1 Stochastic heat equation with additive noise

1.1 Explicit solution


1.2 Properties of solution

Besov regularity. Wavelets and Hölder regularity. Gaussian free field as invariant measure.

1.3 Examples and challenges of nonlinear SPDEs

2 Stochastic heat equation with multiplicative noise

2.1 Itô solution

Itô integral. Existence and uniqueness of solution.

2.2 Weak solution

3 \Phi^4 equation

3.1 Local mild solution in \( d = 1 \)

Fixed point argument.

3.2 Important heuristics: perturbation theory

Wick theorem. Feynman diagrams.

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

4 (Controlled) rough paths

4.1 Fixed point argument via rough paths

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**\(^1\) 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.\(^2\)

A space-time white noise \( \xi \) is a white noise on a space-time domain \( \Lambda \subset \mathbb{R} \times \mathbb{R}^d \) with

\[
E[\xi(f)\xi(g)] = \int_\Lambda f(t,x)g(t,x)dtdx , \quad \text{or} \quad E[\xi(t,x)\xi(s,y)] = \delta(t-s)\delta(x-y) . \quad (1.2)
\]

One example of white noise on \( \mathbb{R}_+ \) is \( \frac{dB(t)}{dt} \), 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. \( L^2(T^d) \)) if we are interested in white noise on \( T^d \), let \((e_i)_{i \in \mathbb{N}}\) be an orthonormal basis. We then define

\[
W(t) = \sum_{k \in \mathbb{N}} b_k(t)e_k \quad (1.3)
\]

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

\[
E[(W(t),f)_H\langle W(t),g \rangle_H] = \sum_{k,\ell} E\langle \delta_k(t)\delta_\ell(s)\langle e_k,f \rangle_H\langle e_\ell,g \rangle_H = (t \wedge s)\sum_k \langle e_k,f \rangle_H\langle e_k,g \rangle_H = (t \wedge s)\langle f,g \rangle_H .
\]

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 \( E\langle W(t),W(t) \rangle_H = t\sum_k \langle e_k,e_k \rangle = \infty \). 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 \( E\langle W(t),W(t) \rangle_H' = t\sum_k \frac{1}{k^2} \langle e_k,e_k \rangle_H < \infty \). (In general, it suffices to have \( \iota : H \hookrightarrow H' \) Hilbert-Schmidt \textit{i.e.} \( \iota \iota^* \) trace class; and \( H \) is the Cameron-Martin space for \( H' \) with a Gaussian measure of covariance \( \iota i \iota^* \) 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

\[\text{Here a distribution refers to a generalized function, as opposed to “probability distribution”}.\]

\[\text{Recall its definition } \delta_n(f) \overset{\text{def}}{=} f(\alpha) \text{ and } \delta \overset{\text{def}}{=} \delta_0. \text{ The notation } \delta(x) \text{ is formal, but it will be convenient to use this notation and write for instance } f \delta(x-y)f(y)dx = f(x). \text{ The formal definition of } \zeta \text{ in terms of } \delta \text{ is “justified” by } E[\zeta(f)\zeta(g)] = \iint E[\zeta(x)\zeta(y)]f(x)g(y)dxdy = \iint \delta(x-y)f(x)g(y)dxdy = \int f(x)g(x)dx.\]
that its “scaling dimension” $[\delta] = [x]^{-d}$. According to this heuristic, the space-time white noise $\xi$ then has “scaling dimension” $[\xi] = [t]^{-\frac{1}{2}}[x]^{-\frac{d}{2}}$. In fact, one has

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

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) $(t, x) \to (\lambda^2t, \lambda x)$. A heuristic dimension counting shows that if the three terms in (1.1) have the same scaling dimension $[\partial_t u] = [\Delta u] = [\xi] = [t]^{-\frac{1}{2}}[x]^{-\frac{d}{2}} = [x]^{-\frac{d+2}{2}}$, then $[u] = [x]^{-\frac{d+2}{2}}$. More precisely, if

$$\tilde{u}(t, x) \overset{\text{def}}{=} \lambda^{\frac{d-2}{2}} u(\lambda^2 t, \lambda x) \quad (1.4)$$

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, $e^{-\delta t}1_{|x| \leq \varepsilon/2}$ is the right approximation for $\delta$ as $\varepsilon \to 0$, as can be checked by testing against $f \in \mathcal{C}_c^\infty$. 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, $e^{-\frac{1}{2}}\zeta(\varepsilon^{-1}x)$ converges to $\xi$ in law.

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

$$u(t, x) = \int_{\mathbb{R}^d} \int_0^t P(t - s, x - y)\xi(s, y)dyds + \int_{\mathbb{R}^d} P(t, x - y)u_0(y)dy \quad (1.5)$$

where $P$ is the heat kernel

$$P(t, x) = (4\pi t)^{-\frac{d}{2}}e^{-\frac{|x|^2}{4t}}. \quad (1.6)$$

We call this the mild solution to (1.1). 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.

