A Bridge Monograph

From Classical PDEs and Numerical Analysis to the Numerics of Stochastic PDEs

The road to solving the stochastic heat equation, from first pictures to the research frontier

Weak Solutions · Finite Elements · Operator Semigroups
Stochastic Calculus · Gaussian Measures · SPDE Discretisation

Prepared as a self-study spine
widen your base; annotate freely; the margins are yours

Palatino · house conventions · compiled July 12, 2026

Preface

This monograph exists to close a specific gap. An Indian M.Sc. in applied mathematics equips you with classical PDEs (characteristics, separation of variables, Green's functions, the maximum principle), with numerical methods as *algorithms* (finite differences, Runge–Kutta, von Neumann stability), with functional analysis through the spectral theorem for compact operators, and with Itô calculus on $\mathbb{R}^n$. The numerics of stochastic partial differential equations sits on top of all four, but it does not use any of them in the form you first learned. It uses their *functional-analytic reformulations*: PDEs as variational problems in Sobolev spaces; numerical methods as approximation theory measured in norms; the heat semigroup as an object with a smoothing estimate; and Itô calculus lifted to Hilbert-space-valued processes driven by a Wiener process that is not even trace-class.

The model problem throughout is the linear stochastic heat equation

$$\frac{\partial u}{\partial t} = \kappa \frac{\partial^2 u}{\partial x^2} + \sigma \dot{W}(x,t),$$

the same equation that adorns the poster that set this project in motion. Every chapter is chosen because it is a girder in the bridge leading to that equation and to its discretisation by classical and neural-network-based methods.

How to read this. Each major idea is presented at three altitudes. A *plain-language picture* gives the intuition with no symbols you could not explain to a bright undergraduate. The *main development* is the real mathematics: definitions, theorems, and proofs, at the level you will be examined on and expected to reproduce. A *thread to the SPDE* closes the loop, showing why the idea is load-bearing for the stochastic heat equation specifically. Boxes flag the four kinds of moment that matter: the intuition, the load-bearing idea, the place where it bites in practice, and the exercise you should attempt before moving on.

A warning about comfort. This document is a spine, not a substitute for doing the work. Read a proof, close the page, and reproduce it. The narrative here is deliberately clean; real understanding is not, and the gap between reading a clean proof and producing a messy one yourself is exactly where competence is built. Where an exercise appears, it is not decoration. Do it on paper first.

Contents

CHAPTER 1

Orientation: the shape of the whole bridge

Before any detail, a map. If you hold the map in your head, every subsequent chapter has a place to attach, and you will never be lost as to *why* a given piece of machinery is being assembled.

1.1 The five spans

The bridge from where you stand to the numerics of the stochastic heat equation has five spans, each a body of theory, each resting on the one before.

1. **Weak PDE theory.** Reformulate the deterministic heat and Poisson equations so that solutions live in Sobolev spaces and are characterised by an integral identity rather than by pointwise differentiation. This is the language in which everything downstream is spoken. Without it, "the solution has spatial regularity $C^{1/2-\varepsilon}$" and "the finite element error is $O(h)$ in the H^1 norm" are not even statements.

2. **Finite element error analysis.** Approximate the weak solution in a finite-dimensional subspace and *prove*, in a norm, how fast the approximation converges as the mesh is refined. Céa's lemma, interpolation estimates, and the Aubin–Nitsche duality argument. This is the native language of the two supervisors whose project motivates the whole exercise.

3. **Operator semigroups.** Recognise the solution operator of the heat equation as a one-parameter semigroup $S(t) = e^{t\Delta}$ generated by the Laplacian, and extract its single most important property: it smooths. The estimate $\|(-\Delta)^{\alpha} S(t)\| \le Ct^{-\alpha}$ drives essentially every error bound in the stochastic theory.

4. **Stochastic calculus and SDE numerics.** Master strong and weak convergence for numerical schemes in finite dimensions — Euler–Maruyama at strong order $1/2$, Milstein at order 1 — because these rates and their proofs lift almost verbatim to infinite dimensions.

5. **Gaussian measures and infinite-dimensional stochastic analysis.** Give rigorous meaning to space–time white noise $\dot{W}$, to the stochastic convolution $\int_0^t S(t-s)\,\mathrm{d}W(s)$, and to the mild solution of the SPDE. This is where the four preceding spans meet.

Load-bearing idea

The stochastic heat equation is not solved by any one of these theories. It is solved at the point where all five meet: the mild solution $u(t) = S(t)u_0 + \int_0^t S(t-s)\,\sigma\,\mathrm{d}W(s)$ is a semigroup (span 3) acting on a Hilbert-space stochastic integral (spans 4–5) whose spatial regularity is measured in Sobolev norms (span 1) and whose numerical approximation is analysed by finite element error theory (span 2).

1.2 The one equation, read five ways

It is worth seeing, in advance, how the same object appears in each span. Take the stochastic heat equation on the interval $\Omega = (0, 1)$ with homogeneous Dirichlet boundary conditions and additive space–time white noise.

- *Classically* (where you start): a PDE with a random forcing term so rough that the classical solution does not exist — u is nowhere differentiable in the strong sense, and the equation cannot hold pointwise.
- *Weakly* (span 1): an identity $\langle u(t), \phi \rangle = \dots$ holding for all smooth test functions ϕ, in which no derivative of u of order higher than the space permits ever appears.
- *Variationally / numerically* (span 2): a problem posed on a finite element space V_h, whose solution u_h approximates u at a rate limited by the low spatial regularity of u.
- *Semigroup / mild* (span 3): the integral equation $u(t) = S(t)u_0 + \int_0^t S(t-s)\,\sigma\,\mathrm{d}W(s)$, meaningful even though the equation has no classical solution, because $S(t)$ smooths the rough noise.
- *Probabilistically* (spans 4–5): a Hilbert-space-valued stochastic process whose increments are Gaussian, whose covariance is computable, and whose sample paths have exactly the Hölder regularity that the numerics must respect.

Thread to the SPDE

Memorise two exponents now; they will recur in every chapter. For space–time white noise in one spatial dimension, the solution of the stochastic heat equation has spatial Hölder regularity $C^{1/2-\varepsilon}$ and temporal Hölder regularity $C^{1/4-\varepsilon}$. These two numbers, $1/2$ and $1/4$, are the ceiling on every strong convergence rate you will ever prove for this equation. When a numerical scheme achieves strong order $1/2$ in space and $1/4$ in time, it is not merely good — it is optimal, because it matches the regularity of the object it approximates.

1.3 Why two spatial dimensions already break

A recurring theme, flagged here so it does not surprise you later: the linear stochastic heat equation with space–time white noise has a function-valued solution *only* in one spatial dimension. In dimension two and above, the solution is merely a distribution, its pointwise values are meaningless, and any nonlinearity requires renormalisation — the circle of ideas for which Hairer received the Fields Medal. This is not a technicality. It is a direct consequence of how the regularity of the stochastic convolution depends on dimension, and it explains why the honest model problem for a first serious study is resolutely one-dimensional. We will see exactly where the dimension enters, in Chapter 9.

1.4 What "neural-network-based methods" must respect

The project title pairs classical methods with neural-network-based ones. A theme we will return to: the solution of the SPDE is *rough*, and neural networks have a well-documented spectral bias toward smooth functions. This is not an engineering inconvenience to be tuned away; it is a structural tension between the object and the approximator. The most defensible research programme — the one this monograph is quietly building toward — is a hybrid in which the rough part of the solution is handled by classical means that respect its regularity, and

the network is asked only to represent a smooth correction, where its bias is harmless. Keep this in mind; every classical span is also preparation for understanding precisely what the neural methods can and cannot be trusted to do.

CHAPTER 2

Weak solutions and Sobolev spaces

2.1 The picture before the machinery

Plain-language picture

Imagine you want to describe the shape of a stretched drumskin pushed by some force. The classical way demands that the skin be smooth enough to have a well-defined curvature at every single point, and then insists the curvature balance the force everywhere. But real forces can be spiky — a finger pressing at one point — and then the skin has a kink, no curvature exists there, and the classical description simply fails at that point.
The weak formulation changes the question. Instead of asking "does the curvature balance the force at every point?" it asks "for every gentle test-poke I apply, does the total work done by the skin's tension match the total work done by the force?" This is a question about integrals, not about pointwise derivatives. A skin with a kink can answer it perfectly well. We have traded a demanding local statement for a forgiving global one — and, remarkably, lost nothing when the classical solution does exist, while gaining solutions in many cases where it does not.

The entire edifice of modern PDE theory — and with it, finite elements and the numerics of SPDEs — rests on this trade. The rough forcing in the stochastic heat equation is precisely the situation in which the classical description fails and the weak one survives, so we must build the weak framework carefully.

2.2 Weak derivatives

Let $\Omega \subseteq \mathbb{R}^d$ be open. Write $C_c^\infty(\Omega)$ for the smooth functions with compact support in Ω; these are the *test functions*. The starting observation is integration by parts: if $u \in C^1(\Omega)$ and $\phi \in C_c^\infty(\Omega)$, then because ϕ vanishes near $\partial\Omega$,

$$\int_\Omega \partial_i u \, \phi \, \mathrm{d}x = -\int_\Omega u \, \partial_i \phi \, \mathrm{d}x.$$

The right-hand side makes sense even when u is not differentiable — it only requires u to be integrable. This motivates promoting the identity to a definition.

Definition 2.1 (Weak derivative). Let $u \in L^1_{\mathrm{loc}}(\Omega)$. We say $g_i \in L^1_{\mathrm{loc}}(\Omega)$ is the *weak i-th partial derivative* of u, written $g_i = \partial_i u$, if

$$\int_\Omega u \, \partial_i \phi \, \mathrm{d}x = -\int_\Omega g_i \, \phi \, \mathrm{d}x \qquad \text{for all } \phi \in C_c^\infty(\Omega).$$

More generally, for a multi-index α, $g = D^\alpha u$ weakly if $\int_\Omega u \, D^\alpha \phi = (-1)^{|\alpha|} \int_\Omega g \, \phi$ for all test ϕ.

Example 2.2 (A kink has a weak derivative). Let $\Omega = (-1,1)$ and $u(x) = |x|$. Then u is not differentiable at 0, but its weak derivative is the sign function $g(x) = \operatorname{sgn}(x)$, which lies in L^1_{loc}. Indeed, for any test ϕ, splitting the integral at 0 and integrating by parts on each side,

$$\int_{-1}^{1} |x|\,\phi'\,\mathrm{d}x = -\int_{-1}^{0}(-1)\phi - \int_{0}^{1}(1)\phi = -\int_{-1}^{1}\operatorname{sgn}(x)\phi\,\mathrm{d}x,$$

the boundary terms at 0 cancelling because ϕ is continuous. So $|\cdot|$ is weakly differentiable although not classically so.

Example 2.3 (A jump has no weak derivative). On the same interval let H be the Heaviside step, $H(x) = 0$ for $x < 0$ and 1 for $x \geq 0$. There is *no* L^1_{loc} function g with $\int H\phi' = -\int g\phi$ for all ϕ: the only candidate for the derivative is the Dirac mass at 0, which is not a function. Thus weak differentiability tolerates kinks but not jumps. This dividing line — continuous-but-cornered is fine, discontinuous is not — is exactly the regularity of the piecewise-linear "hat" functions we will use as finite elements.

Thread to the SPDE

The distinction in these two examples is the seed of the dimensional threshold for SPDEs. Space–time white noise, integrated once by the heat semigroup, produces in $d = 1$ a solution that is like $|x|$ — continuous, cornered, weakly differentiable, function-valued. In $d \geq 2$ the same construction produces something like H or worse — a genuine distribution with no function representative. The weak-derivative dividing line *is* the SPDE dimensional threshold, viewed locally.

