right) = \lim_{||P|| \to 0} \sum_{[s,r] \in P} \left( f(B_s)(B_r - B_s) + f'(B_s) (\frac{1}{2}(B_r - B_s)^2 - c(r - s)) \right) \quad (4.7)$$

Note that a typical sample path of Brownian motion is Hölder $\alpha \in (1/3, 1/2)$, so this limit converges by the above theorem. It is important to understand that

- The statement $(X, X) \in \mathcal{C}^\alpha$ needs probability (in particular Kolmogorov theorem leading to $B \in \mathcal{C}^\alpha$ a.s. for $\alpha \in (1/3, 1/2)$)

- After we take $(X, X) \in \mathcal{C}^\alpha$ the convergence of the above Riemann sum is a purely deterministic statement, and the limit above is deterministic (not “limit of random variables” as in the standard construction of Itô integrals).

To compare with Itô integral, we need to again think of the $B$ in (4.7) as the Brownian motion (rather than a typical sample path of it), and when $c = \frac{1}{3}$, the second term converges to zero in probability in the limit, and this is exactly Itô integral.
Results such as Itô formula can be also proved, see [FH14, Chapter 5].

**Solving differential equation driven by \((X, X)\) in the space \(C_\alpha^\alpha\).** It will turn out that a solution to an SDE

\[
dX_t = f(X_t)dB_t
\]

will also be controlled by \(B\). “\(Y\) controlled by a rough path \(X\)” means that “\(Y\) behaves like \(B\) at small scales”. Indeed, in the above example, if one takes a small portion of the path \(B\) and a small portion of the path \(f(B)\), one can not really tell the difference between the two small portions (although the two paths may look very different at large scales, for instance \(f(B)\) has a drift.) It can be also interpreted as “\(Y\) can be locally approximated by \(B\)”, just like \(Y \in C^1\) means “\(Y\) can be locally approximated by a linear / tangent function”:

\[
Y(t) - Y(s) = Y'(s)(X(t) - X(s)) + R(s, t) \quad X(t) = t
\]

where \(R(s, t) = C^2\). Here \(Y'\) is the derivative and \(R\) is the Taylor Remainder, as the notations indicate.

In Itô’s approach, Under linear growth and Lipschitz conditions, for \(L^2\) initial data, we can prove existence and uniqueness of adapted \(L^2\) solution. See standard book, such as Oksendal’s book Section 5.2. This is essentially our approach for mSHE in Section 2.1.

Fixed point argument via rough paths: [FH14, Chapter 8]. Being typed......

## 5 Regularity structures

As we have seen, the strategy described in Section 3.3 would not work for the \(\Phi^4\) equation in three spatial dimensions. Regularity structure theory as developed by [Hai14] is in a certain sense a far-reaching generalization of the rough paths; the theory is able to give meanings to local solutions for a large class of singular SPDEs including \(\Phi^4\) in three spatial dimensions.

### 5.1 Basic definitions in regularity structure theory

**Definition 5.1** Let \(A \subset \mathbb{R}\) be bounded from below and without accumulation point, and let \(T = \bigoplus_{\alpha \in A} T_\alpha\) be a vector space graded by \(A\) such that each \(T_\alpha\) is a Banach space. \(\uparrow\) Let furthermore \(G\) be a group of continuous operators on \(T\) such that, for every \(\alpha \in A\), every \(\Gamma \in G\), and every \(\tau \in T_\alpha\), one has \(\Gamma \tau - \tau \in \bigoplus_{\beta < \alpha} T_\beta\). The triple \((A, T, G)\) is called a regularity structure with structure group \(G\).

Given \(\tau \in T\), we will write \(\|\tau\|_\alpha\) for the norm of its component in \(T_\alpha\).

A regularity structure is a completely abstract object, with its element \(\tau\) being just abstract symbols. It only becomes useful when endowed with a model, which is a concrete way of associating to any \(\tau \in T\) and \(x_0 \in \mathbb{R}^d\), the actual Taylor polynomial, Brownian motion or some other canonical function / distribution based at \(x_0\) that is represented by \(\tau\).

**Definition 5.2** Given a regularity structure \(\mathcal{T}\) and an integer \(d \geq 1\), a model for \(\mathcal{T}\) on \(\mathbb{R}^d\) consists of maps

\[
\Pi : \mathbb{R}^d \rightarrow \mathcal{L}(T, \mathcal{S}'(\mathbb{R}^d)) \quad \Gamma : \mathbb{R}^d \times \mathbb{R}^d \rightarrow G
\]

\[
x \mapsto \Pi_x \quad (x, y) \mapsto \Gamma_{xy}
\]

\(\uparrow\)In most of applications \(A\) is finite and \(T_\alpha\) is finite dimensional.
Then, there exists a unique linear map

\[ \Gamma_{xy} \Gamma_{yz} = \Gamma_{xz} \]  

(5.1)

and

\[ \Pi_x \Gamma_{xy} = \Pi_y. \]  

(5.2)

Furthermore, given \( r > |\inf A| \), for any compact set \( \hat{\mathcal{R}} \subset \mathbb{R}^d \) and constant \( \gamma > 0 \), there exists a constant \( C \) such that the bounds

\[ \| (\Pi_x \tau)(\varphi^\lambda_n) \| \leq C |\lambda|^r \| \tau \|_\alpha, \quad \| \Gamma_{xy} \tau \|_\beta \leq C |x - y|^{\alpha - \beta} \| \tau \|_\alpha, \]  

(5.3)

hold uniformly over \( \varphi \in \mathcal{B}_r, (x, y) \in \hat{\mathcal{R}}, \lambda \in (0, 1], \tau \in T_\alpha \) with \( \alpha \leq \gamma \), and \( \beta < \alpha \).

With these definitions at hand, it is then natural to define an equivalent in this context of the space of \( \gamma \)-Hölder continuous functions in the following way.

**Definition 5.3** Given a regularity structure \( \mathcal{T} \) equipped with a model \( (\Pi, \Gamma) \) over \( \mathbb{R}^d \), the space \( \mathcal{D}^\gamma = \mathcal{D}^\gamma(\mathcal{T}, \Gamma) \) is given by the set of functions \( f : \mathbb{R}^d \to \bigoplus_{\alpha < \gamma} T_\alpha \) such that, for every compact set \( \hat{\mathcal{R}} \) and every \( \alpha < \gamma \), the exists a constant \( C \) with

\[ \| f(x) - \Gamma_{xy} f(y) \|_\alpha \leq C |x - y|^{\gamma - \alpha} \]  

(5.4)

uniformly over \( x, y \in \hat{\mathcal{R}} \).

### 5.2 Reconstruction theorem

The most fundamental result in the theory of regularity structures then states that given \( f \in \mathcal{D}^\gamma \) with \( \gamma > 0 \), there exists a unique Schwartz distribution \( \mathcal{R} f \) on \( \mathbb{R}^d \) such that, for every \( x \in \mathbb{R}^d, \mathcal{R} f \) “looks like \( \Pi_x f(x) \) near \( x \)”. More precisely, one has

**Theorem 5.4** Let \( \mathcal{T} \) be a regularity structure as above and let \( (\Pi, \Gamma) \) a model for \( \mathcal{T} \) on \( \mathbb{R}^d \). Then, there exists a unique linear map \( \mathcal{R} : \mathcal{D}^\gamma \to \mathcal{S}'(\mathbb{R}^d) \) such that

\[ |(\mathcal{R} f - \Pi_x f(x))(\varphi^\lambda_n)| \lesssim \lambda^\gamma, \]  

(5.5)

uniformly over \( \varphi \in \mathcal{B}_r \) and \( \lambda \) as before, and locally uniformly in \( x \).

**Proof.** Given any \( r > |\inf A| > 0 \), recall the function \( \varphi \in C^r_c(\mathbb{R}^d) \) in the discussion on wavelets below Theorem 1.3. This function \( \varphi \) can be also chosen such that

1. For every polynomial \( P \) of degree \( r \), there exists a polynomial \( \hat{P} \) of degree \( r \) such that, for every \( x \in \mathbb{R}^d \), one has \( \sum_{y \in \mathbb{Z}^d} \hat{P}(y) \varphi(x - y) = P(x) \).

2. One has \( \int \varphi(x) \varphi(x - y) dx = \delta_{y,0} \) for every \( y \in \mathbb{Z}^d \).

3. There exist coefficients \( \{ a_k \}_{k \in \mathbb{Z}^d} \) such that \( 2^{-d/2} \varphi(x/2) = \sum_{k \in \mathbb{Z}^d} a_k \varphi(x - k) \).

We also set \( \Lambda^n = 2^{-n} \mathbb{Z}^d \) and, for \( y \in \Lambda^n \), we set \( \varphi^\lambda_n(x) = 2^{nd/2} \varphi(2^n (x - y)) \). Here, the normalisation is chosen in such a way that the set \( \{ \varphi^\lambda_n \}_{y \in \Lambda^n} \) is again orthonormal in \( L^2 \). We then denote by \( V_n \subset C^r \) the linear span of \( \{ \varphi^\lambda_n \}_{y \in \Lambda^n} \), so that, by the property (3) above, one has \( V_0 \subset V_1 \subset V_2 \subset \ldots \). We furthermore denote by \( \hat{V}_n \) the \( L^2 \)-orthogonal complement of \( V_{n-1} \) in \( V_n \), so that \( V_n = V_0 \oplus V_1 \oplus \ldots \oplus \hat{V}_n \). In order to keep notations compact, it will also be convenient to define the coefficients \( a^\lambda_n \) with \( k \in \Lambda^n \) by \( a^\lambda_n = a_{2^n k} \).
With these notations at hand, we then define a sequence of linear operators $\mathcal{R}^n: \mathcal{D}^\gamma \rightarrow \mathcal{C}^r$ by

$$(\mathcal{R}^n f)(y) = \sum_{x \in \Lambda^n} (\Pi_x f(x))(\varphi^n_x) \varphi^n_x(y).$$

We claim that there then exists a Schwartz distribution $\mathcal{R}f$ such that, for every $\psi \in \mathcal{C}^r$, one has $\langle \mathcal{R}^n f(\psi) \rangle \rightarrow \langle \mathcal{R}f(\psi) \rangle$, and that $\mathcal{R}f$ furthermore satisfies the properties stated in the theorem.