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

$$\mathbb{E}[u(t, x)u(\bar{t}, \bar{x})] = \int_{\mathbb{R}^d} \int_0^t P(t - s, x - y)P(\bar{t} - s, \bar{x} - y)dyds.$$

In particular the “variance”

$$\mathbb{E}[u(t, x)^2] = \int_{\mathbb{R}^d} \int_0^t P(t - s, x - y)^2dyds = \int_0^t \frac{1}{(8\pi(t - s))^\frac{d}{2}} \int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{2(2\pi(t - s))^\frac{d}{2}}}dyds$$

$$= \int_0^t (8\pi(t - s))^{-\frac{d}{2}}ds < \infty \quad \text{if and only if} \quad d = 1.$$
This could also follow by Fourier transform (in \(x\)) and Parseval’s theorem:

\[
\int_0^t |\hat{P}(t - s, k)|^2 ds = \int_0^t e^{-2(t-s)|k|^2} ds = \frac{1 - e^{-2t|k|^2}}{2|k|^2}
\]

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 \(\| \int_{\mathbb{R}^{d+1}} P(t-\cdot, x-\cdot) f(t, x) dt dx \|_{L_2}^2\) which is always finite for \(f \in C_c^\infty\). 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 \(\hat{u}(k) = \hat{u}(-k)\). 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 \(\beta_k = \bar{\beta}_{-k}\). 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\(^4\) SDEs

\[
d\hat{u}(k) = -k^2 \hat{u}(k)dt + d\beta_k .
\]

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 \(k_1, k_2 \neq 0\))

\[
\mathbb{E}[\hat{u}(t, k_1)\hat{u}(t, -k_2)] = 1_{k_1 = k_2} \int_0^t e^{-2(t-s)|k|^2} ds = 1_{k_1 = k_2} \frac{1 - e^{-2|k_1|^2}}{2|k_1|^2}
\]

where we have used Itô isometry.

### 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 \(\text{supp}(\chi) \cap \text{supp}(\varrho(2^{-j}\cdot)) = \emptyset\) for \(j \geq 1\) and \(\text{supp}(\varrho(2^{-1}\cdot)) \cap \text{supp}(\varrho(2^{-j}\cdot)) = \emptyset\) for \(|i-j| \geq 2\). See [] (Proposition 2.10) for existence of such functions. We will write \(\varrho_{-i} = \chi\) and \(\varrho_j = \varrho(2^{-j}\cdot)\) for \(j \geq 0\).

The Littlewood–Paley blocks are now defined as

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

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

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

\(^4\)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 \(\hat{u}(k) = \hat{u}(-k)\).
with the usual interpretation as $\ell^{\infty}$ norm in case $q = \infty$. The space $B^{\alpha}_{p,q}$ does not depend on $(\chi, \varrho)$. We write $C^{\alpha} = B^{\alpha}_{\infty,\infty}$. 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

$$\mathbb{E}\|u_t - u_s\|_{B^{\gamma}_{p,p}}^p \leq C|t - s|^{\delta p/2}.$$  

From this, together with Kolmogorov continuity theorem, and a continuous Besov imbedding $B^{\alpha}_{p,p} \hookrightarrow B^{\alpha - \frac{d}{2}}_{\infty,\infty}$, by taking $p$ large enough we have$^5$

$$u \in C([0, T], B^{\gamma_{-\varepsilon}}_{\infty,\infty}) \quad a.s. \quad \forall \varepsilon > 0. \quad (1.9)$$

To illustrate the main idea, we assume that the underlying space is the torus $T^d$. We will only prove a simpler version of Lemma 1.2:

$$\mathbb{E}\|u(t, \cdot)\|_{B^{\alpha}_{p,p}}^p \leq C \quad \forall \alpha < \gamma. \quad (1.10)$$

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

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

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

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

Now everything boils down to a second moment calculation of Gaussian: by (1.8)

$$\mathbb{E} \left| \sum_{k \in \mathbb{Z}^d} \varrho_j(k) \hat{u}_t(k) e^{ikx} \right|^2 = \sum_{k_1, k_2 \in \mathbb{Z}^d} \varrho_j(k_1) \varrho_j(k_2) \mathbb{E}[\hat{u}_t(k_1)\hat{u}_t(-k_2)] e^{ik_1 x - ik_2 x} = \sum_{k \in \mathbb{Z}^d} \varrho_j(k)^2 \frac{1 - e^{-2|k|^2}}{2|k|^2} \asymp 2^{-2j^d} = 2^{j(d-2)}$$

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

$$\sum_{j \geq -1} 2^{j \alpha p} \mathbb{E}\|\Delta_j u\|_{L^p}^p \asymp \sum_{j \geq -1} 2^{j \alpha p} (2^{j(d-2)})^\frac{p}{2} < \infty \quad \Leftrightarrow \quad \alpha < \gamma = -\frac{d-2}{2}$$

as required by (1.10). \[\square\]

$^5$more precisely, we can find a version of $u$ which is continuous in $t$
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 \( T^d \). Write \( \Lambda = [0, T] \times T^d \). We first introduce some notation. For \( z \in \Lambda \) we define a parabolic distance \( \| z \| = \sqrt{t} + |x| \). For \( \varphi \in C^r_c(\Lambda) \) with some \( r > 0 \), and \( \lambda \in (0, 1) \), we define

\[
\varphi^\lambda_{(s,y)}(t, x) \overset{\text{def}}{=} \lambda^{-(d+2)} \varphi(\lambda^{-2}(t - s), \lambda^{-1}(x - y))
\]

(1.11)