2.3 Sobolev spaces

Definition 2.4 (Sobolev space H^k). For $k \in \mathbb{N}$, the Sobolev space $H^k(\Omega)$ is the set of $u \in L^2(\Omega)$ all of whose weak derivatives up to order k exist and lie in $L^2(\Omega)$, equipped with the norm

$$\|u\|^2_{H^k(\Omega)} = \sum_{|\alpha|\leq k} \int_\Omega |D^\alpha u|^2\,\mathrm{d}x.$$

The case $k = 1$ is the workhorse: $H^1(\Omega) = \{u \in L^2 : \nabla u \in L^2\}$, with $\|u\|^2_{H^1} = \|u\|^2_{L^2} + \|\nabla u\|^2_{L^2}$.

Theorem 2.5 (Completeness). *$H^k(\Omega)$ is a Hilbert space under the inner product $\langle u, v\rangle_{H^k} = \sum_{|\alpha|\leq k}\langle D^\alpha u, D^\alpha v\rangle_{L^2}$.*

Proof sketch. Bilinearity, symmetry and positivity are immediate. For completeness, let (u_n) be Cauchy in H^k. Then $(D^\alpha u_n)$ is Cauchy in L^2 for each $|\alpha| \leq k$, hence converges to some $g_\alpha \in L^2$ by completeness of L^2. Writing the defining identity $\int u_n D^\alpha\phi = (-1)^{|\alpha|}\int D^\alpha u_n\,\phi$ and passing to the limit (both sides converge, since L^2 convergence implies convergence of the L^2 pairing with the fixed $\phi \in L^2$), one gets $\int u\,D^\alpha\phi = (-1)^{|\alpha|}\int g_\alpha\phi$, i.e. $g_\alpha = D^\alpha u$ weakly. Thus $u \in H^k$ and $u_n \to u$ in H^k. □

To impose boundary conditions we need a subspace of functions that "vanish on $\partial\Omega$" — but functions in H^1 are only defined up to null sets, and $\partial\Omega$ is null, so "vanish on the boundary" needs care. Two equivalent routes exist; we take the cleaner one.

Definition 2.6 (H^1_0). $H^1_0(\Omega)$ is the closure of $C^\infty_c(\Omega)$ in the $H^1(\Omega)$ norm.

Intuitively H^1_0 consists of H^1 functions with zero boundary trace. The trace theorem makes the phrase "boundary values of an H^1 function" rigorous; we state it without proof, as we will need only its consequence that the trace is a bounded operation.

Theorem 2.7 (Trace theorem). *Let Ω be a bounded Lipschitz domain. There is a bounded linear* trace *operator $\gamma : H^1(\Omega) \to L^2(\partial\Omega)$ that agrees with restriction $u \mapsto u|_{\partial\Omega}$ on $C^1(\bar{\Omega})$, and $H^1_0(\Omega) = \ker \gamma$.*

2.3.1 The Poincaré inequality: the coercivity engine

The single inequality that makes the Dirichlet problem well-posed is Poincaré's. We prove the one-dimensional case in full because you will need to reproduce it, and because its proof is a model of how boundary vanishing is exploited.

Theorem 2.8 (Poincaré inequality, 1D). *There is a constant $C_P > 0$ such that for all $u \in H^1_0(0,1)$,*

$$\|u\|_{L^2(0,1)} \le C_P \, \|u'\|_{L^2(0,1)}.$$

One may take $C_P = 1/\pi$, but any finite constant suffices for our purposes; the crude $C_P = 1$ is proved below.

Proof. It suffices to prove the bound for $u \in C^\infty_c(0,1)$ and pass to the limit, since such u are dense in H^1_0 by definition and both sides are continuous in the H^1 norm. For such u we have $u(0) = 0$, so by the fundamental theorem of calculus and then Cauchy–Schwarz,

$$u(x) = \int_0^x u'(t)\,\mathrm{d}t, \qquad |u(x)|^2 = \left(\int_0^x u'(t)\,\mathrm{d}t\right)^2 \le x \int_0^x |u'(t)|^2\,\mathrm{d}t \le \int_0^1 |u'|^2\,\mathrm{d}t.$$

Integrating over $x \in (0,1)$ gives $\int_0^1 |u(x)|^2\,\mathrm{d}x \le \int_0^1 |u'|^2\,\mathrm{d}t$, i.e. $\|u\|_{L^2} \le \|u'\|_{L^2}$. □

Load-bearing idea

Poincaré is what converts the "incomplete" seminorm $\|u'\|_{L^2}$ into a genuine norm on H^1_0: it says that if the gradient is small and the function vanishes on the boundary, the function itself is small. Every coercivity estimate in this monograph — the well-posedness of the weak Poisson problem, the stability of the finite element method, the energy estimate for the parabolic equation — runs through Poincaré. When you see coercivity, look for the Poincaré inequality underneath it.

2.4 The weak formulation of the Dirichlet problem

Consider the model boundary-value problem

$$-u'' = f \text{ on } (0,1), \qquad u(0) = u(1) = 0. \tag{2.1}$$

Multiply by a test function $v \in C^\infty_c(0,1)$ and integrate by parts:

$$\int_0^1 (-u'')v\,\mathrm{d}x = \int_0^1 u'v'\,\mathrm{d}x - \underbrace{[u'v]_0^1}_{=0} = \int_0^1 fv\,\mathrm{d}x.$$

The boundary term vanishes because v does. This computation asks only for one derivative of u, in L^2; so it makes sense for $u \in H^1_0$. Extending by density from test functions to all of H^1_0, we arrive at the weak problem.

Definition 2.9 (Weak problem). Given $f \in L^2(0,1)$, find $u \in V := H^1_0(0,1)$ such that

$$a(u,v) := \int_0^1 u'v'\,\mathrm{d}x = \int_0^1 fv\,\mathrm{d}x =: F(v) \qquad \text{for all } v \in V. \tag{2.2}$$

The bilinear form a and the linear functional F are the objects of interest. Existence and uniqueness of u follow from a single abstract theorem, whose two hypotheses — boundedness and coercivity — we now verify for this a, and then state in general.

Proposition 2.10 (Boundedness and coercivity of a). *On $V = H_0^1(0,1)$ the form $a(u,v) = \int_0^1 u'v'$ satisfies:*

(i) ***Boundedness:*** *$|a(u,v)| \le M\|u\|_V\|v\|_V$ with $M = 1$;*

(ii) ***Coercivity:*** *$a(u,u) \ge \alpha\|u\|_V^2$ with $\alpha = (1+C_P^2)^{-1} > 0$.*

Here $\|\cdot\|_V = \|\cdot\|_{H^1}$.

Proof. *(i)* By Cauchy–Schwarz in L^2, $|a(u,v)| = |\langle u', v'\rangle_{L^2}| \le \|u'\|_{L^2}\|v'\|_{L^2} \le \|u\|_{H^1}\|v\|_{H^1}$, the last step because $\|u'\|_{L^2} \le \|u\|_{H^1}$ by definition of the H^1 norm. Thus $M = 1$.

(ii) We must bound $a(u,u) = \|u'\|_{L^2}^2$ below by a multiple of $\|u\|_{H^1}^2 = \|u\|_{L^2}^2 + \|u'\|_{L^2}^2$. By Poincaré (Theorem 2.8), $\|u\|_{L^2}^2 \le C_P^2\|u'\|_{L^2}^2$, so

$$\|u\|_{H^1}^2 = \|u\|_{L^2}^2 + \|u'\|_{L^2}^2 \le (C_P^2+1)\|u'\|_{L^2}^2 = (C_P^2+1)\,a(u,u),$$

whence $a(u,u) \ge (1+C_P^2)^{-1}\|u\|_{H^1}^2$. Set $\alpha = (1+C_P^2)^{-1}$. □

Theorem 2.11 (Lax–Milgram). *Let V be a real Hilbert space, $a : V \times V \to \mathbb{R}$ a bilinear form that is bounded ($|a(u,v)| \le M\|u\|\|v\|$) and coercive ($a(u,u) \ge \alpha\|u\|^2$, $\alpha > 0$), and $F \in V^*$ a bounded linear functional. Then there exists a unique $u \in V$ with $a(u,v) = F(v)$ for all $v \in V$, and moreover $\|u\| \le \alpha^{-1}\|F\|_{V^*}$.*

Proof. For each fixed u, the map $v \mapsto a(u,v)$ is a bounded linear functional, so by the Riesz representation theorem there is a unique $Au \in V$ with $a(u,v) = \langle Au, v\rangle$ for all v; linearity and boundedness of $u \mapsto Au$ (with $\|Au\| \le M\|u\|$) are routine. Likewise $F(v) = \langle f, v\rangle$ for a unique $f \in V$ with $\|f\| = \|F\|_{V^*}$. The problem becomes: find u with $Au = f$.

Coercivity gives $\langle Au, u\rangle = a(u,u) \ge \alpha\|u\|^2$, so $\|Au\| \ge \alpha\|u\|$; hence A is injective with closed range. If the range were not all of V, some $0 \ne w \perp \operatorname{ran} A$; but then $\alpha\|w\|^2 \le a(w,w) = \langle Aw, w\rangle = 0$, forcing $w = 0$, a contradiction. So A is a bijection, $u = A^{-1}f$ exists uniquely, and $\alpha\|u\|^2 \le a(u,u) = F(u) \le \|F\|_{V^*}\|u\|$ gives the bound. □

Corollary 2.12. *The weak Poisson problem (2.2) has a unique solution $u \in H_0^1(0,1)$ for every $f \in L^2(0,1)$, with $\|u\|_{H^1} \le \alpha^{-1}\|f\|_{L^2}$.*

Proof. Combine Proposition 2.10 with Theorem 2.11; $F(v) = \langle f, v\rangle_{L^2}$ is bounded on H_0^1 since $|F(v)| \le \|f\|_{L^2}\|v\|_{L^2} \le \|f\|_{L^2}\|v\|_{H^1}$. □

Thread to the SPDE

This is the template for the elliptic problems that appear at each time step when the stochastic heat equation is discretised implicitly in time. Each such step solves $(I - \Delta t\,\Delta)u^{n+1} = (\text{data})$, whose weak form is coercive by exactly the Poincaré-driven argument above, now with the Δt improving the coercivity constant. The stability of the whole time-stepping scheme is inherited, step by step, from Lax–Milgram.

2.5 Sobolev embedding and the regularity of rough objects

One more circle of results, because it quantifies exactly *how rough* the SPDE solution is. Sobolev embedding theorems say that enough weak differentiability buys genuine (classical, Hölder) continuity.

Theorem 2.13 (Sobolev embedding, representative cases). *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain.*

(i) If $s > d/2$ then $H^s(\Omega) \hookrightarrow C(\bar{\Omega})$ continuously; more precisely $H^s \hookrightarrow C^{0,\gamma}$ for $\gamma = s - d/2$ when $0 < s - d/2 < 1$.

(ii) In $d = 1$: $H^1(0,1) \hookrightarrow C^{0,1/2}(0,1)$, i.e. every H^1 function in one dimension is Hölder continuous with exponent $1/2$.

The fractional Sobolev spaces H^s for non-integer s, defined by interpolation or via the Fourier transform $\|u\|_{H^s}^2 = \int (1 + |\xi|^2)^s |\hat{u}(\xi)|^2 \, d\xi$, are exactly the scale on which the SPDE solution's regularity is measured.

Where this bites

Here is the dimensional threshold, made quantitative. The stochastic convolution — the solution of the linear SPDE with zero initial data — turns out to live in H^s for every $s < 1 - d/2$. In $d = 1$ that means $s < 1/2$, which by Theorem 2.13(i) with $s - d/2 = s - 1/2 < 0$ is *just* below the threshold for continuity: the solution is a function, barely, with Hölder regularity approaching $1/2$ but never reaching it — exactly the $C^{1/2-\varepsilon}$ from the orientation. In $d = 2$, $s < 0$: the solution has *negative* Sobolev regularity, is not a function at all, and only exists as a distribution. The number $d/2$ crossing the value 1 is the whole story of why $d = 1$ is special.

2.6 Negative-order spaces: where the noise lives

One more space is indispensable, because it is the home of the forcing term in the SPDE. If H_0^1 consists of functions with one weak derivative, its *dual* $H^{-1} := (H_0^1)^*$ consists of objects with one *fewer* derivative than L^2 — distributions rough enough that they are not functions, yet tame enough to be tested against H_0^1.