Let us first consider the size of the components of $\mathcal{R}^{n+1}f - \mathcal{R}^nf$ in $V_n$. Given $x \in \Lambda^n$, we make use of properties (3-4), so that

$$\langle \mathcal{R}^{n+1}f - \mathcal{R}^nf, \varphi^n_x \rangle = \sum_{k \in \Lambda^{n+1}} a_k^n \langle \mathcal{R}^{n+1}f, \varphi^{n+1}_{x+k} \rangle - (\Pi_x f(x))(\varphi^n_x)$$

$$= \sum_{k \in \Lambda^{n+1}} a_k^n (\Pi_{x+k} f(x+k))(\varphi^{n+1}_{x+k}) - (\Pi_x f(x))(\varphi^n_x)$$

$$= \sum_{k \in \Lambda^{n+1}} a_k^n ((\Pi_{x+k} f(x+k))(\varphi^{n+1}_{x+k}) - (\Pi_x f(x))(\varphi^{n+1}_{x+k}))$$

$$= \sum_{k \in \Lambda^{n+1}} a_k^n (\Pi_{x+k} (f(x+k) - \Gamma_{x+k,x} f(x)))(\varphi^{n+1}_{x+k}). \quad \text{by (5.2)}$$

Since only finitely many of the coefficients $a_k$ are non-zero, it follows from the definition of $\mathcal{D}^\gamma$ that for the non-vanishing terms in this sum we have the bound

$$\|f(x+k) - \Gamma_{x+k,x} f(x)\|_{\alpha} \lesssim 2^{-n(\gamma - \alpha)},$$

uniformly over $n \geq 0$ and $x$ in any compact set. Furthermore, for any $\tau \in T_\alpha$, it follows from the definition of a model that one has the bound

$$|\langle \Pi_x \tau \rangle (\varphi^n_x)| \lesssim 2^{-an - \frac{nd}{2}},$$

again uniformly over $n \geq 0$ and $x$ in any compact set. Here, the additional factor $2^{-\frac{nd}{2}}$ comes from the fact that the functions $\varphi^n_x$ are normalised in $L^2$ rather than $L^1$. Combining these two bounds, we immediately obtain that

$$|\langle \mathcal{R}^{n+1}f - \mathcal{R}^nf, \varphi^n_x \rangle| \lesssim 2^{-\gamma n - \frac{nd}{2}}, \quad (5.6)$$

uniformly over $n \geq 0$ and $x$ in compact sets. Take now a test function $\psi \in \mathcal{C}^r$ with compact support and let us try to estimate $\langle \mathcal{R}^{n+1}f - \mathcal{R}^nf, \psi \rangle$. Since $\mathcal{R}^{n+1}f - \mathcal{R}^nf \in V_{n+1}$, we can decompose it into a part $\delta \mathcal{R}^nf \in V_n$ and a part $\delta \mathcal{R}^nf \in \tilde{V}_{n+1}$ and estimate both parts separately. Regarding the part in $V_n$, we have

$$|\langle \delta \mathcal{R}^nf, \psi \rangle| = \left| \sum_{x \in \Lambda^{n+1}} \langle \delta \mathcal{R}^nf, \varphi^n_x \rangle \langle \varphi^n_x, \psi \rangle \right| \lesssim 2^{-\gamma n - \frac{nd}{2}} \sum_{x \in \Lambda^{n+1}} |\langle \varphi^n_x, \psi \rangle|, \quad (5.7)$$

where we made use of the bound (5.6). At this stage we use the fact that, due to the boundedness of $\psi$, we have $|\langle \varphi^n_x, \psi \rangle| \lesssim 2^{-nd/2}$. Furthermore, thanks to the boundedness of the support of $\psi$, the number of non-vanishing terms appearing in this sum is bounded by $2^{nd}$, so that we eventually obtain the bound

$$|\langle \delta \mathcal{R}^nf, \psi \rangle| \lesssim 2^{-n}. \quad (5.8)$$
Regarding the second term, we use the standard fact coming from wavelet analysis that a basis of $\mathcal{V}_{n+1}$ can be obtained in the same way as the basis of $\mathcal{V}_n$, but replacing the function $\varphi$ by functions $\hat{\varphi}$ from some finite set $\Phi$. In other words, $\mathcal{V}_{n+1}$ is the linear span of $\{\hat{\varphi}_x^n\}_{x \in \Lambda^n; \hat{\varphi} \in \Phi}$. Furthermore, as a consequence of property (2), the functions $\hat{\varphi} \in \Phi$ all have the property that

$$\int \hat{\varphi}(x) P(x) \, dx = 0,$$  

for any polynomial $P$ of degree less or equal to $r$. In particular, this shows that one has the bound

$$|\langle \hat{\varphi}_x^n, \psi \rangle| \lesssim 2^{-\frac{nr}{2}}.$$

As a consequence, one has

$$|\langle \hat{\mathcal{R}}^n f, \psi \rangle| = \left| \sum_{x \in \Lambda^n} \langle \mathcal{R}^{n+1} f, \hat{\varphi}_x^n \rangle \langle \hat{\varphi}_x^n, \psi \rangle \right| \lesssim 2^{-\frac{nr}{2}} \left| \sum_{x \in \Lambda^n} \langle \mathcal{R}^{n+1} f, \hat{\varphi}_x^n \rangle \right|.$$

At this stage, we note that, thanks to the definition of $\mathcal{R}^{n+1}$ and the bounds on the model $(\Pi, \Gamma)$, we have $|\langle \mathcal{R}^{n+1} f, \hat{\varphi}_x^n \rangle| \lesssim 2^{-\frac{nr}{2} - n\alpha}$, where $\alpha_0 = \inf A$, so that $|\langle \hat{\mathcal{R}}^n f, \psi \rangle| \lesssim 2^{-\frac{nr}{2} - n\alpha}$. Combining this with (5.8), we see that one has indeed $\mathcal{R}^n f \rightarrow \mathcal{R} f$ for some Schwartz distribution $\mathcal{R} f$.

It remains to show that the bound (5.5) holds. For this, given a distribution $\eta \in \mathcal{C}^\alpha$ for some $\alpha > -r$, we first introduce the notation

$$P_n \eta = \sum_{x \in \Lambda^n} \eta(\varphi_x^n) \hat{\varphi}_x^n, \quad \hat{P}_n \eta = \sum_{\hat{\varphi} \in \Phi} \sum_{x \in \Lambda^n} \eta(\varphi_x^n) \hat{\varphi}_x^n.$$

We also choose an integer value $n \geq 0$ such that $2^{-n} \sim \lambda$ and we write

$$\mathcal{R} f - \Pi_x f(x) = \mathcal{R}^n f - P_n \Pi_x f(x) + \sum_{m \geq n} (\mathcal{R}^{m+1} f - \mathcal{R}^m f - \hat{P}_m \Pi_x f(x))$$

$$= \mathcal{R}^n f - P_n \Pi_x f(x) + \sum_{m \geq n} (\hat{\mathcal{R}}^m f - \hat{P}_m \Pi_x f(x)) + \sum_{m \geq n} \delta \mathcal{R}^m f. \quad (5.10)$$

We then test these terms against $\psi^\lambda_x$ and we estimate the resulting terms separately. For the first term, we have the identity

$$\langle \mathcal{R}^n f - P_n \Pi_x f(x) \rangle \langle \psi^\lambda_x \rangle = \sum_{y \in \Lambda^n} \langle \Pi_y f(y) - \Pi_x f(x) \rangle \langle \varphi^\lambda_y \rangle \langle \varphi^\lambda_x \rangle. \quad (5.11)$$

We have the bound $|\langle \varphi^\lambda_y, \psi^\lambda_x \rangle| \lesssim \lambda^{-d} 2^{-dn/2} \sim 2^{dn/2}$. Since one furthermore has $|y - x| \lesssim \lambda$ for all non-vanishing terms in the sum, one also has similarly to before

$$|\langle \Pi_y f(y) - \Pi_x f(x) \rangle \langle \varphi^\lambda_y \rangle | \lesssim \sum_{\alpha < \gamma} \lambda^{\alpha - d} 2^{-dn/2 - \alpha n} \sim 2^{-dn/2 - \gamma n}. \quad (5.12)$$

Since only finitely many (independently of $n$) terms contribute to the sum in (5.11), it is indeed bounded by a constant proportional to $2^{-\gamma n} \sim \lambda^\gamma$ as required.