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

For \( \alpha \in \mathbb{R}_+ \backslash \mathbb{Z}_+ \), let \( C^\alpha_c \) be the completion of \( C^\infty_c \) under

\[
\| f \|_{C^\alpha} = \sum_{|k| < |\alpha|} \sup_{z \in \Lambda} |D^k f(z)| + \sup_{z \in \Lambda} \frac{|D^k f(z) - D^k f(\bar{z})|}{\| z - \bar{z} \|^{\alpha - |\alpha|}}
\]

where \( |k| = 2k_0 + k_1 \) for \((k_0, k_1) \in \mathbb{N} \times \mathbb{N}^d \). When \( \alpha < 0 \), the space \( C^\alpha_c \) is defined as the completion of \( C^\infty_c \) with respect to

\[
\| f \|_{C^\alpha} = \sup_{\lambda \in (0, 1)} \sup_{z \in \Lambda} \sup_{\varphi} \lambda^{-\alpha} |f(\varphi^\lambda_z)|
\]

(1.12)

where \( \sup_{\varphi} \) is over all functions \( \varphi \) which have \( \| \varphi \|_{C^0} \leq 1 \) for \( r_0 = -|\alpha| \) and supported in a unit ball. For \( \alpha < 0 \) it can be shown that \( C^\alpha_c \) 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

\[
\mathbb{E}[\xi(\varphi^\lambda_z)^2] = \int_\Lambda (\varphi^\lambda_z(w))^2 dw \lesssim \lambda^{-2d} \int_{\| w - z \| \leq \lambda} dw = \lambda^{-2d} \lambda^{d+2} = \lambda^{-(d+2)}
\]

(1.13)

where we applied a brutal bound on \( \varphi^\lambda_z(w) \) 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 \( \xi \in C^\alpha_c \) for any \( \alpha < -\frac{d+2}{2} \), by showing \( \mathbb{E}\| \xi \|_{C^\alpha_c}^p < \infty \), 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 \( \Lambda_n \overset{\text{def}}{=} ((2^{-2n} \mathbb{Z}) \times (2^{-n} \mathbb{Z}))^d \cap \Lambda \). Given \( \varphi \) and \( n \in \mathbb{Z}_+ \) we write

\[
\varphi^n_{(s,y)}(t, x) \overset{\text{def}}{=} 2^n \varphi(2^{dn}(t - s), 2^n(x - y))
\]

(1.14)

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 \( \{ \varphi^n_{z \in \Lambda_n} \} \cup \{ \psi^n_{z \in \mathbb{Z}_+, z \in \Lambda_n} \} \) 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

1. For each $k \in \mathbb{Z}$, $\int_{\mathbb{R}} \varphi(x)\varphi(x + k)dx = 1_{k=0}$;

2. there exist “structure constants” $a_k$ such that $\varphi(x) = \sum_{k \in \mathbb{Z}} a_k \varphi(2x + k)$.

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 $\psi(x) = \sum_{k \in \mathbb{Z}} b_k \varphi(x + k) \in C^r$ for some constants $b_k$ such that $\hat{V}_n = \text{span}\{\psi^n_x : x \in \Lambda_n\}$. Therefore, we have an $L^2$ decomposition $L^2(\mathbb{R}) = V_0 \oplus \hat{V}_0 \oplus \hat{V}_1 \oplus \hat{V}_2 \oplus \cdots$, with an orthonormal basis as in Theorem 1.3.

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

**Theorem 1.4** Let $\alpha < 0$. $f \in C^\alpha$ if and only if

$$|f(\psi^0_z)| \lesssim 2^{-\frac{d+2m}{2} - \alpha z} |f(\varphi^0_y)| \lesssim 1$$

where the proportional constants in $\lesssim$ are uniform in $n \in \mathbb{Z}_+, z \in \Lambda_n, y \in \Lambda_0, \psi \in \Psi$.

**Proof.** We only prove necessity (sufficiency is harder). Since $|f(\varphi^0_y)| \lesssim \lambda^\alpha$ for any $\lambda \in (0, 1), z \in \Lambda$ and $\varphi \in C^\alpha$ 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

$$f(\psi^\lambda_z) = 2^{\frac{d+2}{2}} f(\psi^n_z)$$

from which (1.15) follows. We refer to [Hai14, Proposition 3.20] for a complete proof. \qed

Thanks to Theorem 1.4, we can now bound $E||\xi||_{C^\alpha}$:

$$E||\xi||_{C^\alpha}^p = E\left[ \sup_{\psi} \sup_{n \geq 0} \sup_{z \in \Lambda_n} (2^{\frac{d+2m}{2} + n\alpha} \xi(z))^p \right] \lesssim \sum_{\psi} \sum_{n \geq 0} \sum_{z \in \Lambda_n} 2^{\frac{d+2m}{2} + n\alpha} E[\xi(z)^p]$$