Definition 2.14 (The space H^{-1}). $H^{-1}(\Omega)$ is the dual of $H_0^1(\Omega)$: the bounded linear functionals on H_0^1, normed by $\|F\|_{H^{-1}} = \sup_{\|v\|_{H_0^1}=1} |F(v)|$. Via the Riesz map it is characterised, in the eigenbasis of $A = -\Delta$, by

$$H^{-1} = \Big\{ F = \textstyle\sum_k c_k e_k : \ \|F\|_{H^{-1}}^2 = \sum_k \lambda_k^{-1} c_k^2 < \infty \Big\}.$$

The chain $H_0^1 \hookrightarrow L^2 \hookrightarrow H^{-1}$ is a *Gelfand triple*: each embedding is continuous and dense, L^2 is identified with its own dual in the middle, and the duality pairing $\langle \cdot, \cdot \rangle_{H^{-1} \times H_0^1}$ extends the L^2 inner product. This triple is the abstract stage on which the entire variational theory of parabolic and stochastic equations is set.

Thread to the SPDE

Space–time white noise, at a fixed time, is *not* an element of L^2 — we saw in the orientation that its energy is infinite. But it *is* an element of H^{-1} in one dimension, and more generally of $H^{-d/2-\varepsilon}$. The Gelfand triple is precisely the ladder that lets us make sense of an equation whose forcing lives below L^2: the solution sits in H_0^1 or L^2, the equation is tested against H_0^1, and the rough noise is absorbed in the H^{-1} slot. When you read "the solution is H_0^1-valued and the noise is H^{-1}-valued," you are reading the Gelfand triple at work.

2.7 A fully worked eigenvalue computation

Because the eigenbasis of the Dirichlet Laplacian is the single most-used object in the whole monograph, we compute it from scratch once.

Example 2.15 (Eigenpairs of $-\partial_x^2$ on $(0,1)$)**.** Seek $e \neq 0$ and λ with $-e'' = \lambda e$ on $(0,1)$ and $e(0) = e(1) = 0$. The ODE $e'' + \lambda e = 0$ has, for $\lambda > 0$, general solution $e(x) = B\sin(\sqrt{\lambda}\,x) + C\cos(\sqrt{\lambda}\,x)$. The condition $e(0) = 0$ forces $C = 0$. The condition $e(1) = 0$ then requires $B\sin\sqrt{\lambda} = 0$; for a nontrivial solution ($B \neq 0$) we need $\sin\sqrt{\lambda} = 0$, i.e. $\sqrt{\lambda} = k\pi$ for $k \in \mathbb{N}$. Hence

$$\lambda_k = k^2\pi^2, \qquad e_k(x) = \sqrt{2}\,\sin(k\pi x),$$

the normalisation $\sqrt{2}$ chosen so that $\int_0^1 e_k^2 = 1$ (since $\int_0^1 \sin^2(k\pi x)\,\mathrm{d}x = \tfrac{1}{2}$). Orthonormality $\langle e_j, e_k\rangle = \delta_{jk}$ is the standard orthogonality of sines. One checks there are no eigenvalues for $\lambda \leq 0$: for $\lambda < 0$ the solutions are real exponentials, which cannot satisfy both boundary conditions nontrivially, and $\lambda = 0$ forces the affine $e(x) = a + bx$, killed to zero by the boundary conditions. Thus $\{(\lambda_k, e_k)\}_{k\geq 1}$ is the complete spectral resolution, and $\lambda_k = k^2\pi^2$ is the eigenvalue growth that decides every convergence question downstream.

Intuition

Everything specific to the one-dimensional stochastic heat equation — the $1/2$ and $1/4$ exponents, the existence in $d = 1$ and failure in $d = 2$, the exact per-mode Gaussian increments of the exponential Euler scheme — is downstream of the single fact $\lambda_k = k^2\pi^2$, i.e. $\lambda_k \sim k^2$. In d dimensions Weyl's law gives $\lambda_k \sim k^{2/d}$, and the exponent $2/d$ is the one dial that turns every result. Keep Example 2.15 in mind as the concrete anchor beneath all the abstraction: whenever a proof invokes "the eigenvalue growth," it means this.

Work it

(1) Prove the 1D Sobolev embedding $H^1(0,1) \hookrightarrow C^{0,1/2}$ directly: for $u \in C^\infty$, write $u(x) - u(y) = \int_y^x u'$, apply Cauchy–Schwarz, and read off the Hölder-$1/2$ bound $|u(x) - u(y)| \leq \|u'\|_{L^2}|x-y|^{1/2}$; then extend by density. (2) Explain, in one paragraph and with no formulas, why the same argument fails in two dimensions. (3) Verify that $u(x) = x(1-x)$ solves the weak Poisson problem (2.2) with $f \equiv 2$, and check the a-priori bound $\|u\|_{H^1} \leq \alpha^{-1}\|f\|_{L^2}$ numerically for your C_P. (4) Using Example 2.15, expand $f \equiv 2$ in the sine basis and solve the weak problem mode by mode; confirm the series solution agrees with $x(1-x)$.

CHAPTER 3

The finite element method and its error analysis

3.1 The picture before the machinery

Plain-language picture

You want to trace a curved shape but you are only allowed to use straight line segments. The finite element idea is: chop the interval into small pieces, and on each piece draw the best straight line. Join them up and you get a jagged approximation to the smooth curve. Make the pieces smaller and the jagged path hugs the true curve ever more closely.
The deep question is not *that* it gets closer — obviously it does — but *how fast*. If you halve the piece size, does the error halve? Quarter? The answer decides whether the method is usable: a method whose error halves when you double the work is worth far less than one whose error quarters. The finite element error analysis is the machinery that turns "obviously it converges" into "it converges at rate h^2 in this norm", and that number is the difference between a method you can trust and one you cannot.

This chapter is the one with no substitute elsewhere in your training. The finite element method as taught in engineering departments is about assembling stiffness matrices; the finite element method as your prospective supervisors practise it is about *proving convergence rates in norms*. We build the second.

3.2 The Galerkin method

Return to the abstract weak problem: find $u \in V$ with $a(u, v) = F(v)$ for all $v \in V$, where V is a Hilbert space, a is bounded (constant M) and coercive (constant α). The Galerkin method replaces the infinite-dimensional V by a finite-dimensional subspace $V_h \subset V$.

Definition 3.1 (Galerkin approximation)**.** The *Galerkin approximation* $u_h \in V_h$ is the solution of

$$a(u_h, v_h) = F(v_h) \qquad \text{for all } v_h \in V_h. \tag{3.1}$$

Since $V_h \subset V$, the same Lax–Milgram theorem (Theorem 2.11) applied on the Hilbert space V_h guarantees u_h exists and is unique. The method is well-posed for free. The question is how close u_h is to u.

3.2.1 Galerkin orthogonality: the whole subject in one line

Theorem 3.2 (Galerkin orthogonality). *Let u solve the continuous problem and u_h the Galerkin problem* (3.1)*. Then*

$$a(u - u_h,\, v_h) = 0 \qquad \text{for all } v_h \in V_h. \tag{3.2}$$

Proof. The continuous problem holds for *all* $v \in V$, in particular for every $v_h \in V_h \subset V$: thus $a(u, v_h) = F(v_h)$. The discrete problem states $a(u_h, v_h) = F(v_h)$ for all $v_h \in V_h$. Subtract: $a(u, v_h) - a(u_h, v_h) = 0$, and by bilinearity $a(u - u_h, v_h) = 0$. □

Load-bearing idea

Galerkin orthogonality says the error $u - u_h$ is "a-orthogonal" to the entire approximation space V_h. Geometrically, when a is symmetric it is an inner product, and u_h is the *a-orthogonal projection* of u onto V_h — the closest point in V_h to u measured in the energy norm $\|v\|_a := a(v,v)^{1/2}$. The finite element solution is not merely *some* approximation; it is the *best possible* one in the energy norm. Céa's lemma below is the quantitative form of this statement, and every finer error estimate — L^2 rates, a-posteriori bounds — is obtained by squeezing (3.2) further.

3.2.2 Céa's lemma

Theorem 3.3 (Céa). *Under boundedness and coercivity of a, the Galerkin solution satisfies*

$$\|u - u_h\|_V \le \frac{M}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V.$$

Proof. Let $v_h \in V_h$ be arbitrary. Using coercivity, then bilinearity to insert $\pm v_h$, then Galerkin orthogonality (3.2) to kill the term $a(u - u_h, u_h - v_h)$ (legitimate since $u_h - v_h \in V_h$), then boundedness:

$$\begin{aligned}
\alpha \|u - u_h\|_V^2 &\le a(u - u_h,\, u - u_h) \\
&= a(u - u_h,\, u - v_h) + a(u - u_h,\, v_h - u_h) \\
&= a(u - u_h,\, u - v_h) + 0 \\
&\le M \, \|u - u_h\|_V \, \|u - v_h\|_V .
\end{aligned}$$

Divide by $\|u - u_h\|_V$ (if it is zero the bound is trivial): $\|u - u_h\|_V \le (M/\alpha)\|u - v_h\|_V$. Since $v_h \in V_h$ was arbitrary, take the infimum. □

Intuition

Céa's lemma performs a crucial divorce. It separates the *approximation-theoretic* question — how well can functions in V_h approximate the true solution u, i.e. how small is $\inf_{v_h} \|u - v_h\|_V$? — from the *method*, which contributes only the constant M/α. The finite element method is, up to the fixed factor M/α, as good as the best approximation the space allows. So to get a convergence *rate*, we no longer need to think about the PDE at all: we need only estimate how well piecewise polynomials approximate u. That is a question in interpolation theory, and it is where the mesh size h and the smoothness of u finally enter.

3.3 Piecewise-linear finite elements and interpolation

Partition $[0,1]$ into N subintervals by nodes $0 = x_0 < x_1 < \cdots < x_N = 1$, with $h_i = x_i - x_{i-1}$ and $h = \max_i h_i$. Let $V_h \subset H_0^1(0,1)$ be the space of continuous functions that are linear on each subinterval and vanish at the endpoints. A basis is the set of *hat functions* $\{\phi_i\}_{i=1}^{N-1}$, where ϕ_i is the piecewise-linear function equal to 1 at x_i and 0 at every other node.

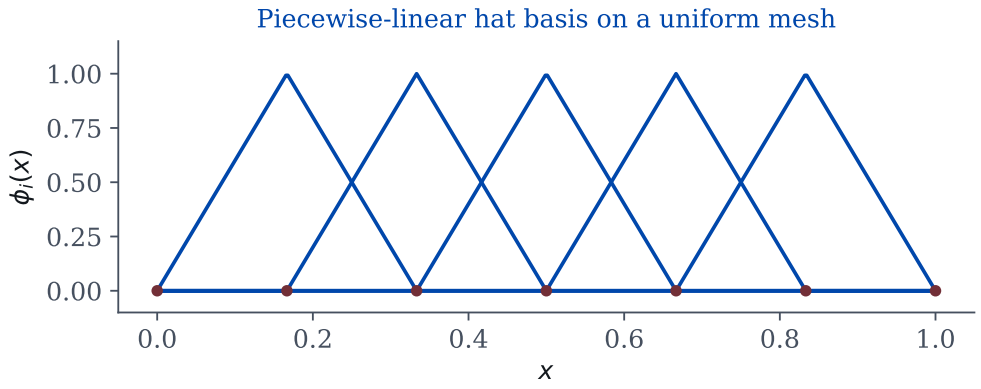

Figure 3.1: The piecewise-linear hat basis $\{\phi_i\}$ on a uniform mesh. Each ϕ_i peaks at node x_i and vanishes at every other node; a general element of V_h is a linear combination whose coefficients are its nodal values. The overlap of adjacent hats — each meets only its two neighbours — is what makes the stiffness matrix tridiagonal.

Definition 3.4 (Nodal interpolant)**.** For continuous u, its nodal interpolant $I_h u \in V_h$ is the unique element of V_h agreeing with u at every node: $I_h u = \sum_i u(x_i)\phi_i$.