We now turn to the second term in (5.10), where we consider some fixed value $m \geq n$. We rewrite this term very similarly to before as

$$\langle \hat{\mathcal{R}}^m f - \hat{P}_m \Pi_x f(x) \rangle \langle \psi^\lambda_x \rangle = \sum_{\hat{\varphi} \in \Phi} \sum_{y,z} \langle \Pi_y f(y) - \Pi_x f(x) \rangle \langle \varphi^{m+1}_y, \hat{\varphi}^m_z \rangle \langle \hat{\varphi}^m_z, \psi^\lambda_x \rangle,$$
where the sum runs over \( y \in \Lambda^{m+1} \) and \( z \in \Lambda^m \). This time, we use the fact that by the property (5.9) of the wavelets \( \hat{\varphi} \), one has the bound
\[
|\langle \varphi^m_y, \psi^\lambda \rangle| \lesssim \lambda^{-d-r}2^{-rm-\frac{md}{2}},
\] (5.13)
and the \( L^2 \)-scaling implies that \( |\langle \varphi^m_y, \hat{\varphi}^m \rangle| \lesssim 1 \). Furthermore, for each \( z \in \Lambda^m \), only finitely many elements \( y \in \Lambda^{m+1} \) contribute to the sum, and these elements all satisfy \( |y - z| \lesssim 2^{-m} \). Bounding the first factor as in (5.12) and using the fact that there are of the order of \( \lambda^{d2md} \) terms contributing for every fixed \( m \), we thus see that the contribution of the second term in (5.10) is bounded by
\[
\sum_{m \geq n} \lambda^{d2md} \sum_{\alpha < \gamma} \lambda^{-\alpha - d - r - 2dm - am - rm} \sim \sum_{m \geq n} \lambda^{-\alpha - r} \sum_{m \geq n} 2^{-am - rm} \sim \lambda^\gamma.
\]
For the last term in (5.10), we combine (5.7) with the bound \( |\langle \varphi^m_y, \psi^\lambda \rangle| \lesssim \lambda^{-d2 - dm/2} \) and the fact that there are of the order of \( \lambda^{d2md} \) terms appearing in the sum (5.7) to conclude that the \( m \)th summand is bounded by a constant proportional to \( 2^{-\gamma m} \). Summing over \( m \) yields again the desired bound and concludes the proof. \( \square \)

5.3 Example: controlled rough paths

Let us see now how the theory of controlled rough paths can be reinterpreted as an example of regularity structure theory. For given \( \alpha \in (\frac{1}{3}, \frac{1}{2}) \) and \( n \geq 1 \), we can define a regularity structure \( \mathcal{S} \) by setting \( A = \{ \alpha - 1, 2\alpha - 1, 0, \alpha \} \). We furthermore take for \( T_0 \) a copy of \( \mathbb{R} \) with unit vector \( 1 \), for \( T_\alpha \) and \( T_{\alpha - 1} \) a copy of \( \mathbb{R}^n \) with respective unit vectors \( B_j \) and \( \Xi_j \), and for \( T_{2\alpha - 1} \) a copy of \( \mathbb{R}^{n \times n} \) with unit vectors \( B_j \Xi_i \). The structure group \( G \) is taken to be isomorphic to \( \mathbb{R}^n \) and, for \( x \in \mathbb{R}^n \), it acts on \( T \) via
\[
\Gamma_x 1 = 1, \quad \Gamma_x \Xi_i = \Xi_i, \quad \Gamma_x B_i = B_i - x^i 1, \quad \Gamma_x (B_j \Xi_i) = B_j \Xi_i - x^j \Xi_i.
\]

Let now \( (X, X) \) be an \( \alpha \)-Hölder continuous rough path with values in \( \mathbb{R}^n \). In other words, the functions \( X \) and \( X \) are as in the introduction, satisfying the relation (4.11) and the analytic bounds \( |X_t - X_s| \lesssim |t - s|^{\alpha}, |X_{s,t}| \lesssim |t - s|^{2\alpha} \). It turns out that this defines a model for \( \mathcal{S} \) in the following way (recall that \( X_{s,t} \) is a shorthand for \( X_t - X_s \)).

**Lemma 5.5** Given an \( \alpha \)-Hölder continuous rough path \( (X, X) \), one can define a model for \( \mathcal{S} \) on \( \mathbb{R} \) by setting \( \Gamma_{su} = \Gamma_{X_{s,u}} \) and
\[
(\Pi_s 1)(t) = 1, \quad (\Pi_s B_j)(t) = X^j_{s,t},
\]
\[
(\Pi_s \Xi)(\psi) = \int \psi(t) dX^j_{t}, \quad (\Pi_s B_j \Xi)(\psi) = \int \psi(t) dX^j_{s,t}.
\]

Here, both integrals are perfectly well-defined Riemann integrals, with the differential in the second case taken with respect to the variable \( t \). Given a controlled rough path \( (Y, Y) \in C^\alpha_X \), this then defines an element \( \hat{Y} \in D^{2\alpha} \) by setting
\[
\hat{Y}(s) = Y(s) 1 + Y'_s(s) B_i,
\]
with summation over \( i \) implied.
We first check that the algebraic properties of Definition 5.2 are satisfied. It is clear that \( \Gamma_{su} \Gamma_{ut} = \Gamma_{st} \) and that \( \Pi_a \Gamma_{su} \tau = \Pi_{a} \tau \) for \( \tau \in \{1, \mathbb{B}_j, \Xi_i\} \). Regarding \( \mathbb{B}_j \Xi_i \), we differentiate Chen’s relations (4.1) which yields the identity
\[
dX_{s, t}^{i,j} = dX_{u, t}^{i,j} + X_{s, u}^{i} \cdot dX_{t}^{j}.
\]
The last missing algebraic relation then follows at once. The required analytic bounds follow immediately from the definition of the rough path space \( D^\alpha \).

Proof. In the same context as above, let \( \alpha \in (\frac{1}{3}, \frac{1}{2}) \), and consider \( \hat{Y} \in D^{2\alpha} \) built as above from a controlled rough path. Then, the map \( \hat{Y} \Xi_i \) given by
\[
(\hat{Y} \Xi_i)(s) = Y(s) \Xi_i + Y_i'(s) \mathbb{B}_i \Xi_j
\]
belong to \( D^{3\alpha - 1} \). Furthermore, there exists a function \( Z \) such that, for every smooth test function \( \psi \), one has
\[
(\mathcal{R} \hat{Y} \Xi_i)(\psi) = \int \psi(t) dZ(t),
\]
and such that
\[
Z_{s, t} = Y(s) X_{s, t}^j + Y_i'(s) X_{s, t}^{i,j} + O(|t - s|^{3\alpha}).
\]

Proof. The fact that \( \hat{Y} \Xi_i \in D^{3\alpha - 1} \) is an immediate consequence of the definitions. Since \( \alpha > \frac{1}{3} \) by assumption, we can apply the reconstruction theorem to it, from which it follows that there exists a unique distribution \( \eta \) such that, if \( \psi \) is a smooth compactly supported test function, one has
\[
\eta(\psi^\lambda_s) = \int \psi^\lambda_s(t) Y(s) dX_t^j + \int \psi^\lambda_s(t) Y_i'(s) dX_{s, t}^{i,j} + O(\lambda^{3\alpha - 1}).
\]

By a simple approximation argument, it turns out that one can take for \( \psi \) the indicator function of the interval \([0, 1]\), so that
\[
\eta(1_{[s, t]}) = Y(s) X_{s, t}^j + Y_i'(s) X_{s, t}^{i,j} + O(|t - s|^{3\alpha}).
\]

Here, the reason why one obtains an exponent \( 3\alpha \) rather than \( 3\alpha - 1 \) is that it is really \( |t - s|^{-1} \mathbb{I}_{[s, t]} \) that scales like an approximate \( \delta \)-distribution as \( t \to s \).
6 Sine-Gordon

6.1 Background

The Berezinskii-Kosterlitz-Thouless transition (BKT transition, named for condensed matter physicists Vadim Berezinskii, John M. Kosterlitz and David J. Thouless) is a phase transition in a large class of 2D physical systems, such as Josephson junction arrays and thin disordered superconducting granular films. The study of this phase transition is a central topic in condensed matter physics; work on the transition led to the 2016 Nobel Prize in Physics being awarded to Thouless, Kosterlitz and Haldane.

These systems are mathematically described by 2D XY model, Coulomb charge model, or sine-Gordon model. In these models one can talk about vortices and anti-vortices, or, positive charges and negative charges. Roughly speaking the transition is from bound vortex-antivortex pairs (or positive-negative-charge pairs called dipoles) at low temperatures to unpaired vortices and anti-vortices (or positive and negative charges) at high temperatures.
Coulomb charge system in 2D. A Coulomb charge system in 2D consists of \( n \) charges at positions \( (x_i)_{i=1}^n \) on a lattice, each with a sign \( e_i \in \{ \pm 1 \} \). The number of charges \( n \) is also random and it is natural to assume Poisson statistics for it. The partition function (so called grand canonical ensemble partition function, due to uncertainty of \( n \)) is:

\[
Z = \sum_{n \geq 0} \frac{z^n}{n!} \sum_{x_1, \ldots, x_n} e^{-\beta \sum_{i,j=1}^{n} e_i e_j (-\Delta)^{-1}(x_i-x_j)}
\]

where \((-\Delta)^{-1}\) is the Coulomb potential and \( e_i e_j \in \{ \pm 1 \}\) tells us if the two charges are repulsive or attractive.