where $\psi$ ranges in the finite collection $\Psi \cup \{\varphi\}$ given in Theorem 1.3. This is finite if and only if $\alpha < -\frac{d+2}{2} - \frac{d+2}{p}$. Therefore, given $\alpha = -\frac{d+2}{2} - \varepsilon$ with $\varepsilon > 0$, one can choose $p$ large enough so that $-\varepsilon < -\frac{d+2}{p}$, which ensures $E||\xi||_{C^\alpha}^p < \infty$. \footnote{\textsuperscript{3}It is easy to find a discontinuous one, e.g. the Haar wavelet, but finding one in $C^\alpha$ is very nontrivial.\textsuperscript{4}A student in the class asked that given $\alpha = -\frac{d+2}{2} - \varepsilon$ with $\varepsilon > 0$, why it happens that $E||\xi||_{C^\alpha}^p < \infty$, 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 $E||\xi||_{C^\alpha}^p < \infty$ for a sufficiently large $p$ we do have $E||\xi||_{C^\alpha}^{p'} < \infty$ 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 $2^{n(d+2)}$; if we did not have this factor $2^{n(d+2)}$, we would be able to conclude that $\sum_{n \geq 0} 2^{\frac{d+2m}{2} + n\alpha} < \infty$ if and 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 $E||\xi||_{C^\alpha}^p < \infty$.} Here, an important input of the
bound is $E[\xi(\psi^n_x)^p] \lesssim E[\xi(\psi^n_x)^2]^{p/2} = \|\psi^n_x\|_{p/2}^p \lesssim 1$, 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 $E[\zeta(\varphi^n_x)^2] \lesssim \lambda^{2\alpha}$ for some $\tilde{\alpha} < 0$ then $\zeta \in C^\infty_{\lambda\alpha}$ for any $\alpha < \tilde{\alpha}$.

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

**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 $\mathcal{N}(0, \frac{1}{|k|^2})$ as an invariant measure, namely $e^{-\frac{1}{2}|k|^2|a(k)|^2}/Z$ (where $Z$ is the suitable normalization).

The Gaussian free field $\Phi$ on $T^d$ is a random distribution, with $\Phi(0) = 0$ (namely $\Phi(1) = \int_{T^d} \Phi = 0$) which is given by

$$\Phi = \sum_{k \in \mathbb{Z}^d \setminus \{0\}} \frac{a_k}{|k|} e^{ik \cdot x}$$

(1.18)

where $\{a_k\}$ are independent complex standard Gaussians, except for a constraint $a_k = \bar{a}_{-k}$, namely $\text{Re}(a_k), \text{Im}(a_k) \sim \mathcal{N}(0, \frac{1}{2})$ independent s.t. $E|a_k|^2 = 1$. Since

$$E|\Phi(k)|^2 = \frac{E|a_k|^2}{|k|^2} = \frac{1}{|k|^2}$$

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 $\int_{T^d} f = \int_{T^d} g = 0$,

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

This leads to the formal notation of Gaussian free field in some literature $\Phi \sim e^{-\frac{1}{2} f(\nabla \Phi)^2} dx$. We refer to [She07, Section 2] for a more systematic discussion on Gaussian free field (via “abstract Wiener space” approach and “Gaussian Hilbert space” approach).

The stationary solution to the SHE (1.1) then reads

$$\hat{u}_t(k) = \int_{-\infty}^t e^{-(t-s)|k|^2} d\beta_k(s) \quad \text{or} \quad u(t, x) = \int_{-\infty}^t \int_{T^d} p(t-s, x-y)\xi(s, y) ds dy$$
Hölder regularity for stationary solution. Recall the parabolic distance \( \|z\| \) defined above. We have the following result about the degree of singularity of heat kernel at origin:

\[
|P(z)| \lesssim \|z\|^{-d} \quad \text{for} \quad \|z\| \leq \frac{1}{2}
\]

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

\[
P(t, x) = (4\pi t)^{-\frac{d}{2}} P(1, x/\sqrt{t}) \lesssim |t|^{-\frac{d}{2}} e^{-\frac{|x|^2}{4t}} \lesssim |t|^{-\frac{d}{2}} \lesssim \|z\|^{-d}
\]

and when \( \sqrt{t} \leq |x| \), one has

\[
P(t, x) = \frac{|x|^d}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}} \lesssim \frac{|x|^d}{(4\pi t)^{\frac{d}{2}}} (\frac{|x|^2}{4t})^{-d/2} \lesssim 1,
\]

so

\[
P(t, x) = \begin{cases} |x|^{-d} P(t/x^2, 1) & \text{if } |x|^{-d} \lesssim \|z\|^{-d}. \end{cases}
\]

Another useful result is:

**Lemma 1.6** If \( f \) and \( g \) are functions on space-time 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 \( |(f * g)(z)| \lesssim \|z\|^{\alpha + \beta + (d + 2)}. \)

**Proof.** \((f * g)(z) = \int f(z - w)g(w)dw.\) Let \( r \) be defined as \( \frac{\|z\|}{2}.\) We consider three regimes in space-time. First, \( \|w\| < r, \) in which case \( \|z - w\| > C\|z\| \) for some \( C > 0, \) so we get

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

The second regime is

\[\ldots\]

We can now prove \( u \in C^\alpha_s \) for any \( \alpha < -\frac{d+2}{2}. \) Proof: TO TYPE...

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 ubiquitously found 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 \( (\partial_x^2 H)^2 \)).

\[
\partial_t H = \partial_x^2 H + (\partial_x H)^2 + \zeta.
\]

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 \( d \geq 2.\)

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

\[
\partial_t u = \Delta u + f(u)\xi. \tag{1.19}
\]
The specialization \( f(u) = u \), i.e.
\[
\partial_t u = \Delta u + u\xi \tag{1.20}
\]
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 (1.20). 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\(^\text{10}\); 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
\[
\partial_t u = \Delta u + f(u)\zeta \tag{1.21}
\]
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 [K16] for further discussion and references regarding long time behaviors of PAM.