Example 3.5 (The stiffness and mass matrices, computed)**.** Write $u_h = \sum_j U_j \phi_j$ and test against each ϕ_i in the discrete problem (3.1). This turns the weak equation into the linear system $AU = b$, where $A_{ij} = a(\phi_j, \phi_i) = \int_0^1 \phi_j' \phi_i'$ is the *stiffness matrix* and $b_i = \int_0^1 f\phi_i$. On a uniform mesh of width h, each hat has slope $\pm 1/h$ on its two supporting elements, so a direct computation gives

$$A_{ii} = \int (\phi_i')^2 = \frac{2}{h}, \qquad A_{i,i\pm1} = \int \phi_i' \phi_{i\pm1}' = -\frac{1}{h}, \qquad A = \frac{1}{h}\,\mathrm{tridiag}(-1,\, 2,\, -1).$$

This is, up to the factor $1/h$, exactly the second-difference matrix of the finite difference method — a first hint that finite elements and finite differences coincide for this simplest problem, though they diverge sharply on non-uniform meshes and in higher dimensions. The *mass matrix* $M_{ij} = \int \phi_i \phi_j = \frac{h}{6}\,\mathrm{tridiag}(1, 4, 1)$ will appear in the parabolic problem below, multiplying the time derivative.

The whole rate now hinges on one interpolation estimate.

Theorem 3.6 (Interpolation error)**.** *There is a constant C independent of h such that for all $u \in H^2(0,1) \cap H_0^1(0,1)$,*

$$\|u - I_h u\|_{L^2} \le C h^2 \, \|u''\|_{L^2}, \qquad \|(u - I_h u)'\|_{L^2} \le C h \, \|u''\|_{L^2}.$$

Proof of the H^1-seminorm bound on one element. Fix a subinterval $K = [x_{i-1}, x_i]$ of length h_K. Let $e = u - I_h u$. Since $I_h u$ interpolates u at both endpoints of K, $e(x_{i-1}) = e(x_i) = 0$; by Rolle's theorem e' vanishes at some $\xi \in K$. Then for $x \in K$, $e'(x) = \int_\xi^x e'' = \int_\xi^x u''$ (as $I_h u$ is linear, its second derivative is zero), so by Cauchy–Schwarz $|e'(x)| \le \int_K |u''| \le h_K^{1/2} \|u''\|_{L^2(K)}$. Squaring and integrating over K gives $\|e'\|_{L^2(K)}^2 \le h_K^2 \|u''\|_{L^2(K)}^2$. Summing over all elements yields

$\|e'\|_{L^2} \le h\|u''\|_{L^2}$. The L^2 bound follows by a Poincaré-type argument on each element (an extra factor h_K). □

Corollary 3.7 (Energy-norm convergence). *If the weak solution u lies in H^2, the piecewise-linear finite element solution satisfies*

$$\|u - u_h\|_{H^1} \le \frac{M}{\alpha}\,\|u - I_h u\|_{H^1} \le Ch\,\|u''\|_{L^2} = O(h).$$

Proof. Céa (Theorem 3.3) bounds $\|u-u_h\|_{H^1}$ by $(M/\alpha)\inf_{v_h}\|u-v_h\|_{H^1}$; choose the particular competitor $v_h = I_h u$ and apply Theorem 3.6. □

3.4 The Aubin–Nitsche duality trick

Corollary 3.7 gives $O(h)$ in the H^1 norm. In the weaker L^2 norm one expects — and gets — one order better, $O(h^2)$. The mechanism is a duality argument of surpassing elegance, due to Aubin and Nitsche.

Theorem 3.8 (Aubin–Nitsche). *Assume the domain and coefficients are such that the dual problem is H^2-regular: for every $g \in L^2$, the solution w of $a(v,w) = \langle g, v\rangle$ for all v satisfies $w \in H^2$ with $\|w\|_{H^2} \le C_{\mathrm{reg}}\|g\|_{L^2}$. Then*

$$\|u - u_h\|_{L^2} \le C\,h\,\|u - u_h\|_{H^1} = O(h^2).$$

Proof. Represent the L^2 norm of the error by duality. Let $e = u - u_h$ and take $g = e$ in the dual problem, giving $w \in H^2$ with $a(v,w) = \langle e, v\rangle$ for all $v \in V$ and $\|w\|_{H^2} \le C_{\mathrm{reg}}\|e\|_{L^2}$. Put $v = e$:

$$\|e\|_{L^2}^2 = \langle e, e\rangle = a(e, w).$$

Now subtract $a(e, I_h w)$, which is zero by Galerkin orthogonality (Theorem 3.2, since $I_h w \in V_h$):

$$\|e\|_{L^2}^2 = a(e,\, w - I_h w) \le M\,\|e\|_{H^1}\,\|w - I_h w\|_{H^1} \le M\,\|e\|_{H^1}\cdot Ch\|w''\|_{L^2},$$

using boundedness and the interpolation estimate. Finally $\|w''\|_{L^2} \le \|w\|_{H^2} \le C_{\mathrm{reg}}\|e\|_{L^2}$. Assembling, $\|e\|_{L^2}^2 \le Ch\|e\|_{H^1}\|e\|_{L^2}$; cancel one factor of $\|e\|_{L^2}$ to get $\|e\|_{L^2} \le Ch\|e\|_{H^1} = O(h^2)$ by Corollary 3.7. □

Intuition

The trick is to measure the error against itself *through* an auxiliary smooth problem. The dual solution w is smoother than the error, so subtracting its interpolant — free of charge, by Galerkin orthogonality — gains one power of h. The hypothesis that made it work is *elliptic regularity*: the dual solution must be genuinely H^2. On a non-convex domain or with rough coefficients this can fail, and then the L^2 rate degrades below h^2. Remember this: the extra order is borrowed from the regularity of the dual problem, and when that regularity is unavailable, the order is not there to borrow.

Thread to the SPDE

This is the exact mechanism that reappears, perturbed by noise, in the stochastic setting. In Chapter 10 the deterministic Ritz projection and its error splitting $u - u_h = (u - R_h u) + (R_h u - u_h)$ carry over to the semidiscrete finite element approximation of the stochastic heat equation. The weak (as opposed to strong) convergence rate for the SPDE is proved by a stochastic analogue of Aubin–Nitsche duality — which is why weak rates are roughly double strong rates, just as the L^2 rate here is double the H^1 rate. The doubling you see in this deterministic chapter is the same doubling you will meet in the stochastic one.

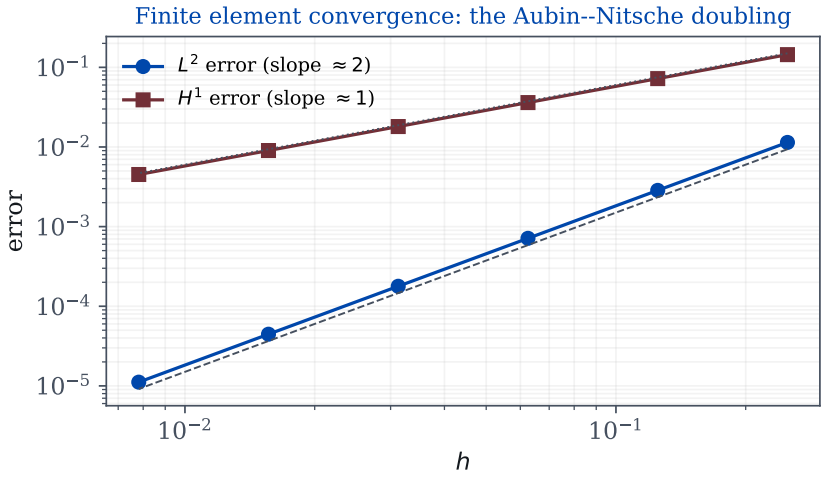

Figure 3.2: Finite element convergence for $-u'' = 2$ on $(0,1)$, exact solution $u = x(1-x)$, computed in the author's house style. The H^1 error falls at rate h (slope 1 on the log–log axes) and the L^2 error one order faster at rate h^2 (slope 2): the Aubin–Nitsche doubling, made visible. Halving h halves the H^1 error but quarters the L^2 error. The measured slopes are 2.00 and 1.00 to two decimals.

3.5 Parabolic problems: the semidiscrete method

The stochastic heat equation is parabolic, so we need the finite element treatment of the deterministic heat equation as a rehearsal. Consider $\partial_t u - \Delta u = f$ with $u(0) = u_0$ and Dirichlet conditions. The semidiscrete (discrete in space, continuous in time) finite element method seeks $u_h(t) \in V_h$ with

$$\langle \partial_t u_h, v_h \rangle + a(u_h, v_h) = \langle f, v_h \rangle \quad \forall v_h \in V_h, \qquad u_h(0) = P_h u_0,$$

where P_h is the L^2 projection onto V_h. The analysis introduces the *Ritz projection* $R_h : V \to V_h$ defined by $a(R_h u, v_h) = a(u, v_h)$ for all v_h — i.e. the finite element solution of the *elliptic* problem with the same data — and splits the error as

$$u - u_h = \underbrace{(u - R_h u)}_{\rho,\ \text{elliptic},\ O(h^2)} + \underbrace{(R_h u - u_h)}_{\theta,\ \text{controlled by an energy estimate}}.$$

The first piece ρ is bounded by the elliptic theory just developed. The second piece $\theta \in V_h$ satisfies a discrete evolution equation whose right-hand side is $-\partial_t \rho$, and a Grönwall energy estimate closes the bound. The upshot is $O(h^2)$ convergence in L^2 for smooth data.

Load-bearing idea

The Ritz-projection error splitting is *the* technique of parabolic finite element analysis, and it is exactly the skeleton that Kruse's stochastic monograph fleshes out with a noise term. If you understand this splitting deterministically — the smooth part handled by elliptic theory, the discrete part by an energy estimate — then the stochastic version in Chapter 10 will read as "the same, plus a stochastic convolution one must estimate." Master it here, where there is no randomness to distract you.

Work it

(1) Reproduce the proof of Céa's lemma with the page closed. (2) Assemble, in your NumPy house style, the stiffness matrix $A_{ij} = \int_0^1 \phi_i' \phi_j'$ and mass matrix $M_{ij} = \int_0^1 \phi_i \phi_j$ for uniform hat functions; verify A is tridiagonal $\frac{1}{h}\,\mathrm{tridiag}(-1, 2, -1)$. (3) Solve the discrete Poisson problem for $f \equiv 2$, compare to the exact $u = x(1-x)$, and confirm numerically that halving h quarters the L^2 error (rate h^2) but only halves the H^1 error (rate h). This single experiment makes the Aubin–Nitsche doubling tangible.

CHAPTER 4

First problems: the onboarding sequence

Intuition

This short chapter is different from the others. It contains no exposition — only four problems, in a deliberate order, that you should attempt on paper *before* reading their solutions in Appendix B. They are the entry ritual to the whole subject: coercivity, boundedness, Galerkin orthogonality, and Céa's lemma. The proofs also appear woven into Chapters 2–3 as part of the running development, but here they are isolated as *your* first task, because the gap between reading a clean proof and producing a messy one yourself is exactly where competence is built. Attempt each one cold. Only when you have a real attempt — including the crossings-out — turn to the solution.

Throughout, $V = H_0^1(0,1)$ with norm $\|\cdot\|_V = \|\cdot\|_{H^1}$, and $a(u,v) = \int_0^1 u'v'\,\mathrm{d}x$ is the bilinear form of the weak Dirichlet problem (2.2). Recall the Poincaré inequality (Theorem 2.8): there is $C_P > 0$ with $\|u\|_{L^2} \le C_P \|u'\|_{L^2}$ for all $u \in V$.

Problem 1 — Coercivity

Work it

Prove that a is *coercive* on V: there exists $\alpha > 0$ such that

$$a(u,u) \ge \alpha \, \|u\|_V^2 \qquad \text{for all } u \in V.$$

Identify the constant α in terms of the Poincaré constant C_P.