The famous sine-Gordon transformation (or Siegert-Kac transformation) turns this partition function to a field theory. To see this, let \( f(x) = \sum_i e_i \delta_{x_i}(x) \) be the “charge distribution”. Writing \((\cdot, \cdot)\) for the \(L^2\) product on the lattice, we have

\[
Z = \sum_{n \geq 0} \frac{z^n}{n!} \sum_{x_1, \ldots, x_n} e^{-(\sqrt{\beta} f, (-\Delta)^{-1}\sqrt{\beta} f)} = \sum_{n \geq 0} \frac{z^n}{n!} \sum_{x_1, \ldots, x_n} E e^{i(\sqrt{\beta} f, \Phi)}
\]

where \( \Phi \sim N(0, (-\Delta)^{-1}) \) is the Gaussian free field. Therefore,

\[
Z = \sum_{n \geq 0} \frac{z^n}{n!} \sum_{x_1, \ldots, x_n} E \prod_{j=1}^n e^{i\sqrt{\beta} \delta_{x_j}(x_j)} = E \sum_{n \geq 0} \frac{z^n}{n!} \prod_{j=1}^n \sum_{e_i} e^{i\sqrt{\beta} \delta_{x_j}(x_j)} = E \sum_{n \geq 0} \frac{z^n}{n!} \left( \sum_{x} 2 \cos(\sqrt{\beta} \Phi(x)) \right)^n
\]

This leads to the sine-Gordon quantum field theory, formally written as

\[
\exp \left( -\frac{1}{2} \int |\nabla \Phi(x)|^2 - z \cos(\sqrt{\beta} \Phi(x)) dx \right) \mathcal{D} \Phi.
\]

We have two remarks:

1) In order to pass from lattice to continuum one immediately sees a problem: \((-\Delta)^{-1}(x_i-x_j)\) is infinite if \( x_i = x_j \). Therefore, one should not have self-interaction for each charge (namely, should shift the energy by an overall infinite constant \( n(-\Delta)^{-1}(0) \)). We will see a similar phenomena (i.e. Wick renormalization) when we discuss the SPDE.

2) We didn’t treat boundary conditions very carefully. On the torus, in fact, \((-\Delta)^{-1}\) is ill-defined due to the zero mode. To cure this problem, one can assume that \( f \) is mean zero - and this means that the system is neutral (equal numbers of positive and negative charges). Alternatively, one can take a small mass \( m \) and consider the potential \((m^2 - \Delta)^{-1}\), and then take the limit as \( m \to 0 \) for \( Z \), which will effectively lead to the same neutrality constraint.

Phase transition. There are two phases:

1) At high temperatures (\( \beta \) small), the system is in the “plasma phase” (or “Debye screening phase”): the correlation length is expected to be finite, namely the correlation decays exponentially. This is because with large probability, the long-range electrostatic
interactions generated by a uniform density of particles of opposite charges almost cancel each others, resulting in an effective short range interaction.

2) At low temperatures ($\beta$ large), the system is in the “dipole phase” (or “KT phase”): the correlation length is expected to be infinite. The system favors the formation of pairs of opposite charges, namely the dipoles, and the effective range of the interactions remains long (recall that the Coulomb interaction we start with is a long range one).

Here the correlations often refer to “charge correlations” for two probing charges. Namely, consider two probes: $p_1$, which is a charge $\eta \in (0, 1)$ at the lattice site $x$; and $p_2$, which is a particle of charge $-\eta$ at the lattice site $y$. We denote by $Z_{p_1, p_2}$ the partition function defined as above but with these two extra charges, and define the charge correlation $\varrho_\eta(x, y)$ as (the infinite volume limit of) $Z_{p_1, p_2}/Z$.

It is expected that there is a critical temperature $\beta_c(z)$ depending on $z$, such that if $\beta < \beta_c(z)$, $\varrho_\eta$ decays exponentially, while if $\beta > \beta_c(z)$,

$$\varrho_\eta(x, y) \sim \frac{C}{|x - y|^{2\kappa}}$$

where $\kappa$ is called the correlation critical exponent, which is proportional to $\eta^2$; more precisely,

$$\kappa = \frac{\beta_{\text{eff}}}{4\pi} \eta^2$$

where the so called $\beta_{\text{eff}}$ is a function of $(\beta, z)$. If $\beta = \beta_c(z)$,

$$\varrho_\eta(x, y) \sim \frac{C}{|x - y|^{2\kappa}} (\log |x - y|)^\alpha$$

where $\alpha = \frac{1}{2}$ if $\eta = \frac{1}{2}$ and $\alpha = -\kappa$ otherwise.

The critical value at $z = 0$, namely $\beta_c(0)$ is equal to $8\pi$.

Approaching $\beta_c(z)$ from higher temperatures $\beta \nearrow \beta_c(z)$, the correlation length is expected to diverge as

$$\xi \sim e^{\alpha(z)|\beta - \beta_c(z)|^{-1/2}}$$

Note that this is very different from $\xi \sim |\beta - \beta_c|^{-\nu}$ in the second-order phase transitions (such as in Ising model).

6.2 Dynamical sine-Gordon

In the following we replace the $\sqrt{\beta}$ above by $\beta$ for notational convenience. The sine-Gordon equation reads:

$$\partial_t u = \frac{1}{2} \Delta u + \sin(\beta u) + \zeta,$$  \hspace{1cm} (6.1)

over $\mathbb{R}_+ \times \mathbb{T}^2$, where $\zeta$ denotes space-time white noise.

Equation (6.1) is related to many models of equilibrium statistical mechanics in two dimensions. Most directly it is the natural (Langevin) dynamic for the sine-Gordon (Euclidean) quantum field theory in two spatial dimensions.

[HS16] showed that for $\beta^2 \in (0, 4\pi)$ the methods of Da Prato and Debussche [DPD02, DPD03] suffice for showing well-posedness while for $\beta^2 \in [4\pi, \frac{16\pi}{3}]$ local well-posedness can be obtained via the theory of regularity structures. [CHS18] then use the theory of regularity structures to give a proof of local well-posedness for the entire subcritical regime $\beta^2 \in (0, 8\pi)$.  

Equation (6.1) fails to be classically well-posed as soon as $\beta$ is non-zero. Formally, writing $u = \Phi + v$ where $\Phi$ solves the linear equation (i.e. the equation with $\beta = 0$), $v$ should solve the equation

$$\partial_t v = \frac{1}{2} \Delta v - \frac{i}{2} (e^{i\beta v} e^{i\beta \Phi} - e^{-i\beta v} e^{-i\beta \Phi}) .$$

(6.2)

(Equivalently we can also write $\partial_t v = \frac{1}{2} \Delta v + \sin(\beta \Phi) \cos(\beta v) + \cos(\beta \Phi) \sin(\beta v)$ and the analysis will be the same.)

The field $\Phi$ is, at small scales, a logarithmically correlated Gaussian random field over space-time. In particular, realizations of $\Phi$ are not functions on space-time but rather distributions. However one can still try to give meaning to the expressions $e^{\pm i\beta \Phi}$ by employing Wick renormalization.

**Wick renormalization for $e^{\pm i\beta \Phi}$.** We start by a second moment calculation:

$$E[e^{i\beta \Phi_z(z)} e^{-i\beta \Phi_z(z')} ] = E[e^{i\beta \Phi_z(z)}] e^{-\frac{\beta^2}{2} E[(\Phi_z(z) - \Phi_z(z'))^2]} .$$

(6.3)

In fact we have

$$E[\Phi_z(z) \Phi_z(z')] \sim -\frac{1}{2\pi} \log(|z - z'| + \varepsilon)$$

and in particular

$$E[\Phi_z(z)^2] \sim -\frac{1}{2\pi} \log(\varepsilon) \to \infty$$

which means that $e^{i\beta \Phi_z(z)}$ converges to the trivial limit 0. To obtain a nontrivial limit, we renormalize:

$$:e^{i\beta \Phi_z(z)}: \overset{\text{def}}{=} C_{\varepsilon} e^{i\beta \Phi_z(z)} \quad \text{where } C_{\varepsilon} \overset{\text{def}}{=} e^{\frac{\beta^2}{2} E[\Phi_z(z)^2]} \sim e^{-\frac{\beta^2}{4\pi}}$$

(in certain sense we blow it up to get an “$O(1)$ object”, namely we force its expectation to be 1.) Now re-doing the above calculation, we have

$$E[ :e^{i\beta \Phi_z(z)} : e^{-i\beta \Phi_z(z')} : ] = e^{-\frac{\beta^2}{2\pi} \log(|z - z'| + \varepsilon)} = (|z - z'| + \varepsilon)^{-\frac{\beta^2}{2\pi}} .$$

(6.4)

This function has a well-defined limit as $\varepsilon \to 0$; and, it is space-time integrable for any $\beta^2 < 8\pi$.

The following picture illustrates the above renormalization procedure: we can interpret $e^{i\beta \Phi}$ and $e^{-i\beta \Phi}$ as a positive and a negative charge. In the exponent of (6.3) there are 3 terms: a $(z, z')$ interaction term (black line in the picture), and two self-interaction terms which lead to logarithmic divergence (red lines in the picture). The Wick renormalization precisely cancels the self-interactions.

The two-point correlation formula (6.4) already indicates that $:e^{\pm i\beta \Phi_z}:$ should be a well-defined space-time distribution for any $\beta^2 < 8\pi$, and should have Hölder-Besov regularity $C^\alpha$ for any $\alpha < -\frac{\beta^2}{4\pi}$, but the proof of this via a Kolmogorov argument requires bounds on higher moments.