**Stochastic Navier-Stokes equation.** This is a vector valued equation:
\[
\partial_t \vec{u} + \vec{u} \cdot \nabla \vec{u} = \Delta \vec{u} - \nabla p + \vec{\zeta}, \quad \text{div} \ \vec{u} = 0 \tag{1.22}
\]
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 $\xi$: 

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

(1.23) 

Here $\frac{\delta \mathcal{H}(\Phi)}{\delta \Phi}$ is the variational derivative of the functional $\mathcal{H}(\Phi)$; for instance, when $\mathcal{H}(\Phi) = \frac{1}{2} \int (\nabla \Phi)^2dx$ is the Dirichlet form, $\frac{\delta \mathcal{H}(\Phi)}{\delta \Phi} = -\Delta \Phi$ and (1.23) 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” (1.23) is that given a Hamiltonian $\mathcal{H}(\Phi)$, the formal measure

$$\frac{1}{Z} e^{-\mathcal{H}(\phi)} D\Phi$$

(1.24)

is formally an invariant measure for Eq. (1.23). Here $D\Phi$ is the formal Lebesgue measure and $Z$ is a “normalization constant”. We emphasize that (1.24) 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 $u \in C([0, T], H^{-\frac{d+2-\varepsilon}{2}})$ 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 $H^{-\frac{d-2}{2}}$, or $B_{\infty, \infty}^{-\frac{d-2}{2}-\varepsilon}$, for any $\varepsilon > 0$.

5. Let $\zeta$ be the spatial white noise in (1.21), that is, $\zeta$ is constant in time, and $\mathbb{E} \zeta(x)\zeta(y) = \delta(x-y)$. Prove using similar criteria as in Lemma 1.5 that $\zeta$ has Hölder regularity below $-\frac{d}{2}$.

\[\text{Being invariant means that if the initial condition of (1.23) is random with “probability law” given by (1.24), then the solution at any } t > 0 \text{ will likewise be distributed according to this same “probability law”. For readers familiar with stochastic ordinary differential equations, one simple example is given by the Ornstein-Uhlenbeck process } dX_t = -\frac{1}{2} X_t dt + dB_t, \text{ where } B_t \text{ is the Brownian motion, and its invariant measure is the (one-dimensional) Gaussian measure } \frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}} dX.\]
We call $Z$ a mild solution if
\[
Z(t, x) = \int_{\mathbb{R}^d} \int_0^t P(t-s, x-y)Z(s, y)\xi(s, y)dy ds + \int_{\mathbb{R}^d} P(t, x-y)u_0(y)dy \tag{2.2}
\]
In this, we must have that $\int_0^t \int_{\mathbb{R}} P^2(t-s, x-y)E[Z^2(s, y)]dy ds < \infty$ for the Itô integrals to make sense and be finite.

### 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 $\varphi \in C_c^\infty(\mathbb{R})$ we can define $\int_{\mathbb{R}_+ \times \mathbb{R}} 1_{(0,t)}(s)\varphi(x)\xi(s, x)dx ds$, 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 $(t_1 \wedge t_2) \int \varphi^2(x)dx$.

Let $\mathcal{F}_0 = \emptyset$ and for each $t > 0$ define $\mathcal{F}_t$ to be the $\sigma$–field generated by
\[
\left\{ \int_{\mathbb{R}_+ \times \mathbb{R}} 1_{(0,s]}(u)\varphi(u)\xi(u, x)dx du : 0 \leq s \leq t, \varphi \in C_c^\infty(\mathbb{R}) \right\}.
\]
It is clear that $\mathcal{F}_t$ is a filtration. As the next step, we consider “piece-wise constant” processes. Let $S$ be the set of functions of the form
\[
f(t, x, \omega) = \sum_{i=1}^n X_i(\omega)1_{(a_i, b_i]}(t)\varphi_i(x),
\]
where $X_i$ is a bounded $\mathcal{F}_{a_i}$–measurable random variable and $\varphi_i \in C_c^\infty(\mathbb{R})$. For $f \in S$, define
\[
\int_{\mathbb{R}_+ \times \mathbb{R}} f(t, x)\xi(t, x)dx dt = \sum_{i=1}^n X_i \int_{\mathbb{R}_+ \times \mathbb{R}} 1_{(a_i, b_i]}(t)\varphi_i(x)\xi(t, x)dx dt.
\]
It is easy to check that the integral is linear and satisfy Itô isometry from $L^2(\mathbb{R}_+ \times \mathbb{R} \times \Omega)$ to $L^2(\Omega)$, that is
\[
E \left[ \left( \int_{\mathbb{R}_+ \times \mathbb{R}} f(t, x)\xi(t, x)dx dt \right)^2 \right] = \int_{\mathbb{R}_+ \times \mathbb{R}} E[f^2(t, x)]dx dt.
\]
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. \( \square \)

Thus, if \( f \in L^2(\mathbb{R}_+ \times \mathbb{R} \times \Omega, \mathcal{F}, \mathbb{P}) \) there exist \( f_n \in S \) such 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 \( L^2(\Omega, \mathcal{F}, \mathbb{P}) \). Hence there is a limit point \( I \in L^2(\Omega, \mathcal{F}, \mathbb{P}) \) 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 \( L^2(\Omega, \mathcal{F}, \mathbb{P}) \) bounded initial data and solutions:

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

**Theorem 2.3** For \( L^2(\Omega, \mathcal{F}, \mathbb{P}) \) bounded initial data \( z_0 \) there exists a unique \( L^2(\Omega, \mathcal{F}, \mathbb{P}) \) 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 \( L^2(\Omega, \mathcal{F}, \mathbb{P}) \) boundedness. Thus, by Itô isometry,

\[
f(t) = \sup_{x \in \mathbb{R}} f(t, x) = \sup_{x \in \mathbb{R}} \int_{0}^{t} \int_{\mathbb{R}} \left( \frac{Z(y, s) - Z'(y, s)}{\sqrt{t-s}} \right)^2 \cdot p^2(t-s, x-y)dyds
\]

\[
\leq \sup_{x \in \mathbb{R}} \int_{0}^{t} \int_{\mathbb{R}} f(s, y)p^2(t-s, x-y)dyds
\]

\[
\leq c \int_{0}^{t} f(s) \frac{ds}{\sqrt{t-s}}
\]

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

\[
f(t) \leq c^2 \int_{0}^{t} \int_{0}^{s} \frac{f(u) du ds}{\sqrt{(s-u)(t-s)}} = c^2 \int_{0}^{t} f(u) \int_{u}^{t} \frac{ds}{\sqrt{(s-u)(t-s)}} du.
\]  \hspace{1cm} (2.3)

By Exercise \[\] we can show that \( f(t) \equiv 0 \) hence \( Z = Z' \). \( \square \)
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 \( \tilde{z}^n(t, x) := z^{n+1}(t, x) - z^n(t, x) \) satisfies

\[
\tilde{z}^{n+1}(t, x) = \int_0^t \int_\mathbb{R} p(t - s, x - y)\tilde{z}^n(s, y)\xi(s, y)dyds.
\]

Hence, as before, via the Itô isometry

\[
\mathbb{E}\left[(\tilde{z}^{n+1}(t, x))^2\right] = \int_0^t \int_\mathbb{R} p^2(t - s, x - y)\mathbb{E}\left[(\tilde{z}^n(s, y))^2\right]dyds.
\]

Define

\[
f^n(t) = \sup_{s \in [0, t]} \mathbb{E}\left[(\tilde{z}(s, x))^2\right]
\]

and note that \( f^0(t) < \infty \) by hypothesis. Then,

\[
f^{n+1}(t) = \sup_{u \in [0, t]} \mathbb{E}\left[(\tilde{z}^{n+1}(t, x))^2\right] = \sup_{u \in [0, t]} \int_\mathbb{R} p^2(u - s, x - y)\mathbb{E}\left[(\tilde{z}(s, y))^2\right]dyds
\]

\[
\leq \sup_{u \in [0, t]} \int_\mathbb{R} p^2(u - s, x - y)f^n(s)dyds
\]

\[
\leq \sup_{u \in [0, t]} c \int_0^u f^{n-1}(s)\frac{ds}{\sqrt{u - s}} \leq c \int_0^t f^{n-1}(s)\frac{ds}{\sqrt{t - s}}.
\]

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 \( \tilde{z}^n(\cdot, \cdot) \) form a Cauchy sequence in \( L^2(\Omega, \mathcal{F}, \mathbb{P}) \). 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 11 such as the mild solution given by (15), 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 \( (\mathbb{E}(M_t^2) < \infty \) for all \( t \) martingale then the quadratic variation \( \langle M \rangle_t \) is the unique non-decreasing stochastic process such that

\[13\]The notion of “weak solution” in SPDE is totally different from “weak solution” in classical PDE.
$M_t^2 - \langle M \rangle_t$ is a martingale.\footnote{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.} 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 $\varphi \in C^\infty_c(\mathbb{R})$,

$$
\int_{\mathbb{R}^+ \times \mathbb{R}} 1_{(0,t)}(s) \varphi(x) \xi(s,x) dx ds
$$

(2.4)

considered in Section 2.1 is a Brownian motion with variance $\| \varphi \|^2_{L^2}$ (its quadratic variation is $t \| \varphi \|^2_{L^2}$).

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

$$
\partial_t u = \Delta u + \xi \quad u(0) = u_0
$$

(2.5)

can be also defined in a similar way. To this end, given $\varphi \in C^\infty(\mathbb{T}^d)$ writing $(u, \varphi)$ for the evaluation of the distribution $u \in \mathcal{S}'(\mathbb{T}^2)$ 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 $t \| \varphi \|^2_{L^2}$.

This motivates the following definition.

**Definition 2.4** $v \in \mathcal{C}(\mathbb{R}^+, \mathcal{S}'(\mathbb{T}^2))$ is called a weak solution to (2.5) if

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

(2.6)

and

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

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

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

$$
\partial_t Z = \Delta Z + Z \xi \quad Z(0) = Z_0
$$

(2.7)

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

**Definition 2.5** $v \in \mathcal{C}(\mathbb{R}^+, \mathcal{S}'(\mathbb{T}^2))$ 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 $\varphi \in C^\infty(\mathbb{T}^d)$.