Hint (uncover only if stuck). $a(u,u) = \|u'\|_{L^2}^2$, but $\|u\|_V^2 = \|u\|_{L^2}^2 + \|u'\|_{L^2}^2$ has the extra L^2 piece. You must absorb $\|u\|_{L^2}^2$ into a multiple of $\|u'\|_{L^2}^2$ — which is precisely what Poincaré buys you. This is the single place the boundary condition (hidden inside H_0^1, hence inside Poincaré) does its work.

Problem 2 — Boundedness

Work it

Prove that a is *bounded* on V: there exists $M < \infty$ such that

$$|a(u,v)| \le M \|u\|_V \|v\|_V \qquad \text{for all } u, v \in V,$$

and state the value of M. (This one is easy; do it anyway, and note exactly which inequality you use and where the definition of the H^1 norm enters.)

Hint. One application of Cauchy–Schwarz in L^2, then the observation that $\|u'\|_{L^2} \le \|u\|_{H^1}$ by definition of the H^1 norm.

Problem 3 — Galerkin orthogonality

Work it

Let $u \in V$ solve the continuous problem $a(u,v) = F(v)$ for all $v \in V$, and let $V_h \subset V$ be any finite-dimensional subspace with $u_h \in V_h$ solving the discrete problem $a(u_h, v_h) = F(v_h)$ for all $v_h \in V_h$. Prove *Galerkin orthogonality*:

$$a(u - u_h,\, v_h) = 0 \qquad \text{for all } v_h \in V_h.$$

Hint. The continuous identity holds for *every* $v \in V$; in particular for every $v_h \in V_h$, since $V_h \subset V$. Write down the continuous and discrete identities for the same test function v_h, subtract, and use bilinearity. The entire content is the containment $V_h \subset V$ — this is why the Galerkin space must be a genuine subspace, not merely some finite-dimensional space nearby.

Problem 4 — Céa's lemma (the target)

Work it

Using *only* coercivity (Problem 1), boundedness (Problem 2), and Galerkin orthogonality (Problem 3), prove Céa's lemma:

$$\|u - u_h\|_V \le \frac{M}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V.$$

Aim for four lines. If you reach it unaided, you have earned the right to read Chapter 3 at speed; if you stall, note precisely where, and *that* is the informative outcome.

Hint (last resort). Start from $\alpha\|u - u_h\|_V^2 \le a(u - u_h, u - u_h)$. For an arbitrary $v_h \in V_h$, split the second argument as $(u - v_h) + (v_h - u_h)$. The term $a(u - u_h,\, v_h - u_h)$ vanishes — why? because $v_h - u_h \in V_h$ and Problem 3 applies. Bound what remains by boundedness, cancel one factor of $\|u - u_h\|_V$, and take the infimum over v_h.

Load-bearing idea

The order is not arbitrary. Problems 1–2 establish that the abstract machine (Lax–Milgram) applies, so both u and u_h exist and are unique. Problem 3 is the geometric heart: the error is a-orthogonal to the whole discrete space, i.e. u_h is the a-projection of u onto V_h. Problem 4 converts that geometry into a *quantitative* bound: the finite element error is, up to the constant M/α, the best approximation the space V_h can offer. Everything else in the finite element error analysis — interpolation estimates, the Aubin–Nitsche L^2 improvement, the parabolic Ritz splitting, and their stochastic descendants in Chapter 10 — is this four-step ritual, refined. Master these four and the rest of the subject has a spine to hang on.

Where this bites

Do not read the solutions in Appendix B until you have a genuine attempt at all four on paper. A fluent reading of a clean proof produces the *sensation* of understanding without the substance; the only reliable test is whether you can produce the argument yourself, on a blank page, a week later. When you send these to a reader — or sit them in front of a supervisor — it is the reproduction that counts, not the recognition. Attempt first. The crossings-out are where the learning is.

CHAPTER 5

The deterministic heat equation as rehearsal

5.1 The picture before the machinery

Plain-language picture

Before adding randomness, we should understand the equation without it, thoroughly, because almost every difficulty in the stochastic case is already present — in tamer form — in the deterministic one. The deterministic heat equation describes how temperature spreads through a rod: heat flows from hot to cold, sharp differences blur, and everything tends toward a smooth equilibrium.
The three things we most need to steal from the deterministic theory are: *how* the solution smooths (the semigroup and its estimates), *how fast* a numerical method converges to it (the parabolic error analysis), and *how* the two combine in the "method of lines" that turns a PDE into a big system of ODEs. Get these three clean here, in the absence of noise, and the stochastic chapter becomes an exercise in adding one carefully-estimated random term to a structure you already own.

5.2 The equation and its modal solution

On $\Omega = (0,1)$ with Dirichlet conditions, the deterministic heat equation is

$$\partial_t u = \Delta u = \partial_x^2 u, \qquad u(x,0) = u_0(x), \qquad u(0,t) = u(1,t) = 0. \tag{5.1}$$

Expanding in the eigenbasis $e_k = \sqrt{2}\sin(k\pi x)$ of Example 2.15 and writing $u(t) = \sum_k \hat{u}_k(t) e_k$, each mode obeys the scalar linear ODE $\hat{u}_k'(t) = -\lambda_k \hat{u}_k(t)$, so $\hat{u}_k(t) = e^{-\lambda_k t}\hat{u}_k(0)$ and

$$u(x,t) = \sum_{k\geq 1} e^{-\lambda_k t} \langle u_0, e_k \rangle \, e_k(x) = S(t)u_0. \tag{5.2}$$

This *is* the heat semigroup of Chapter 6, now derived by separation of variables rather than abstract generation theory — the two viewpoints coincide, and it is worth holding both.

Intuition

The factor $e^{-\lambda_k t} = e^{-k^2\pi^2 t}$ damps high modes (k large) far faster than low ones: after any positive time, the sharp features carried by high modes have vanished and only the smooth low modes survive. This is smoothing, seen in the clearest possible terms — a diagonal exponential decay whose rate grows quadratically with the mode number. When randomness is added, the *same* damping fights the *same* high modes of the noise; the entire question of SPDE regularity is whether the damping wins, mode by mode, and by how much.

5.3 The method of lines and stiffness

Discretising (5.1) in space alone — by finite elements or finite differences — but leaving time continuous produces the *method of lines*: a large system of coupled ODEs

$$M\,\dot{U}(t) = -A\,U(t), \qquad U(0) = U_0,$$

where A and M are the stiffness and mass matrices of Example 3.5. The eigenvalues of the pencil (A, M) approximate the continuous $\lambda_k = k^2\pi^2$, so after refining to N nodes the largest discrete eigenvalue is of order N^2. This is *stiffness*: the system contains dynamics on timescales from $O(1)$ (slow, low modes) down to $O(N^{-2})$ (fast, high modes), and an explicit time-stepper must resolve the fastest to remain stable.

Load-bearing idea

Stiffness is the reason the choice of time-stepper matters so much for parabolic problems, deterministic or stochastic. An explicit Euler step is stable only for $\Delta t \lesssim \lambda_{\max}^{-1} \sim N^{-2}$ — prohibitively small. Implicit and exponential integrators remove this restriction by treating the stiff linear part without an explicit stability constraint. The exponential integrator does so *exactly*, via the matrix exponential $e^{-\Delta t\, A}$, which is the discrete shadow of the semigroup $S(\Delta t)$. Every remark about time-stepping the stochastic heat equation in Chapter 10 is inherited, unchanged, from this deterministic stiffness analysis.

5.4 Energy estimates: stability without eigenvalues

The modal solution is available only because the domain and operator are simple. The technique that survives into general domains, variable coefficients, and nonlinearities — and into the stochastic theory — is the *energy method*. Multiply (5.1) by u and integrate over Ω:

$$\int_\Omega u\,\partial_t u = \int_\Omega u\,\Delta u \implies \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_{L^2}^2 = -\|\nabla u\|_{L^2}^2 \le 0,$$

using integration by parts and the boundary conditions. Thus $\|u(t)\|_{L^2}$ is non-increasing — the fundamental *energy dissipation* of the heat equation. Integrating in time and using Poincaré gives exponential decay $\|u(t)\|_{L^2} \le e^{-\lambda_1 t}\|u_0\|_{L^2}$.

Thread to the SPDE

The energy method is the workhorse of the parabolic finite element error analysis (Chapter 3's Ritz-splitting energy estimate) and of the stochastic theory (the moment bounds on the SPDE solution). Its virtue is that it needs no eigenbasis: it works on any domain and survives the addition of a stochastic forcing, where the Itô correction (Chapter 7) contributes an extra $+\frac{1}{2}\|\cdot\|_{\mathrm{HS}}^2\,\mathrm{d}t$ term to the energy balance — the deterministic dissipation, plus a noise-driven energy injection, in perpetual competition. That competition is exactly the stationary variance $\frac{1}{2\lambda_k}$ we computed mode by mode in Chapter 9.

Work it

(1) Derive the energy dissipation identity $\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|^2 = -\|\nabla u\|^2$ and combine with Poincaré to get the decay rate $\lambda_1 = \pi^2$. (2) For the method-of-lines system $M\dot{U} = -AU$ with the uniform-mesh A, M of Example 3.5, estimate the largest eigenvalue as a function of N and confirm the $O(N^2)$ stiffness. (3) Implement explicit Euler and observe the instability when $\Delta t > c\,N^{-2}$; then implement the exponential step $U^{n+1} = e^{-\Delta t\,M^{-1}A}U^n$ and confirm unconditional stability. This is the deterministic dress rehearsal for the SPDE capstone.

CHAPTER 6

Operator semigroups and the heat kernel

6.1 The picture before the machinery

Plain-language picture

Drop a blob of ink in still water and photograph it every second. The pictures form a sequence, each obtained from the last by the same physical rule: "let diffusion act for one more second." Crucially, applying the rule for three seconds is the same as applying it for one second and then for two more — time adds. This "family of operations indexed by elapsed time, that compose by adding the times" is what a semigroup is.
The heat equation's solution operator is exactly such a family: $S(t)$ takes an initial temperature profile and returns the profile after time t. It has one property that makes the whole stochastic theory work: it *smooths*. However jagged the initial data, after any positive time the profile is infinitely smooth — the sharp ink boundary blurs instantly. This smoothing is quantitative, and its rate is the hinge on which every SPDE error estimate turns: the noise is rough, but the semigroup smooths it just fast enough that the solution is a function at all.

6.2 The abstract Cauchy problem and the resolvent

The reason semigroups matter for us is that they *solve* evolution equations. The link is the abstract Cauchy problem

$$u'(t) = Au(t)\ (t > 0), \qquad u(0) = u_0, \tag{6.1}$$

whose solution, whenever A generates a C_0-semigroup, is simply $u(t) = S(t)u_0$. For the heat equation, $A = \Delta$ and $S(t) = e^{t\Delta}$ is the heat semigroup; (6.1) is then the heat equation in disguise, and the semigroup is its solution operator. When a forcing term is present, $u' = Au + f$, the variation-of-constants formula

$$u(t) = S(t)u_0 + \int_0^t S(t-s)f(s)\ \mathrm{d}s$$

gives the solution — and this is the deterministic ancestor of the mild solution (9.2), with the stochastic forcing $\mathrm{d}W$ replacing $f\ \mathrm{d}s$.

The *resolvent* $R(\lambda, A) = (\lambda I - A)^{-1}$ is the algebraic bridge between A and $S(t)$, encoded in the Laplace-transform identity

$$R(\lambda, A) = \int_0^\infty e^{-\lambda t} S(t)\ \mathrm{d}t \qquad (\mathrm{Re}\,\lambda \text{ large}),$$

the operator analogue of $\frac{1}{\lambda - a} = \int_0^\infty e^{-\lambda t} e^{at}\,\mathrm{d}t$. For a self-adjoint A with eigenpairs (λ_k, e_k) the resolvent, like the semigroup, acts diagonally: $R(\lambda, A)$ multiplies mode k by $(\lambda - \lambda_k)^{-1}$, and is

bounded precisely when λ avoids the spectrum $\{\lambda_k\}$. The Hille–Yosida theorem is, at bottom, the statement that controlling the resolvent norm for all $\lambda > 0$ is equivalent to the semigroup being a contraction.