**Higher order correlations of $:e^{\pm i\beta \Phi_z}:$.** Note that $:e^{i\beta \Phi_z}:$ is not an element of any finite order Wiener chaos, so to obtain its regularity in the limit using Kolmogorov argument we have to analyze its higher order correlations.
Let’s define $J_\varepsilon(z) = (|z| + \varepsilon)^{i\pi}$ so that by (6.4)

$$E[ :e^{i\beta \Phi(z)} : :e^{-i\beta \Phi(z')} :] = J_\varepsilon(z - z')^{-1}$$

and by similar calculation as in (6.4)

$$E[ :e^{i\beta \Phi(z)} : :e^{i\beta \Phi(z')} :] = E[ :e^{-i\beta \Phi(z)} : :e^{-i\beta \Phi(z')} :] = J_\varepsilon(z - z').$$

By these two identities we have

$$E \left[ :e^{i\beta \Phi(z_1)} : :e^{-i\beta \Phi(y_1)} : \ldots :e^{i\beta \Phi(z_N)} : :e^{-i\beta \Phi(y_N)} : \right] = \prod_{\ell < m} J_\varepsilon(z_\ell - z_m) \prod_{n < k} J_\varepsilon(y_n - y_k) \prod_{i,j} J_\varepsilon(z_i - y_j)$$

namely, we have a kernel $J_\varepsilon$ between every two charges of the same sign, and a kernel $J_\varepsilon^{-1}$ between every two charges of opposite signs. The following picture illustrates this formula for $N = 3$

![Diagram](image1)

where each solid line denotes the kernel $J_\varepsilon^{-1}$ and each dashed line denotes $J_\varepsilon$.

One needs a multi-scale argument to analyze the behavior of this function. Roughly speaking, we can show the following sort of bound: the function denoted by the above graph is bounded up to a proportional constant only depending on $N$ by

![Diagram](image2)

which is a factorized product of $N$ factors $J_\varepsilon^{-1}$. The idea of this multi-scale argument is a type of multi-scale clustering, which roughly goes as follows. We gather the charges which are separated from each other by distance or order $2^n$ into a “cluster”. We start from a small enough integer $n$, in which case every cluster has only one charge; we then increase $n$ inductively, and see what happens when two clusters at scale $\sim 2^n$ merge into one cluster at scale $\sim 2^{n+1}$. The following picture illustrates such a merge.

![Diagram](image3)
Here, consider the charge $g$ on the right cluster, and a pair of opposite charges $a, b$ on the left cluster. Since $a, b$ are relatively close, and $g$ is relatively far, the $a - g$ distance and the $b - g$ distance are more or less equal. So $\mathcal{F}(a, g) \cdot \mathcal{F}(b, g)^{-1}$ is bounded by some universal constant. By proceeding $n$ from small to large, we can make such cancellations happen and when $n$ is sufficiently large, we end up with a factorized function as shown in (6.5), with each factor being the two-point correlation.

Using a Kolmogorov argument, one can prove that $e^{\pm i\beta \phi_\epsilon}$ is indeed a well-defined space-time distribution for any $\beta^2 < 8\pi$, and has Hölder-Besov regularity $C^\alpha$ for any $\alpha < -\frac{1}{4\pi}$. We write $\Psi_\pm = e^{\pm i\beta \phi_\epsilon}$.

**Local solution for $\beta^2 < 4\pi$.** Now consider the renormalized equation

$$\partial_t v = \frac{1}{2} \Delta v - \frac{i}{2} (e^{i\beta \epsilon \Psi_+} - e^{-i\beta \epsilon \Psi_-}) .$$

When $\beta^2 < 4\pi$, the renormalized object $\Psi$ is a distribution of regularity above $-1$, so if we assume $v \in C^1$ then classical results in harmonic analysis can be applied to define the products and the two products both have regularity above $-1$. By Schauder’s estimate one concludes that the fixed point map brings $v$ back to $C^1$. This is essentially the same fixed point argument for dynamical $\Phi^4$ equation in $d = 2$.

**Local solution for $\beta^2 \in [4\pi, 8\pi)$.** When $\beta^2 \geq 4\pi$ however, the above argument fails, since $\Psi_\pm$ are more singular. One needs to apply the regularity structure theory. The idea of the theory is that $v$ should locally “behave like” $P \ast \Psi$ where $P$ is the heat kernel. Assuming $v \sim P \ast \Psi$ and iterate this ansatz back to (6.6), one obtains the “next order” behavior of $v$ which is $P \ast (\Psi (P \ast \Psi))$. We then proved that for $\beta^2 < 8\pi$ the product $\Psi \cdot (P \ast \Psi)$ can be defined via renormalization and correlation analysis. This pushes $\beta^2$ to the next threshold $\frac{16\pi}{3}$, at which an even higher order iteration becomes necessary.

In fact there is a sequence of thresholds $\beta^2 = 4\pi, \frac{16\pi}{3}, 6\pi, \ldots \rightarrow 8\pi$ (converging to $8\pi$) such that more relevant “perturbative objects” arising from such iterations need to be controlled as $\beta^2$ crosses each of them (this sequence turns out to be the same as the one found in constructive QFT.) The correlation analysis of these objects is very technical, because, for polynomials of Gaussian fields (arising in previously studied equations such as KPZ) a second moment bound suffices due to results in Malliavin calculus, but for sine-Gordon one needs to estimate arbitrary moments. The analysis for the first two perturbative objects $\Psi$ and $\Psi (P \ast \Psi)$ (which suffices for $\beta^2 < \frac{16\pi}{3}$) already spans over fifty pages in our paper [HS16], using a Coulomb charge system interpretation, a multi-scale analysis and certain cancellation mechanisms; but as $\beta^2$ becomes arbitrarily close to $8\pi$, one has arbitrarily many such perturbative objects.

The paper [CHSTR] manages to use a systematic way to construct all these (arbitrarily many) objects, and thus prove local well-posedness of sine-Gordon equation for any $\beta^2 < 8\pi$. Interestingly, although all the perturbative objects requires renormalizations, all the renormalizations end up being canceled except for the Wick renormalization used to define $\Psi$. 


7 Yang-Mills

7.1 Lattice gauge theory

Yang-Mills is also called gauge theory. For probabilists it is perhaps easier to motivate its definition starting from the lattice gauge theory.

The gauge field $U$ and gauge transformation $g$. A reader may be familiar with some spin models such as Ising model, where a configuration assigns each site a number $\pm 1$, and it has a nearest neighbor interaction. In lattice Yang-Mills, a configuration assigns each edge $(x, y)$ an element $U_{xy}$ in a Lie group $G$ - for simplicity we just assume that it is a unitary matrix. We also assume $U_{xy} = U_{yx}^{-1}$. For the analogy of “nearest neighbor interaction” in this context, we consider a plaquette $p$ with vertices $x_1, x_2, x_3, x_4$ in anti-clockwise order and define

$$U(p) = U_{x_1x_2}U_{x_2x_3}U_{x_3x_4}U_{x_4x_1}.$$ 

(Here, more precisely, a plaquette $p = (x_1, j, k)$ depends on a site $x_1$ and the first moving direction $j$ and the second moving direction $k$, i.e. $x_2 = x_1 + e_j, x_3 = x_1 + e_j + e_k$.)

We will define a probability measure on the space of such configurations; and the main condition to achieve is that we want this probability measure to be invariant under gauge transformations.

A gauge transformation is a transformation which turns one configuration $U$ to another configuration, defined as follows. Given a $G$-valued function $g$ on the lattice sites, and a configuration $U$ (which we recall is a $G$-valued function on the edges),

$$(g \circ U)_{xy} \overset{def}{=} g(x)U_{xy}g(y)^{-1}.$$ 

g is called a gauge transformation. Note that under a gauge transformation $U(p)$ transforms by conjugation $U(p) \mapsto g(x_1)U(p)g(x_1^{-1}).$

Example 7.1 Abelian case: $G = U(1)$. In this case $U_{xy} = e^{iA_{xy}}$ where $A$ is $\mathbb{R}$-valued. We can think of $A$ as the field (rather than $U$). Then $U(p) = e^{iA(p)}$ where $A(p) = A_{x_1x_2} + A_{x_2x_3} + A_{x_3x_4} + A_{x_4x_1}$. In $d = 2$ this is the discrete curl of the vector field $A$. In arbitrary $d \geq 2$ this is the discrete exterior differential of the one-form $A$. In particular, in $d = 4$ (the physical “space-time”), $A$ has 4 components and its exterior differential $dA$ has 6 components - three representing the electric field and three representing the magnetic field.

The gauge transformation $g(x) = e^{if(x)}$ where $f$ is $\mathbb{R}$-valued. We have

$$(g \circ U)_{xy} = e^{if(x)}e^{iA_{xy}}e^{-if(y)}.$$ 

Namely, the gauge transformation acts on $A$ by $A \mapsto A + \nabla f$ where $\nabla$ is the finite difference. Roughly speaking in physics one can only observe the electromagnetic field $dA$, and the field $A$ depends on a particular gauge (i.e. depends on $f$).
**Gauge invariant measures on the space of gauge fields.** A probability measure on the space of such configurations which is invariant under gauge transformations can be defined as follows:

\[
\frac{1}{Z} \prod_p Q(U(p)) dU
\]

where \( \prod_p \) is over all plaquettes, \( Q \) is a real valued “class function” on the Lie group \( G \), i.e. \( Q(yxy^{-1}) = Q(x) \), for all \( x, y \in G \), and \( dU \) is the product Haar measure,\(^{[8]} \) and \( Z \) is a normalization. However, there are other conditions one would like the model to satisfy, for instance as lattice spacing vanishes it should approximate (at least formally) the continuum Yang-Mills theory discussed below.