**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

$$f^n(t) \leq \left(\frac{ct}{n/2}\right)^n (n/2)!.$$ 

3. $\Phi^4$ equation

3.1 Local mild solution in $d = 1$

$$\partial_t u = \Delta u - u^3 + \xi \quad u(0) = u_0.$$ 

$u$ is called a mild solution if

$$u(t, x) = \int_{T^d} \int_0^t P_{t-s}(x-y)\xi(ds, dy) + P_t u_0(x) - \int_{T^d} \int_0^t P_{t-s}(x-y)u(s, y)^3 ds dy.$$ 

where $P_t u_0(x) = \int_{T^d} P_t(x-y)u_0(y)dy$ is the solution to heat equation from initial condition $u_0$. We call $u$ a local mild solution if there exists a stopping time $\tau$ s.t. (1) $\tau > 0$ a.s. and (2) \ref{3.2} holds a.s. for every $t$ such that $t < \tau$ a.s. We call $(u, \tau)$ a maximal local mild solution if for any local mild solution $(\tilde{u}, \tilde{\tau})$, $\tilde{\tau} \leq \tau$ a.s.

**Theorem 3.1** 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 $g \in C(R_+, C^\alpha)$ be given by

$$g(t, x) \triangleq \int_{T^d} \int_0^t P_{t-s}(x-y)\xi(ds, dy) + P_t u_0(x).$$

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

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

A fixed point of $M_{g,T}$ will be a mild solution. We will 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. \qed
3.2 Important heuristics: perturbation theory

3.3 $\Phi^4$ equation in $d=2$

Consider the renormalized $\Phi^4$ equation

$$\partial_t u_\varepsilon = \Delta u_\varepsilon - (u_\varepsilon^3 - 3C_\varepsilon u_\varepsilon) + \xi_\varepsilon .$$  \hspace{1cm} (3.3)

**Da Prato - Debussche argument.** Write $u_\varepsilon = \Phi_\varepsilon + v_\varepsilon$, 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

$$\partial_t v_\varepsilon = \Delta v_\varepsilon - \left(v_\varepsilon^3 + 3\Phi_\varepsilon v_\varepsilon^2 + 3(\Phi_\varepsilon^2 - C_\varepsilon)v_\varepsilon + \Phi_\varepsilon^3 - 3C_\varepsilon \Phi_\varepsilon \right) .$$  \hspace{1cm} (3.4)

In this subsection we aim to show:

1. $\Phi_\varepsilon^2 - C_\varepsilon$ and $\Phi_\varepsilon^3 - 3C_\varepsilon \Phi_\varepsilon$ converge in probability in $C^\alpha$ for any $\alpha < 0$. We will call the limits $\Phi^2$ and $\Phi^3$: respectively. (The fact that their integrations against smooth functions have bounded second moment was essentially already seen in Section 3.2. Here we identify the distribution space $C^\alpha$ in which they converge.)

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

$$\partial_t v = \Delta v - \left( v^3 + 3\Phi v^2 + 3\Phi^2 v + v + 3\Phi^3 \right)$$  \hspace{1cm} (3.5)

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 [?]

Before proving the above two results, one might ask: why has the constant $C_\varepsilon$ in (3.3) been precisely distributed to the “right places” in (3.4) to make the powers of $\Phi_\varepsilon$ converge? We first give a better interpretation to the renormalization in this context.

**Wiener chaos.** 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

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

where $\mathbb{R}[z_1, \cdots, z_k]$ 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 $H^n$, denoted by $\bar{H}^{n-1}$, is called the homogeneous chaos of order $n$. A useful fact is that if $Z \in H$, then $h_n(Z) \in H^{n}$: where $h_n$ is the $n$-th Hermite polynomial. (This is due to the fact that the Hermite polynomials form orthonormal basis of $L^2(\Omega, \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx).$)
4 (Controlled) rough paths

So far, we have reviewed the Itô / Stratonovich calculus, and discussed the definitions of stochastic integrals of predictable and $L^2$ bounded $H$ with respect to a semimartingale $X$, in particular, with respect to Brownian motion $X = B$, that is $\int H dB$. Then as a special and simple example, we calculated $\int B dB$. Namely

$$\int H dB$$

It turns out that there is a different approach, which in certain sense goes in the opposite way, namely from the simple integral $\int X dB$ one can get the general ones $\int H dB$. This is called “rough path theory”. The class of $X$ will be called rough paths (which of course include Brownian motion), and the class of $H$ will be called controlled rough paths. The true power of “rough path theory” is that it gives an almost sure pathwise definition of stochastic differential equation, and the solution to the equation, when lifted onto the level of rough paths, is continuous in the driven noise.

We denote by $C_2([0, T], \mathbb{R}^n)$ the space of continuous functions from $\mathbb{R}^2$ into $\mathbb{R}^n$ that vanish on the diagonal. Usually, we will omit the base space $[0, T]$ and the target space $\mathbb{R}^n$ in our notations for the sake of simplicity. We also define a difference operator $\delta: C \to C_2$ by

$$\delta X(x, y) = X(y) - X(x) .$$

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” $X \in C_2([0, T], \mathbb{R}^{n \times n} \times n)$ such that the algebraic relations

$$X^{ij}(r, t) - X^{ij}(r, s) - X^{ij}(s, t) = \delta X^i(r, s) \delta X^j(s, t),$$

hold for every triple of points $(r, s, t)$ and every pair of indices $(i, j)$. One should think of $X$ as postulating the value of the quantity

$$\int_s^t \delta X^i(s, r) dX^j(r) \overset{\text{def}}{=} X^{ij}(s, t),$$

where we take the right hand side as a definition for the left hand side. (And not the other way around!) The aim of imposing (4.1) is to ensure that (4.2) 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.