Example 6.1 (The scalar heat mode as a one-parameter semigroup)**.** The simplest instance: on $X = \mathbb{R}$ take $A = -\lambda$ (multiplication by a negative constant). Then $S(t) = e^{-\lambda t}$, obviously a C_0-semigroup of contractions, with generator $-\lambda$, resolvent $(\mu + \lambda)^{-1}$, and smoothing estimate $|(-\lambda)^\gamma S(t)| = \lambda^\gamma e^{-\lambda t}$ — exactly the k-th diagonal entry of the heat semigroup. The full heat semigroup on L^2 is nothing but the direct sum over k of these scalar semigroups, one per mode, with $\lambda = \lambda_k$. Everything proved abstractly in this chapter reduces, in the eigenbasis, to this one-line scalar computation repeated over all modes — which is why the eigenbasis is worth its weight in gold.

6.3 C_0-semigroups and their generators

Definition 6.2 (C_0-semigroup)**.** A family $\{S(t)\}_{t \geq 0}$ of bounded linear operators on a Banach space X is a *strongly continuous (or C_0-) semigroup* if

(i) $S(0) = I$;

(ii) $S(t+s) = S(t)S(s)$ for all $t, s \geq 0$ (the semigroup law);

(iii) $t \mapsto S(t)x$ is continuous from $[0, \infty)$ to X for each fixed $x \in X$.

Definition 6.3 (Generator)**.** The *infinitesimal generator* A of $\{S(t)\}$ is

$$Ax = \lim_{t \downarrow 0} \frac{S(t)x - x}{t},$$

with domain $\mathcal{D}(A)$ the set of x for which this limit exists in X.

The generator is to the semigroup what the derivative at 0 is to the exponential: formally $S(t) = e^{tA}$, and $u(t) = S(t)u_0$ solves the abstract Cauchy problem $u'(t) = Au(t)$, $u(0) = u_0$. The heat semigroup is generated by the Laplacian.

Theorem 6.4 (Hille–Yosida, contraction case)**.** *A densely defined operator A generates a C_0-semigroup of contractions ($\|S(t)\| \leq 1$) if and only if A is closed and, for every $\lambda > 0$, $\lambda \in \rho(A)$ (the resolvent set) with $\|(\lambda I - A)^{-1}\| \leq 1/\lambda$.*

We will not prove Hille–Yosida — its role here is conceptual — but we will use the resolvent $(\lambda I - A)^{-1}$ heavily, as it is the bridge between the semigroup and the spectral decomposition of A.

6.4 The heat semigroup by spectral decomposition

On $\Omega = (0,1)$ with Dirichlet conditions, let $A = -\Delta = -\partial_x^2$ with domain $\mathcal{D}(A) = H^2(0,1) \cap H_0^1(0,1)$. The operator A is self-adjoint, positive, with compact resolvent, so by the spectral theorem it has an orthonormal basis of eigenfunctions in $L^2(0,1)$:

$$Ae_k = \lambda_k e_k, \qquad e_k(x) = \sqrt{2}\sin(k\pi x), \qquad \lambda_k = k^2\pi^2, \quad k \geq 1.$$

Every $u \in L^2$ expands as $u = \sum_k \langle u, e_k \rangle e_k$, and the semigroup acts diagonally:

Proposition 6.5 (Heat semigroup in the eigenbasis)**.** *The heat semigroup $S(t) = e^{-tA}$ is*

$$S(t)u = \sum_{k \geq 1} e^{-\lambda_k t} \langle u, e_k \rangle \, e_k.$$

It is a C_0-semigroup of contractions on L^2; indeed $\|S(t)u\|_{L^2}^2 = \sum_k e^{-2\lambda_k t} |\langle u, e_k \rangle|^2 \leq \|u\|_{L^2}^2$.

Proof. The semigroup law follows from $e^{-\lambda_k(t+s)} = e^{-\lambda_k t}e^{-\lambda_k s}$ acting mode by mode; $S(0) = I$ is the completeness of $\{e_k\}$; strong continuity and the contraction bound follow from dominated convergence on the Fourier coefficients, each factor $e^{-\lambda_k t} \in (0, 1]$. That its generator is $-A$ is a direct computation of the limit $t^{-1}(S(t)u - u) \to \sum_k (-\lambda_k)\langle u, e_k\rangle e_k = -Au$, valid on $\mathcal{D}(A)$. □

6.4.1 Fractional powers and the smoothing estimate

Because A is positive self-adjoint, its fractional powers are defined mode by mode:

$$A^\gamma u = \sum_k \lambda_k^\gamma \langle u, e_k\rangle e_k, \qquad \mathcal{D}(A^\gamma) = \Big\{u : \sum_k \lambda_k^{2\gamma} |\langle u, e_k\rangle|^2 < \infty\Big\}.$$

These domains are the "smoothness scale": $\mathcal{D}(A^{1/2}) = H_0^1$, $\mathcal{D}(A) = H^2 \cap H_0^1$, and in general $\mathcal{D}(A^{s/2}) = \dot{H}^s$, the homogeneous Sobolev space of order s. Now the estimate that governs everything.

Theorem 6.6 (Analytic smoothing estimate). *For every $\gamma \geq 0$ there is a constant C_γ such that*

$$\|A^\gamma S(t)\|_{\mathcal{L}(L^2)} \leq C_\gamma\, t^{-\gamma} \qquad \text{for all } t > 0.$$

Consequently $S(t)$ maps L^2 into $\mathcal{D}(A^\gamma)$ for every γ and every $t > 0$: the semigroup is smoothing.

Proof. Acting in the eigenbasis, $A^\gamma S(t)$ multiplies the k-th mode by $\lambda_k^\gamma e^{-\lambda_k t}$. The operator norm on L^2 of a diagonal multiplier is the supremum of the multipliers:

$$\|A^\gamma S(t)\| = \sup_{k \geq 1} \lambda_k^\gamma e^{-\lambda_k t} \leq \sup_{\lambda > 0} \lambda^\gamma e^{-\lambda t}.$$

Maximising $g(\lambda) = \lambda^\gamma e^{-\lambda t}$ over $\lambda > 0$: $g'(\lambda) = 0$ at $\lambda_* = \gamma/t$, giving $g(\lambda_*) = (\gamma/t)^\gamma e^{-\gamma} = (\gamma^\gamma e^{-\gamma})\, t^{-\gamma}$. Set $C_\gamma = \gamma^\gamma e^{-\gamma}$ (and $C_0 = 1$). □

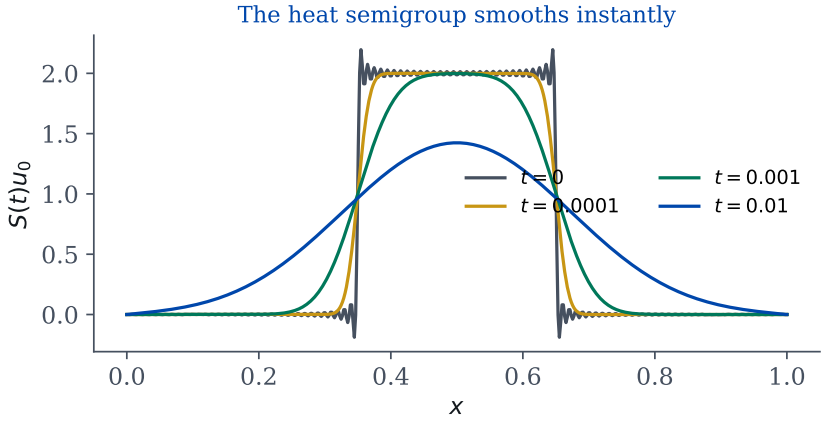

Figure 6.1: The heat semigroup $S(t) = e^{-tA}$ applied to a rough (indicator-like) initial profile, expanded in 200 sine modes. At $t = 0$ the profile has jumps; for any $t > 0$ it is already analytic, and the corners are gone almost instantly. The high modes, damped by $e^{-\lambda_k t}$ with $\lambda_k = k^2\pi^2$, are annihilated fastest — this selective damping of high frequencies is precisely the smoothing that Theorem 6.6 quantifies and that tames white noise in the SPDE.

Load-bearing idea

The estimate $\|A^\gamma S(t)\| \le C_\gamma t^{-\gamma}$ is the most-used inequality in the numerical analysis of parabolic SPDEs. Read it as a trade: the semigroup will hand you γ extra derivatives of smoothness, but it charges you a singularity $t^{-\gamma}$ at time zero. The whole art of proving SPDE convergence rates is to choose γ large enough that the noise becomes manageable, yet small enough that the $t^{-\gamma}$ singularity remains integrable against $\mathrm{d}t$ (which requires $\gamma < 1$). The optimal exponents $1/2$ and $1/4$ from the orientation are precisely the outcome of balancing this trade at the borderline of integrability.

Thread to the SPDE

Here is the estimate doing its job, in preview. The stochastic convolution is $\int_0^t S(t-s)\,\mathrm{d}W(s)$. Its variance involves $\int_0^t \|S(t-s)\|_{\mathrm{HS}}^2\,\mathrm{d}s$, where the Hilbert–Schmidt norm counts how the semigroup damps each noise mode. Because the noise excites all modes equally (white noise), the sum over modes converges only thanks to the exponential damping $e^{-\lambda_k(t-s)}$ inside $S(t-s)$ — and the rate at which it converges, controlled by Theorem 6.6, is exactly what yields the $s < 1/2$ spatial regularity in $d = 1$ and its failure in $d = 2$. The smoothing estimate is the quantitative heart of the dimensional threshold.

6.5 Exponential integrators: exploiting the linear part exactly

A practical payoff of the semigroup viewpoint, central to state-of-the-art SPDE numerics. When time-stepping $u' = Au + g$, a naive Euler scheme treats Au by a difference quotient and inherits the stiffness of A (whose eigenvalues $\lambda_k = k^2\pi^2 \to \infty$). The *exponential integrator* instead solves the linear part exactly using $S(\Delta t) = e^{\Delta t A}$:

$$u^{n+1} = S(\Delta t)u^n + \int_{t_n}^{t_{n+1}} S(t_{n+1} - s)\, g(s)\,\mathrm{d}s,$$

approximating only the integral of the (smoother) nonlinearity g. For the stochastic heat equation this becomes the *exponential Euler* scheme, which handles the stiff linear diffusion without a time-step restriction and achieves the optimal temporal rate. We construct it explicitly in Chapter 10; the point here is that the semigroup is not merely an analytical device but the computational core of the best schemes.

Work it

(1) Verify that $g(\lambda) = \lambda^\gamma e^{-\lambda t}$ is maximised at $\lambda = \gamma/t$ and confirm the constant $C_\gamma = \gamma^\gamma e^{-\gamma}$. (2) For $A = -\partial_x^2$ on $(0,1)$, implement $S(t)$ in the sine basis and verify numerically that $\|A^{1/2}S(t)u\|_{L^2} \sim t^{-1/2}$ as $t \downarrow 0$ for generic $u \in L^2$. (3) Confirm that $\mathcal{D}(A^{1/2}) = H_0^1$ by checking that $\sum_k \lambda_k |\langle u, e_k\rangle|^2 < \infty$ is equivalent to $\|u'\|_{L^2}^2 < \infty$ via Parseval.

CHAPTER 7

Stochastic calculus and the numerics of SDEs

7.1 The picture before the machinery

Plain-language picture

A pollen grain in water jitters, kicked randomly by molecules. Its path is continuous — it never teleports — yet so ragged that at no instant does it have a well-defined velocity. Ordinary calculus, built on velocities, cannot describe it. Itô's calculus is the replacement: a way to do calculus along paths that are continuous but infinitely wiggly.
The one rule you must internalise is that for such paths, small time steps and small space steps are not comparable in the usual way. Over a time step Δt, the random displacement is of size $\sqrt{\Delta t}$, not Δt. Squared, that displacement is of size Δt — it does *not* vanish faster than Δt the way $(\Delta t)^2$ would. This single fact, that "the square of the noise is of order $\mathrm{d}t$," is the entire difference between Itô calculus and ordinary calculus, and it is why numerical schemes for random equations converge at fractional rates like $1/2$ rather than the integer rates of deterministic methods.