A canonical choice is the Wilson model

\[
\frac{1}{Z} e^{-\beta S(U)} , \quad S(U) = \sum_p \text{Re}(\text{tr}(I - U(p))) ,
\]

namely, \( Q \) above is chosen as \( Q(x) = e^{-\beta \text{Re}(\text{tr}(I - x))} \) which is a class function (since trace is).

**Coupling with Higgs field.** The above model with only the field \( U \) is called “pure Yang-Mills”; in general one is also interested in models where the field \( U \) is coupled with other fields. One example is the Higgs field.\(^{[9]} \) A Higgs field \( \Phi \) is a \( \mathbb{C}^N \) valued function on the lattice sites, i.e. \( \Phi(x) \in \mathbb{C}^N \) for every lattice site \( x \).

The field \( U \) tells us how to take the derivative for the field \( \Phi \). (This is a discrete analogue of the fiber bundle in differential geometry: \( \Phi \) plays the role of a section of a vector bundle and \( U \) plays the role of a connection.) More precisely, the **covariant derivative** \( D_U \Phi \) along the edge \( (x, y) \) is defined as

\[
U_{xy} \Phi(y) - \Phi(x)
\]

which means that one first “transports” \( \Phi \) from \( y \) to \( x \) and then compare with its value at \( x \).

A gauge transformation \( g \) can also act on the field \( \Phi \), simply by multiplication

\[
(g \circ \Phi)(x) \overset{\text{def}}{=} g(x) \Phi(x) .
\]

The covariant derivative \( D_U \Phi \) has a covariance property: \( D_{g \circ U}(g \circ \Phi) = g \circ D_U \Phi \). The full Yang-Mills-Higgs action can be defined by

\[
S(U, \Phi) = \sum_p \text{Re}(\text{tr}(I - U(p))) + \sum_e |D_U \Phi(e)|^2 + \lambda \sum_x |\Phi(x)|^4
\]

where \( | \cdot | \) is the complex amplitude, and the term \( |D_U \Phi(e)|^2 \) is invariant under gauge transformations due to covariance property; other terms such as \( |\Phi|^2n \) can be also added into the action. If \( U \) were identically equal to identity \( I \), this would be the discrete \( \Phi^4 \) model.

---

\(^{[8]} \) By definition Haar measure \( \mu \) on a Lie group \( G \) is bi-translation-invariant: for all Borel subsets \( S \subset G \) and all \( g \in G \), \( \mu(gS) = \mu(Sg) = \mu(S) \).

\(^{[9]} \) In the 1960s, Higgs proposed that broken symmetry in electroweak theory (which is a gauge theory with \( G = U(1) \times SU(2) \)) could explain the origin of mass of elementary particles (and in particular the W and Z bosons). This so-called Higgs mechanism and the mathematical model is basically what we describe here. Higgs and several other physicists predicted the existence of a new particle, the Higgs boson (which is basically described by the \( \Phi \) field). On 4 July 2012, CERN announced the discovery of the boson at the Large Hadron Collider. Higgs received the Nobel Prize in 2013.
Wilson loop observables. A physical theory should have observables of interest. For lattice gauge theories, the definition of a Wilson loop observable (or sometimes called Wilson loop variable) is very simple. Given a loop $\gamma$ with directed edges $e_1, \cdots, e_m$ (being a “loop” means that the start point of $e_1$ is the same point as the end point of $e_m$), the Wilson loop variable $W_\gamma$ is defined as

$$W_\gamma \overset{\text{def}}{=} \text{tr}(U_{e_1} \cdots U_{e_m}).$$

$W_\gamma$ is gauge invariant: i.e. under a gauge transformation $g$, since $\gamma$ is a loop the product $U_{e_1} \cdots U_{e_m}$ transforms by conjugation.

Important questions. There are several important questions on lattice gauge theories

1. Infinite volume limit (uniqueness or non-uniqueness): the uniqueness (or non-uniqueness) is in general unknown for lattice gauge theories in $d > 2$ when $\beta$ is large.

2. Continuum limit as lattice spacing vanishes.

3. We assume the Lie group $G = U(N)$ here - so there is a question on asymptotic behavior of certain observables as $N \to \infty$, called large $N$ problem.

4. Quark confinement. Wilson [Wil74] argued that the potential between a static quark and antiquark separated by distance $R$ is given by the formula

$$V(R) = -\lim_{T \to \infty} \frac{1}{T} \log E_{\gamma(R,T)}$$

where $\gamma(R,T)$ is a rectangle of length $T$ and breadth $R$. If $V(R) \to \infty$ as $R \to \infty$, the quark-antiquark pair cannot separate beyond a fixed distance - this is the phenomenon called quark confinement. The key to prove this is to show the so called area law: for compact non-Abelian Lie group $G$ and consider any infinite volume limit of $d = 4$ lattice gauge theory, then

$$|E_{\gamma(R,T)}| \leq Ce^{-cRT}$$

where $C, c > 0$ only depend on $\beta$. Osterwalder and Seiler proved the area law at sufficiently small $\beta$. Guth and Fröhlich - Spencer showed that $U(1)$ lattice gauge theory in $d = 4$ is not confining for large $\beta$. For $U(1)$ theory, Göpfert and Mack proved the area law at arbitrary $\beta$ in $d = 3$.

5. Mass gap:

$$E(W_{\gamma_1}W_{\gamma_2}) - E(W_{\gamma_1})E(W_{\gamma_2}) \sim \exp(-d(\gamma_1, \gamma_2)/\xi)$$

where $\xi$ is called the correlation length and $d(\gamma_1, \gamma_2)$ is a suitable Euclidean distance between the two loops. Moreover $\lim_{\beta \to \infty} \xi = \infty$.

7.2 Some geometry backgrounds and Yang-Mills in continuum

Fix a Lie group $G$ with Lie algebra $\mathfrak{g}$. For simplicity we again assume $G = U(N)$ and $\mathfrak{g} = u(N)$, but our discussion can be easily extended to general Lie groups and Lie algebras.

$\mathfrak{g}$-valued differential forms and connections. These are just differential forms (or “forms” for simplicity) for which the coefficients take values in $\mathfrak{g}$. We denote the space of $\mathfrak{g}$ valued $k$-forms by $\Omega^k(\mathfrak{g})$. For example $A(x) = \sum_{j=1}^d A_j(x) \, dx_j$ is a $\mathfrak{g}$-valued 1-form, where
When $N = 1$ i.e. $g$ is abelian, $[A_i, A_j] = 0$ and thus $A \wedge A = 0$.

To connect to the discrete picture, one should think of the discrete field $U$ as $U = e^{\varepsilon A}$ where $A$ is the $g$ valued 1-form here.

**Covariant differentiations on $\Omega^r(C^N)$ and $\Omega^r(g)$**. Given a connection form $A$, for an $C^N$-valued $r$-form $\Phi$ (denoted by $\Phi \in \Omega^r(C^N)$), we define the covariant differentiation

$$d_A \Phi = d\Phi + A \wedge \Phi \in \Omega^{r+1}(C^N).$$

This definition yields a $C^N$-valued $r + 1$-form. For example, for $\Phi \in \Omega^0(C^N)$ one has

$$d_A \Phi = \sum_{j=1}^d d^A_j \Phi dx_j, \quad d^A_j \Phi = \partial_j \Phi + A_j \Phi$$

and for $\Phi \in \Omega^1(C^N)$, namely $\Phi = \sum_j \Phi_j dx_j$, one has

$$d_A \Phi = \sum_{i,j} (\partial_i \Phi_j + A_i \Phi_j) dx_i \wedge dx_j.$$

Note that for a $C^N$ valued function, (7.2) is the continuum analogue of $D_U$:

$$D_U \Phi = \frac{1}{\varepsilon} (U_{xy} \Phi(y) - \Phi(x)) = \frac{1}{\varepsilon} (e^{\varepsilon A_x} \Phi(y) - \Phi(x)) \approx \partial_j \Phi + A_j \Phi$$

The covariant derivative can be extended to act on $B \in \Omega^r(g)$, as follows

$$d_A B = dB + [A, B] \in \Omega^{r+1}(g), \quad [A, B] \overset{\text{def}}{=} A \wedge B - (-1)^r B \wedge A$$

In particular we have $D_A A = dA + [A, A] = F_A + \frac{1}{2} [A, A]$ where $F_A$ is the curvature defined below.

**Curvature operator and curvature form**. The curvature operator $F_A$ is defined as acting by $d_A$ twice on $\Phi \in \Omega^0(C^N)$, i.e.

$$F_A \Phi \overset{\text{def}}{=} d_A d_A \Phi = (d + A \wedge)(d\Phi + A\Phi) = (dA)\Phi - A \wedge d\Phi + A \wedge d\Phi + (A \wedge A)\Phi = (dA + A \wedge A)\Phi \in \Omega^2(C^N)$$

where $dd = 0$ is used. Note that we initially view $d_A d_A$ as a (double) differentiation operator, but it turns out to be a multiplication operator by $dA + A \wedge A$; so this calculation motivates the other definition of curvature $F_A$ as a $g$-valued 2-form

$$F_A \overset{\text{def}}{=} dA + A \wedge A \in \Omega^2(g).$$

20Strictly speaking, a 1-form is just a local representation of a connection; here we just identify them for simplicity of exposition.
More explicitly,
\[ F_A = \sum_{i<j} F_{ij} \, dx^i \wedge dx^j \quad F_{ij} = \partial_i A_j - \partial_j A_i + [A_i, A_j]. \] 
(7.6)
When \( N = 1 \) i.e. \( \mathfrak{g} \) is abelian, \([A_i, A_j] = 0\) and thus \( F_{ij} = \partial_i A_j - \partial_j A_i \) (if \( d = 2 \), this is the curl of \( A \)).