**Example 4.1** $n = 1$, $X(t) = B_t$, and $X(s, t) = \frac{1}{2}(B_t - B_s)^2 - c(t - s)$ where $c \in \mathbb{R}$. $(X, X)$ is a rough path. In fact, we are here just postulating that

$$\int_0^t B_s dB_s = \frac{1}{2}B_t^2 - ct$$

because (4.3) is equivalent (Exercise) with

$$\int_s^t \delta X(s, r) dX(r) = \int_s^t (B_r - B_s) dB_r = \frac{1}{2}(B_t - B_s)^2 - c(t - s)$$

**Exercise:** Prove (4.1) holds for this example.
Another important notion (proposed by Gubinelli 2004) is that of a path \( Y \) **controlled** by a rough path \( X \). Given a rough path \( X \) taking values in \( \mathbb{R}^m \), we say that a *pair* of functions \((Y, Y')\) is a rough path controlled by \( X \) if the “remainder term” \( R \) given by

\[
R(s, t) \overset{\text{def}}{=} \delta Y(s, t) - Y'(s) \delta X(s, t),
\]

has better regularity properties than \( Y \). Typically, we will assume that \( \|Y\|_\alpha < \infty \) and \( \|Y'\|_\alpha < \infty \) for some Hölder exponent \( \alpha \), but that \( \|R\|_\beta < \infty \) for some \( \beta > \alpha \). Here, \( Y(s, t), R(s, t) \in \mathbb{R}^m \) and the second term is a matrix-vector multiplication.\footnote{A priori there could be many distinct “derivative processes” \( Y' \) associated to a given path \( Y \). However, if \( X \) is a typical sample path of Brownian motion and if we impose the bound \( \|R\|_\beta < \infty \) for some \( \beta > \frac{1}{2} \), then it was shown by Hairer-Pillai that there can be at most one derivative process \( Y' \) associated to every \( Y \).}

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

\[
\delta Y(s, t) = f(B_t) - f(B_s) = f(B_s + (B_t - B_s)) - f(B_s)
\]

\[
= f'(B_s)(B_t - B_s) + f''(B_s)(B_t - B_s)^2 + \cdots
\]

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

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.

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 \( \int_a^b \langle Y_t, dX(t) \rangle \), 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”:

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

where \( P \) denotes a partition of the integration interval, and \( \|P\| \) is the length of its longest element.

With these notations at hand, we quote the following result by Gubinelli:

**Theorem 4.3** Let \((X, X)\) satisfy (4.1) and let \((Y, Y')\) be a rough path controlled by \( X \) with a remainder \( R \) given by (4.4). Assume furthermore that

\[
\|X\|_\alpha + \|X\|_\beta + \|Y'\|_\alpha + \|R\|_\beta < \infty,
\]

\( \|\| \) means that...
for some exponents $\alpha, \bar{\alpha}, \beta, \bar{\beta} > 0$. Then, provided that $\alpha + \bar{\beta} > 1$ and $\bar{\alpha} + \beta > 1$, the compensated Riemann sum in \((4.5)\) converges. Furthermore, one has the bound

$$\left| \int_s^t \delta Y(s, r), dX(r) \right| \leq |t - s|^\gamma (\|X\|_\alpha \|R\|_\beta + \|X\|_\beta \|Y'\|_\alpha) \quad (4.7)$$

with $\gamma = (\alpha + \bar{\beta}) \wedge (\bar{\alpha} + \beta)$, for some proportionality constant depending only on the dimensions of the quantities involved and the values of the exponents $\alpha, \bar{\alpha}, \beta, \bar{\beta}$.

**Example 4.4** Take the above examples of Brownian motion, we have

$$\|B\|_\alpha + \|X\|_\beta + \|Y'\|_\alpha + \|R\|_\bar{\beta} < \infty,$$

for any $\alpha < \frac{1}{2}$ (regularity of Brownian motion), $\beta < 1$ (regularity of $\frac{1}{2}(B_t - B_s)^2 - c(t - s)$), $\bar{\alpha} < \frac{1}{2}$ (regularity of $f'(B)$) and $\bar{\beta} < 1$ (regularity of the “higher order terms” $f''(B_s)(B_t - B_s)^2 + \cdots$). So the condition $\alpha + \bar{\beta} > 1$ and $\bar{\alpha} + \beta > 1$ can be satisfied. We can then define stochastic integral against $B$ as

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

When $c = \frac{1}{2}$, the second term will not contribute to the limit and this is exactly Itô integral. When $c = 0$, the above limit is the same as

$$\lim_{\|P\| \to 0} \sum_{[s,r] \in P} \frac{f(B_s) + f(B_r)}{2} (B_r - B_s)$$

which is Statonovich integral.

The basic idea of proof of Theorem 4.3 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. $X$. The integration of $X(t) - X(s)$ w.r.t. $X$ is postulated by $X$, so as long as $\bar{\alpha} + \beta > 1$ (i.e. the Young’s condition holds) one can integrate $Y'(s)(X(t) - X(s))$ w.r.t. $X$. Regarding $R$, since $\alpha + \bar{\beta} > 1$, its integration against $X$ is negligible. For detailed proofs, we refer to [?, Section 4.1-4.3].

Results such as Itô formula can be also proved, see [?, Chapter 5].

### 4.1 Fixed point argument via rough paths

Two approaches for solving SDE: Itô’s approach, and rough path approach.

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

Rough path approach: [?, Chapter 8]. Being typed...
References