We develop finite-dimensional stochastic calculus and its numerics first because — this is the reason the chapter exists — the convergence *rates* and their *proofs* lift almost unchanged to the infinite-dimensional SPDE setting, where the analysis is cluttered by functional-analytic overhead. Learn the mechanism where it is clean.

7.2 Brownian motion and the Itô integral

Definition 7.1 (Brownian motion)**.** A standard *Brownian motion* $\{W(t)\}_{t\geq 0}$ is a real-valued process with $W(0) = 0$, independent increments, $W(t) - W(s) \sim \mathcal{N}(0, t-s)$ for $s < t$, and continuous sample paths.

Brownian paths have infinite total variation but finite *quadratic* variation: partitioning $[0,t]$ ever more finely, $\sum_i |W(t_{i+1}) - W(t_i)|^2 \to t$ in probability. This is the rigorous form of "the square of the noise is of order $\mathrm{d}t$," usually written $(\mathrm{d}W)^2 = \mathrm{d}t$.

Because Brownian paths have unbounded variation, the integral $\int_0^t \sigma(s)\,\mathrm{d}W(s)$ cannot be defined pathwise as a Riemann–Stieltjes integral. Itô's construction defines it for adapted, square-integrable integrands as an L^2-limit of left-endpoint Riemann sums, $\int_0^t \sigma\,\mathrm{d}W = \lim \sum_i \sigma(t_i)(W(t_{i+1}) - W(t_i))$. The left-endpoint choice is essential and is what makes the integral a martingale.

Theorem 7.2 (Itô isometry). *For adapted σ with $\mathbb{E}\int_0^t \sigma(s)^2\,\mathrm{d}s < \infty$,*

$$\mathbb{E}\Big[\Big(\int_0^t \sigma(s)\,\mathrm{d}W(s)\Big)^2\Big] = \mathbb{E}\int_0^t \sigma(s)^2\,\mathrm{d}s.$$

The isometry is the workhorse: it converts the second moment of a stochastic integral into an ordinary (Lebesgue) integral of the second moment of the integrand. Every variance computation in the theory — including, in Chapter 9, the regularity of the stochastic convolution — is an application of it.

Theorem 7.3 (Itô formula). *If X satisfies $\mathrm{d}X = b\,\mathrm{d}t + \sigma\,\mathrm{d}W$ and $f \in C^2$, then*

$$\mathrm{d}f(X) = f'(X)\,\mathrm{d}X + \tfrac{1}{2}f''(X)\,\sigma^2\,\mathrm{d}t = \big(f'(X)b + \tfrac{1}{2}f''(X)\sigma^2\big)\,\mathrm{d}t + f'(X)\sigma\,\mathrm{d}W.$$

The second-derivative term $\frac{1}{2}f''\sigma^2\,\mathrm{d}t$ is the Itô correction; it is present precisely because $(\mathrm{d}W)^2 = \mathrm{d}t$ rather than 0. It has no counterpart in ordinary calculus and is the source of every "surprise" in stochastic analysis.

7.3 Stochastic differential equations

Definition 7.4 (SDE). A *stochastic differential equation* is

$$\mathrm{d}X(t) = b(X(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t), \qquad X(0) = X_0,$$

shorthand for the integral equation $X(t) = X_0 + \int_0^t b(X)\,\mathrm{d}s + \int_0^t \sigma(X)\,\mathrm{d}W$.

Theorem 7.5 (Existence and uniqueness). *If b, σ are globally Lipschitz and satisfy a linear growth bound, the SDE has a unique (strong) solution with continuous paths and $\mathbb{E}\sup_{t\le T}|X(t)|^2 < \infty$.*

The proof is a Picard iteration in which the Itô isometry replaces the deterministic estimate on the integral term — structurally identical to the Cauchy–Lipschitz theorem for ODEs, with Theorem 7.2 doing the work.

7.4 Numerical schemes: strong and weak convergence

Two distinct notions of "the numerical solution is close to the true one" coexist, and keeping them apart is essential.

Definition 7.6 (Strong and weak convergence). A scheme producing approximations X^N (with N steps, step size $\Delta t = T/N$) converges with *strong order* p if

$$\big(\mathbb{E}|X(T) - X^N|^2\big)^{1/2} \le C(\Delta t)^p \quad \text{(pathwise accuracy)},$$

and with *weak order* q if for all smooth bounded φ,

$$|\mathbb{E}\varphi(X(T)) - \mathbb{E}\varphi(X^N)| \le C(\Delta t)^q \quad \text{(distributional accuracy)}.$$

Intuition

Strong convergence asks the numerical path to track the *same* Brownian path's true trajectory — pathwise fidelity. Weak convergence asks only that the numerical solution have approximately the right *statistics* — the right mean, variance, distribution — regardless of whether any individual path is tracked. Weak is easier, and weak orders are typically *double* the strong orders. For pricing an option you need only weak accuracy (you want $\mathbb{E}$ of a payoff); for pathwise control you need strong. In the SPDE setting the same dichotomy holds, and the doubling reappears — the same doubling seen in Chapter 3's H^1-versus-L^2 rates. Three different doublings, one underlying reason: measuring in a weaker sense borrows an extra order.

7.4.1 Euler–Maruyama

Definition 7.7 (Euler–Maruyama). $X^{n+1} = X^n + b(X^n)\Delta t + \sigma(X^n)\Delta W^n$, where $\Delta W^n = W(t_{n+1}) - W(t_n) \sim \mathcal{N}(0, \Delta t)$ are the exact Brownian increments.

Theorem 7.8 (Order of Euler–Maruyama). *Under global Lipschitz and linear growth hypotheses, Euler–Maruyama has strong order $p = 1/2$ and weak order $q = 1$.*

Idea of the strong estimate. Let $e_n = X(t_n) - X^n$. Over one step, the local error splits into a drift part (of size $(\Delta t)^2$, negligible) and a diffusion part. The diffusion increment's error comes from freezing $\sigma(X(s))$ at its left endpoint $\sigma(X(t_n))$ over the step. By the Itô isometry the mean-square of this frozen-coefficient error is $\int_{t_n}^{t_{n+1}} \mathbb{E}|\sigma(X(s)) - \sigma(X(t_n))|^2 \, \mathrm{d}s$; Lipschitz continuity of σ and the $\sqrt{\Delta t}$-size of Brownian increments bound $\mathbb{E}|X(s) - X(t_n)|^2 \le C(s - t_n)$, giving a local mean-square error of order $(\Delta t)^2$ per step. Summed over $N = T/\Delta t$ steps and closed by a discrete Grönwall inequality, the global mean-square error is $O(\Delta t)$, i.e. strong order $1/2$ in the root-mean-square. □

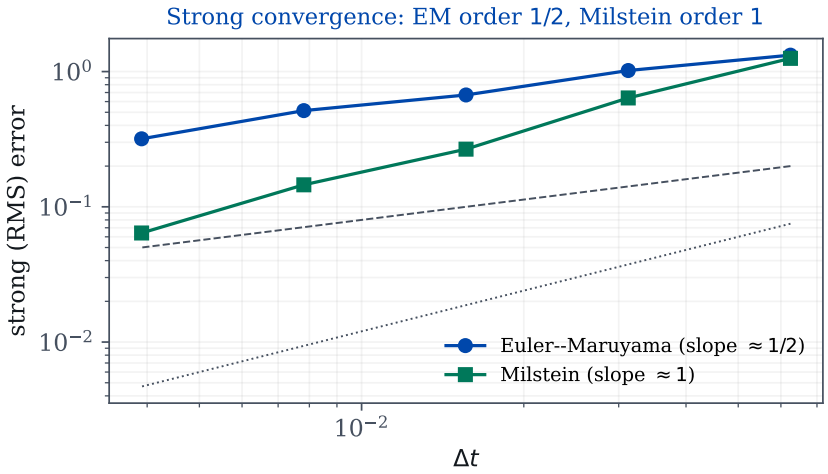

Figure 7.1: Strong convergence for geometric Brownian motion $\mathrm{d}X = \mu X \, \mathrm{d}t + \sigma X \, \mathrm{d}W$ (exact solution known), 2000 sample paths per step size, in the author's house style. Euler–Maruyama converges at strong order $1/2$ (slope ≈ 0.5); adding the single Milstein term $\frac{1}{2}\sigma\sigma'((\Delta W)^2 - \Delta t)$ lifts the order to 1 (slope $\approx$ 1.0). This is the finite-dimensional rehearsal for every SPDE convergence experiment: the same log–log regression, the same fractional exponents imposed by the $\sqrt{\Delta t}$ scale of Brownian increments.

Load-bearing idea

The strong order $1/2$ is not a defect of the method; it is imposed by the $\sqrt{\Delta t}$ scale of Brownian increments. To beat it one must add information about how the noise varies *within* a step — that is exactly what the Milstein scheme does, by including the next term of the stochastic Taylor expansion. In infinite dimensions the analogue of this order barrier is the spatial/temporal Hölder regularity of the SPDE solution, and the analogue of "add the next Taylor term" is the exponential-integrator and Milstein-type SPDE schemes of Chapter 10.

7.4.2 Milstein

Expanding $\sigma(X(s))$ by Itô's formula inside the diffusion integral produces one more computable term, the *Lévy area* contribution $\frac{1}{2}\sigma\sigma'((\Delta W)^2 - \Delta t)$:

$$X^{n+1} = X^n + b\Delta t + \sigma\Delta W^n + \tfrac{1}{2}\sigma\sigma'\big((\Delta W^n)^2 - \Delta t\big).$$

Theorem 7.9 (Order of Milstein). *The Milstein scheme has strong order $p = 1$ (and weak order $q = 1$).*

The extra term captures the leading intra-step fluctuation of the diffusion coefficient, lifting the strong order from $1/2$ to 1. In dimension one it requires only $\sigma\sigma'$; in higher dimensions it requires simulating iterated Itô integrals (Lévy areas) unless a commutativity condition holds — a genuine practical obstruction, and a first taste of why multi-dimensional and infinite-dimensional stochastic numerics are hard.

7.5 Multilevel Monte Carlo

When only weak (statistical) accuracy is needed — computing $\mathbb{E}\varphi(X(T))$ — a variance-reduction idea of Giles slashes the cost. Naive Monte Carlo with bias $O(\Delta t)$ and statistical error requires balancing discretisation against sample count, costing $O(\varepsilon^{-3})$ for accuracy ε. *Multilevel Monte Carlo* writes the fine-level expectation as a telescoping sum over a hierarchy of step sizes,

$$\mathbb{E}[\varphi(X^L)] = \mathbb{E}[\varphi(X^0)] + \sum_{\ell=1}^{L} \mathbb{E}[\varphi(X^\ell) - \varphi(X^{\ell-1})],$$

and estimates each difference with its own sample count. Because coarse and fine paths driven by the same Brownian motion are strongly correlated, the differences have small variance, and most samples can be taken cheaply at coarse levels. The cost falls to $O(\varepsilon^{-2})$ — optimal. This is the single most important variance-reduction technique for both SDEs and SPDEs, and it rests directly on the *strong* convergence rate (Theorem 7.8) controlling the level variances.

Work it

(1) Simulate geometric Brownian motion $\mathrm{d}X = \mu X\,\mathrm{d}t + \sigma X\,\mathrm{d}W$ (exact solution known) and empirically confirm Euler–Maruyama strong order $1/2$ by regressing $\log(\text{RMS error})$ on $\log\Delta t$; expect slope ≈ 0.5. (2) Add the Milstein term and confirm the slope jumps to ≈ 1.0. (3) Confirm weak order 1 for Euler–Maruyama by tracking the error in $\mathbb{E}[X(T)]$ rather than the pathwise RMS. These three plots are the finite-dimensional rehearsal for every SPDE convergence experiment you will run.