**Yang-Mills action.** The Yang-Mills action on \( T^d \) is the functional
\[ S_{YM}(A) \overset{\text{def}}{=} \int_{T^d} |F_A(x)|^2 \, d^d x \]
where the norm \(|F_A(x)|\) is given by an Ad-invariant inner product on the Lie algebra \( \mathfrak{g} \). Since we restrict to \( \mathfrak{g} = \mathfrak{u}(N) \), and \( \mathbb{C}^N \) is endowed with the standard metric, the inner product on \( \mathfrak{g} \) boils down to Hilbert-Schmidt inner product \( \langle A, B \rangle_\mathfrak{g} = Tr(A^* B) = \sum_{i=1}^N \langle Ae_i, Be_i \rangle_{\mathbb{C}^N} \).

More generally we can define the Yang-Mills-Higgs action
\[ S_{YMH}(A, \Phi) \overset{\text{def}}{=} \int_{T^d} |F_A(x)|^2 + |d_A \Phi|^2 + \frac{1}{4} |\Phi|^4 + \frac{1}{2} m^2 |\Phi|^2 \, d^d x \]
The Euclidean quantum Yang-Mills-Higgs theory on \( T^d \) is informally described as the probability measure
\[ d\mu(A, \Phi) = \frac{1}{Z} \exp \left( -\frac{1}{4g_0^2} S_{YMH}(A, \Phi) \right) \text{D}AD\Phi, \]
where \( A \in \Omega^1(\mathfrak{g}), \Phi \in \Omega^0(\mathbb{C}^N), \text{D}A \) and \( \text{D}\Phi \) are “infinite-dimensional Lebesgue measures”, \( g_0 \) is a positive coupling constant, and \( Z \) is the formal normalizing constant.

**Gauge transformation.** A gauge transformation \( g \) is a \( G \)-valued function. It acts on \( \Phi \in \Omega^r(\mathbb{C}^N) \) and a connection form \( A \) by
\[ g(\Phi) = g\Phi, \quad g(A) = gAg^{-1} + gdg^{-1}. \] 
(7.7)
The following properties are well-known

1. Covariance of covariant derivative: for \( \Phi \in \Omega^r(\mathbb{C}^N) \),
\[ d_{g(A)}(g\Phi) = gd_A\Phi, \] 
(7.8)
and \( ^{21} \) for \( B \in \Omega^r(\mathfrak{g}) \),
\[ d_{g(A)}(gBg^{-1}) = g(d_A B)g^{-1}. \] 
(7.9)
In other words, \( d_A \) is always covariant when it acts on either the bundle for the natural representation of \( G \) on \( \mathbb{C}^N \) or the bundle for the adjoint representation of \( G \) on \( \mathfrak{g} \).

2. Curvature transforms by conjugation:
\[ F_{g(A)} = gF_A g^{-1} \] 
(7.10)
namely, for each \( j, k \), the transformation \( g \) acts on \( F_{jk} \) by conjugation.

3.
\[ d_{g(A)}^* F_{g(A)} = g(d_A^* F_A)g^{-1} \] 
(7.11)
where each side belongs to \( \Omega^1(\mathfrak{g}) \).

By part (1)(2) of this lemma, \( S_{YMH}(A, \Phi) \) is invariant under (7.7).

\(^{21}\) One can view \( T^d \times \mathfrak{g} \) as a Lie Algebra bundle associated with the vector bundle \( T^d \times \mathbb{C}^N \), with \( G \) having the adjoint representation on \( T^d \times \mathfrak{g} \).
7.3 Stochastic quantization

**Stochastic quantization.** The associated Langevin dynamic for $S_{YM}$ is given by

$$\partial_t A = - d_A^* F_A + \xi \quad (7.12)$$

where $d_A^*$ denotes the adjoint of $d_A$. Indeed $d_A^* F_A$ is $(2)$ times the gradient of $S_{YM}$ w.r.t. the above $Ad$-invariant inner product on the Lie algebra $g$: for any $B \in \Omega^1(g)$

$$\langle d_A^* F_A, B \rangle_g = \langle F_A, d_A B \rangle_g.$$  

The variation of $F_A = dA + A \wedge A$ in direction $B$ is, by (7.12), $dB + A \wedge B + B \wedge A$, and by (7.3) this is precisely $d_A B$. So the variation of $S_{YM}(A) = \int_{\mathbb{T}^d} |F_A(x)|^2 dx$ in direction $B$ is $(2$ times) $\langle F_A, d_A B \rangle_g$.

The associated Langevin dynamic for $S_{YMH}$ is given by

$$\partial_t A = - d_A^* F_A + B(d_A \Phi \otimes \Phi) + \xi, \quad (7.13)$$
$$\partial_t \Phi = - d_A^* d_A \Phi - m^2 \Phi - |\Phi|^2 \Phi + \zeta,$$

where the $\mathbf{R}$-linear map $B : \mathbf{C}^N \otimes \mathbf{C}^N \rightarrow g$ is given by $B(u \otimes v) = uv^* - vu^*$ and $u^*$ denotes the conjugate-transpose of $u$. Equivalently, $B$ is characterized by enforcing that, for all $u,v \in \mathbf{C}^N$ and $D \in g$, one has $\langle B(u \otimes v), D \rangle_g = 2Re(\langle u, DV \rangle_{\mathbf{C}^N}.$

To see $(d_A \Phi, \Phi)$ comes from gradient of $\int_{\mathbb{T}^d} |d_A \Phi|^2 dx$, for any $B \in \Omega^1(g)$, the variation of $d_A \Phi = d \Phi + A \Phi$ in direction $B$ is $B \Phi$, so the variation of $\int_{\mathbb{T}^d} |d_A \Phi|^2 dx$ in direction $B$ is $\langle d_A \Phi, B \Phi \rangle_{\mathbf{C}^N}$. On the other hand

$$\langle B(d_A \Phi \otimes \Phi), B \rangle_g = \langle d_A \Phi, B \Phi \rangle_{\mathbf{C}^N}$$

by the above definition of $B$. The other terms in (7.13) can be derived similarly.

The noise $\xi = \sum_0^d \xi_i dx_i$, where each $\xi_i$ and $\zeta$ are $g$ and $\mathbf{C}^N$ valued space-time white noises respectively:

$$E \langle \xi_i, B \rangle_\mathbb{R} \langle \xi_i, D \rangle_\mathbb{R} = \langle B, D \rangle_\mathbb{R} \quad E \langle \zeta, \Psi \rangle_\mathbb{R} \langle \zeta, \Phi \rangle_\mathbb{R} = \langle \Psi, \Phi \rangle_{\mathbf{C}^N}$$

and Dirac correlated in space-time as usual. $\xi_i, \ldots, \xi_d$ and $\zeta$ are independent.

In spatial coordinates (7.13) reads, for $i = 1, \ldots, d$ and with summation over $j$ implicit,

$$\partial_t A_i = \Delta A_i - \partial^2_{ij} A_j + [A_j, 2\partial_j A_i - \partial_i A_j + [A_j, A_i]]$$
$$+ [\partial_j A_i, A_j] + B((\partial_i \Phi + A_i \Phi) \otimes \Phi) + \xi_i \quad (7.14)$$
$$\partial_t \Phi = \Delta \Phi + (\partial_j A_j) \Phi + 2A_j \partial_j \Phi + A_j^2 \Phi - m^2 \Phi - |\Phi|^2 \Phi + \zeta.$$

**(Formal) gauge invariance.** Note that, at a formal level, if $(A, \Phi)$ a solution to (7.13) then, for any fixed (time-independent) gauge transformation $g$ we have that $(A^g, \Phi^g)$ solves (7.13) with $(\xi, \zeta)$ replaced by the rotated noise $(Ad_g \xi, g\zeta)$ which is equal to $(\xi, \zeta)$ in law. Thus we expect gauge covariance in distribution for the dynamic (7.13) under (time-independent) gauge transformations.

Let’s derive this without the Higgs field $\Phi$. To this end we recall that for $B \in \Omega^1(g)$,

$$(D^A)^* B = \sum_{j=1}^d (-\partial_j B_j + [B_j, A_j]), \quad (7.15)$$
Yang–Mills–Higgs heat equation which in coordinates reads
\[(D^A)^* B = \sum_{i=1}^{d} \left( \sum_{j=1}^{d} \left( - \partial_j B_{ji} + [B_{ji}, A_j] \right) \right) dx^i \] (7.16)

where we define \( B_{ji} \equiv -B_{ij} \) for \( j \geq i \). Therefore,
\[ - (d_A^* F_A)_i = - \sum_j \left( - \partial_j F_{ji} + [F_{ji}, A_j] \right) \]
\[ = \sum_j \left( \partial_j (\partial_j A_i - \partial_i A_j + [A_j, A_i]) \right) - \left[ \partial_j A_i - \partial_i A_j + [A_j, A_i], A_j \right] \]
\[ = \Delta A_i - \partial_i \partial_j A_j + \partial_j [A_j, A_i] + [A_j, [\partial_j A_i - \partial_i A_j]] - [[A_j, A_i], A_j] \]
\[ = \Delta A_i - \partial_i \partial_j A_j + [\partial_j A_j, A_i] + [A_j, [2\partial_j A_i - \partial_i A_j]] + [A_j, [A_j, A_i]] . \]

**Donaldson–DeTurck term.** A major problem with (7.13) is the lack of ellipticity in the equations for \( A \), which is a reflection of the invariance of the action \( S_{YMH} \) under the gauge group. This problem can be circumvented using the following observation: if we take any sufficiently regular functional \( (A, \Phi) \mapsto H(A, \Phi) \in C^\infty(T^2, \mathfrak{g}) \) and consider instead of (7.13) the equation
\[ \partial_t A = - d_A^* F_A + d_A H(A, \Phi) + \xi , \]
\[ \partial_t \Phi = - d_A^* d_A \Phi - H(A, \Phi) \Phi - m^2 \Phi - |\Phi|^2 \Phi + \zeta \] (7.17)
then, at least formally, solutions to (7.17) are gauge equivalent in law to those of (7.13) under a time-dependent gauge transformation.

In order to get a parabolic flow for \( A \) in (7.17) a convenient choice of \( H \) is given by \( H(A, \Phi) = - d_A^* A \) which yields the so-called Donaldson–DeTurck–Zwanziger term in the first equation of (7.17). After making this choice our main focus will be the stochastic Yang–Mills–Higgs heat equation which in coordinates reads
\[ \partial_t A_i = \Delta A_i + [A_j, 2\partial_j A_i - \partial_i A_j + [A_j, A_i]] + (\partial_i \Phi + A_i \Phi, \Phi) + \xi_i , \]
\[ \partial_t \Phi = \Delta \Phi + 2 A_j \partial_j \Phi + A_j^2 \Phi - m^2 \Phi - |\Phi|^2 \Phi + \zeta . \] (7.18)

**Example 7.2** In the abelian case, a gauge transformation simply reads \( A \mapsto A + df \) for a function \( f \). Dropping the Higgs field \( \Phi \), the equation (7.13) becomes
\[ \partial_t A = - d^* dA + \xi , \]
and the equation (7.17) becomes
\[ \partial_t B = - d^* dB - d d^* B + \xi = \Delta B + \xi . \]

One has \( B = A + df \) (so \( A, B \) are gauge equivalent) where \( f \) satisfies
\[ \partial_t f = - d^* B = - d^* A - d^* df . \]
(Formal) gauge covariance. (7.17) also has a gauge covariance property, as long as the gauge transformation is allowed to depend on time. Namely, if \((A, \Phi)\) is a solution to (7.17) then, given a time evolving gauge transformation \(g\) solving
\[
(\partial_t g)g^{-1} = H(A, \Phi) - \text{Ad}_g H(A, \Phi), \quad g(0) = g_0 ,
\]
then \((A^g, \Phi^g)\) solves (7.17) with \((\xi, \zeta)\) replaced by \((\text{Ad}_g \xi, g \zeta)\) which, if \(g\) is adapted, is again equal in law to \((\xi, \zeta)\).

**Example 7.3** In the abelian case, dropping the Higgs field \(\Phi\), consider the parabolic equation
\[
\partial_t A = -d^* dA - dd^* A + \xi = \Delta A + \xi .
\]
If \(f\) satisfies
\[
\partial_t f = -d^* df = \Delta f
\]
then \(A \overset{\text{def}}{=} A + df\) also satisfies the parabolic heat equation
\[
\partial_t A = \partial_t A + d\partial_t f = -d^* dA - dd^* A - dd^* df + \xi = \Delta A + \xi .
\]

To construct the solutions to (7.17) driven by white noises, there are two ways to proceed:

1) Study the Langevin dynamics of Wilson’s lattice gauge theory. This will be along the line of the discussion in the previous section. The advantage of this approach is that this lattice discretization preserves the gauge symmetry, and the aim is to prove that with only a mass renormalization for \(\Phi\) (but not a mass renormalization for \(A\) since it breaks gauge symmetry), a continuum limit can be obtained. The limit is expected to be the “right” (i.e. gauge invariant) limit (with the mass for \(\Phi\) being the only degree of freedom) since the sequence of regularizations preserve exact gauge symmetry.

2) Study a continuum version of the problem by a smooth regularization \((\xi_\varepsilon, \zeta_\varepsilon)\) of the white noises. This would certainly avoid lots of technical complexities caused by the lattice, but since the exact gauge symmetry is in general broken with such smooth regularizations, the challenge is then to demonstrate a way in which one can still obtain the “right” limit as \(\varepsilon \to 0\).

While we are working on both approaches, we focus on approach 2) in the following exposition. We denote by \(J_\varepsilon\) a smooth mollification operator \((J_\varepsilon \to id\ as \ \varepsilon \to 0)\), for instance convolution with an approximate Dirac distribution. We can prove the following results over short time interval \([0, T]\) for random time \(T > 0\). To lighten the notation and illustrate the main idea, we only state the following results without the field \(\Phi\). Consider the following regularized, mass renormalized equation:
\[
\partial_t A^\varepsilon = -D_{A^\varepsilon}^* F_{A^\varepsilon} - D_{A^\varepsilon} D_{A^\varepsilon}^* A^\varepsilon - C_{\varepsilon} A^\varepsilon + J_\varepsilon \xi . \tag{7.20}
\]

**Theorem 7.4** (7.20) Let \(d = 2, 3\). There is a choice of constant \(C_{\varepsilon}\) such that

1) The solutions \(A^\varepsilon\) to (7.20) starting from initial data \(A_0\) converge as \(\varepsilon \to 0\).

2) Let \(g_0 \in \mathcal{G}\). Then there is a process \(g_\varepsilon : [0, T] \to \mathcal{G}\) such that \(g_\varepsilon(t)\) and \(A^\varepsilon \circ g_\varepsilon(t)\) satisfy the following coupled equations
\[
\begin{align*}
g_\varepsilon \partial_t g_\varepsilon^{-1} &= -D_{A^\varepsilon}^* (g_\varepsilon dg_\varepsilon^{-1}) \\
\partial_t A^\varepsilon &- \quad -D_{A^\varepsilon} F_{A^\varepsilon} - D_{A^\varepsilon} D_{A^\varepsilon}^* A^\varepsilon + g_\varepsilon (J_\varepsilon \xi) g_\varepsilon^{-1} - C_{\varepsilon} (A^\varepsilon - g_\varepsilon dg_\varepsilon^{-1}) \tag{7.21}
\end{align*}
\]
with initial data \( g_0 \) and \( A_0^g = g_0 \circ A_0 \). Moreover, \( (g_\varepsilon, A^{g_\varepsilon}) \) converges to some limit \( (g, A^g) \) as \( \varepsilon \to 0 \).

3) Let \( \tilde{A}^\varepsilon \) solve the same equation (7.20) but with initial data \( A_0^g \). By part 1), \( \tilde{A}^\varepsilon \) converges to some limit denoted by \( \tilde{A} \) as \( \varepsilon \to 0 \). Then for each \( t \in (0, T) \),

\[
A^g(t) \overset{\text{law}}{=} \tilde{A}(t).
\]

We make some remarks on the significances of the above statements.

(1) The second statement is reminiscent to the aforementioned DeTurck trick, but different in the sense that it links two parabolic equations with different initial data, whereas the standard DeTurck trick links the (non-parabolic) Yang-Mills equation with the gauge fixed parabolic equation.

(2) The mass renormalization \(-\bar{C}_\varepsilon A^g_{\varepsilon} \) in (7.20) is indeed not gauge invariant: after a gauge transformation, the renormalization becomes \(-\bar{C}_\varepsilon (A^{g_\varepsilon}_{\varepsilon} - g_\varepsilon d g_{\varepsilon}^{-1}) \). The origin of the problem is that the smooth mollification operator \( J_\varepsilon \) breaks gauge symmetry. (In Wilson’s lattice setting one would expect that the mass renormalization for \( A \) should cancel out.)

(3) For the \( g \)-valued space-time white noises \( \xi \) and any \( g_\varepsilon \in \mathcal{G} \), one has \( g_\varepsilon \xi g_{\varepsilon}^{-1} \overset{\text{law}}{=} \xi \), i.e. the gauge invariance (in law). However, this is not true for \( J_\varepsilon \xi \). In (7.21), the term \( g_\varepsilon (J_\varepsilon \xi) g_{\varepsilon}^{-1} \) is not simply distributed as \( J_\varepsilon \xi \) and the products between \( g_\varepsilon \) and \( J_\varepsilon \xi \) lack classical meaning in the limit. This shows that the term \( g_\varepsilon (J_\varepsilon \xi) g_{\varepsilon}^{-1} \) also needs renormalization. It turns out that the extra renormalization \( C_\varepsilon g_\varepsilon d g_{\varepsilon}^{-1} \) precisely does the job.

(4) The last statement states a “gauge equivariance” property. It roughly states that solutions starting from gauge equivalent initial data remain gauge equivalent for positive time. This allows us to construct gauge invariant dynamical observables.

(5) One may wonder if there is any freedom for the choice of \( C_\varepsilon \) and which choice leads to the “right” limit. The most interesting point of the above theorem is that (7.21) needs renormalizations of the form \(-\bar{C}_\varepsilon A^{g_\varepsilon}_{\varepsilon} \) and \(-\bar{C}_\varepsilon g_\varepsilon d g_{\varepsilon}^{-1} \) for separate reasons, so a priori \( \bar{C}_\varepsilon \) and \( \tilde{C}_\varepsilon \) could be two unrelated constants; however they turn out to be the same \( C_\varepsilon \), and there is only one choice of it such that this is true!

References