CHAPTER 8

Gaussian measures and Wiener processes in Hilbert space

8.1 The picture before the machinery

Plain-language picture

In finite dimensions, a Gaussian random vector is described by a mean and a covariance matrix; the matrix's eigenvectors are the independent directions of randomness and its eigenvalues are the variances along them. To randomise an entire *function* — a temperature profile over an interval, say — we need infinitely many such directions, one for each vibration mode of the interval. The covariance "matrix" becomes a covariance *operator* on an infinite-dimensional space.

The subtlety that has no finite-dimensional analogue: for the randomness to be an honest function (rather than an unmanageably rough object), the variances along the infinitely many directions must *sum to something finite* — the covariance operator must be trace-class. Space–time white noise is the extreme case where every direction has *equal* variance, so the variances do not sum — the covariance operator is the identity, which is not trace-class. That single failure of summability is why white noise is so rough, why the stochastic heat equation is delicate, and why dimension one is the last dimension in which the equation has a function-valued solution.

8.2 Gaussian measures and covariance operators

Let $\mathcal{H}$ be a separable Hilbert space (think $L^2(0,1)$).

Definition 8.1 (Gaussian measure). A probability measure μ on $\mathcal{H}$ is *Gaussian* if for every $h \in \mathcal{H}$ the real random variable $x \mapsto \langle x, h\rangle$ (under $x \sim \mu$) is Gaussian. Its *mean* $m \in \mathcal{H}$ and *covariance operator* $Q : \mathcal{H} \to \mathcal{H}$ are defined by $\langle m, h\rangle = \mathbb{E}\langle x, h\rangle$ and $\langle Qh, k\rangle = \mathbb{E}[\langle x - m, h\rangle\langle x - m, k\rangle]$.

Theorem 8.2 (Characterisation). *Q is a covariance operator of a Gaussian measure on $\mathcal{H}$ if and only if it is self-adjoint, non-negative, and* trace-class*: $\operatorname{Tr} Q = \sum_k \langle Qe_k, e_k\rangle < \infty$ for one (hence every) orthonormal basis $\{e_k\}$.*

Load-bearing idea

Trace-class is the dividing line. If Q has eigenvalues $q_k \geq 0$ with $\sum_k q_k < \infty$, a genuine $\mathcal{H}$-valued Gaussian random element exists, representable as $x = m + \sum_k \sqrt{q_k}\,\xi_k e_k$ with i.i.d. standard normals ξ_k — the *Karhunen–Loève expansion*. The sum converges in $\mathcal{H}$ precisely because $\mathbb{E}\|x - m\|^2 = \sum_k q_k = \operatorname{Tr} Q < \infty$. When $\sum_k q_k = \infty$ — as for white noise, where every $q_k = 1$ — the series does not converge in $\mathcal{H}$, and the "random function" lives only in a larger space of distributions. Every difficulty in the SPDE traces to this one divergent sum.

Example 8.3 (Karhunen–Loève is PCA for random functions)**.** You have already implemented the finite-dimensional case of this expansion under another name. Principal component analysis diagonalises a covariance matrix and writes a random vector as a sum of uncorrelated components along the eigenvectors, weighted by the square roots of the eigenvalues. Karhunen–Loève is exactly this for a random *function*: the covariance operator Q replaces the covariance matrix, its eigenfunctions e_k replace the principal directions, and its eigenvalues q_k are the variances along them. A sample of the random function is drawn by generating independent standard normals ξ_k and forming $\sum_k \sqrt{q_k}\,\xi_k e_k$ — literally the code you would write for PCA sampling, with the sum now infinite. The trace-class condition $\sum_k q_k < \infty$ is the requirement that the total variance across all infinitely many directions be finite, without which the "sample" has infinite expected norm and is not a function.

Intuition

This connection is worth internalising because it demystifies the entire infinite-dimensional Gaussian theory: it is finite-dimensional Gaussian intuition, plus a convergence condition on the sum of variances. A Q-Wiener process is a time-evolving Karhunen–Loève expansion in which each coefficient ξ_k is replaced by an independent Brownian motion $\beta_k(t)$; a cylindrical Wiener process is the same with all variances equal to one, and the convergence condition fails, which is the whole content of "white noise is not a function." When the abstraction feels heavy, return to the PCA picture: it is never more than that, plus bookkeeping about which sums converge.

8.3 Q-Wiener and cylindrical Wiener processes

Definition 8.4 (Q-Wiener process)**.** For trace-class Q, a *Q-Wiener process* is an $\mathcal{H}$-valued process

$$W(t) = \sum_k \sqrt{q_k}\,\beta_k(t)\,e_k,$$

where $\{e_k\}$ are the eigenvectors of Q, q_k the eigenvalues, and β_k independent standard real Brownian motions. It has continuous $\mathcal{H}$-valued paths and $\mathbb{E}\|W(t)\|^2 = t\,\operatorname{Tr} Q$.

Definition 8.5 (Cylindrical Wiener process / space–time white noise)**.** Taking $Q = I$ (all $q_k = 1$) gives the *cylindrical Wiener process*

$$W(t) = \sum_k \beta_k(t)\,e_k.$$

This series does *not* converge in $\mathcal{H}$ (since $\sum_k 1 = \infty$), but it converges in any larger Hilbert space $\mathcal{H}_1 \supset \mathcal{H}$ into which $\mathcal{H}$ embeds by a Hilbert–Schmidt map. Its time derivative $\dot{W}$ is *space–time white noise*: formally, a Gaussian field with $\mathbb{E}[\dot{W}(x,t)\dot{W}(y,s)] = \delta(x-y)\delta(t-s)$.

Intuition

Space–time white noise is "equal randomness in every mode and at every instant, all independent." Its total energy is infinite — that is the divergent $\sum_k 1$ — so it cannot be a function. But when it is fed through the heat semigroup, which damps mode k by $e^{-\lambda_k t}$ with $\lambda_k = k^2\pi^2 \to \infty$, the high modes are suppressed fast enough that the *output* has finite energy in one spatial dimension. The stochastic heat equation is, in one phrase, "infinitely rough noise, smoothed just enough by diffusion to become a function — but only in $d = 1$."

8.4 The stochastic integral in Hilbert space

To make sense of $\int_0^t \Phi(s)\,\mathrm{d}W(s)$ for operator-valued integrands Φ, one repeats Itô's construction. The correct isometry involves the *Hilbert–Schmidt* norm.

Definition 8.6 (Hilbert–Schmidt operators). An operator $\Phi : \mathcal{H} \to \mathcal{H}$ is *Hilbert–Schmidt* if $\|\Phi\|_{\mathrm{HS}}^2 := \sum_k \|\Phi e_k\|^2 < \infty$. The Hilbert–Schmidt operators form a Hilbert space.

Theorem 8.7 (Itô isometry in Hilbert space). *For a suitable adapted operator-valued integrand Φ and a Q-Wiener process W,*

$$\mathbb{E}\Big\|\int_0^t \Phi(s)\,\mathrm{d}W(s)\Big\|_{\mathcal{H}}^2 = \mathbb{E}\int_0^t \|\Phi(s)Q^{1/2}\|_{\mathrm{HS}}^2\,\mathrm{d}s.$$

For a cylindrical Wiener process ($Q = I$) this reads $\mathbb{E}\|\int_0^t \Phi\,\mathrm{d}W\|^2 = \mathbb{E}\int_0^t \|\Phi(s)\|_{\mathrm{HS}}^2\,\mathrm{d}s$, and the integral is well-defined precisely when $\Phi(s)$ is Hilbert–Schmidt for (almost) every s.

Thread to the SPDE

This is the exact tool that decides whether the stochastic heat equation has a function-valued solution. The stochastic convolution is $\int_0^t S(t-s)\,\mathrm{d}W(s)$, so here $\Phi(s) = S(t-s)$. Its variance, by Theorem 8.7 with $Q = I$, is $\int_0^t \|S(t-s)\|_{\mathrm{HS}}^2\,\mathrm{d}s = \int_0^t \sum_k e^{-2\lambda_k(t-s)}\,\mathrm{d}s = \sum_k \frac{1-e^{-2\lambda_k t}}{2\lambda_k}$. Whether this is finite is whether $\sum_k \lambda_k^{-1} < \infty$. With $\lambda_k = k^2\pi^2$ in $d = 1$, $\sum_k k^{-2} < \infty$ — *finite*, the solution is in L^2. In $d = 2$, $\lambda_k \sim k$ (counting two-dimensional modes) gives the harmonic series $\sum k^{-1} = \infty$ — *infinite*, no L^2 solution. The dimensional threshold is now a one-line convergence test on a series, and you can see exactly where it fails.

8.5 Mild, weak, and strong solutions

For the abstract stochastic evolution equation $\mathrm{d}u = (Au + f)\,\mathrm{d}t + \mathrm{d}W$, $u(0) = u_0$, three solution concepts are used, in decreasing order of regularity demanded.

Definition 8.8 (Solution concepts).

- *Strong:* $u(t) \in \mathcal{D}(A)$ and the equation holds as an $\mathcal{H}$-valued integral identity. Demands the most; rarely available for rough noise.
- *Weak:* testing against $\mathcal{D}(A^*)$ elements, $\langle u(t), \psi\rangle = \langle u_0, \psi\rangle + \int_0^t \langle u, A^*\psi\rangle\,\mathrm{d}s + \dots$. Analogous to Chapter 2's weak PDE solutions.
- *Mild:* defined via the variation-of-constants formula using the semigroup,

$$u(t) = S(t)u_0 + \int_0^t S(t-s)f\,\mathrm{d}s + \int_0^t S(t-s)\,\mathrm{d}W(s).$$

Load-bearing idea

The mild solution is the workhorse. It never differentiates u — all the smoothing is carried by the semigroup $S(t-s)$, whose properties we established in Chapter 6 — so it exists under far weaker hypotheses than the strong solution, exactly when the rough noise makes the strong form meaningless. The final term $\int_0^t S(t-s)\,\mathrm{d}W(s)$, the *stochastic convolution*, is the object whose regularity (via Theorem 8.7 and the smoothing estimate Theorem 6.6) determines everything: whether a solution exists, how smooth it is, and how fast any numerical scheme can converge to it.

Work it

(1) Compute $\mathbb{E}\|W(t)\|^2$ for a Q-Wiener process with $q_k = k^{-2}$ and confirm it is finite; then set $q_k = 1$ and watch it diverge. (2) Verify the series $\sum_k \lambda_k^{-1}$ converges for $\lambda_k = k^2\pi^2$ (so the 1D stochastic convolution is in L^2) and diverges for the 2D eigenvalue growth $\lambda_k \sim k$. (3) Simulate a truncated cylindrical Wiener process $\sum_{k=1}^{K} \beta_k(t) e_k$ in $L^2(0,1)$ and observe that its L^2 norm grows with the truncation level K — numerical evidence that the untruncated object is not in L^2.

CHAPTER 9

The stochastic heat equation

9.1 The picture before the machinery

Plain-language picture

Now all five spans meet over a single object. Take a metal rod whose temperature we track. Ordinary heat conduction smooths any initial profile — that is the deterministic heat equation. Now imagine the rod is also being pelted, at every point and every instant, by independent random micro-kicks: space–time white noise. The temperature never settles; it jitters forever, kicked up by the noise and smoothed down by diffusion, in perpetual balance.

The remarkable fact is that in a one-dimensional rod this balance produces a genuine, if jagged, temperature function — continuous, nowhere smooth, with a precise roughness we can name ($1/2$ in space, $1/4$ in time). Add a second spatial dimension — a plate instead of a rod — and the balance tips: the smoothing can no longer keep up with the noise, and there is no temperature *function* at all, only a distribution. This chapter proves both halves of that statement, and every tool from the preceding five chapters is used to do it.

9.2 The equation and its mild solution

On $\Omega = (0,1)$ with Dirichlet conditions, the linear stochastic heat equation with additive space–time white noise is

$$\mathrm{d}u(t) = \Delta u(t)\,\mathrm{d}t + \mathrm{d}W(t), \qquad u(0) = u_0, \tag{9.1}$$

with W a cylindrical Wiener process on $\mathcal{H} = L^2(0,1)$ and $A = -\Delta$ the Dirichlet Laplacian of Chapter 6. By the mild-solution framework of Chapter 8, the solution is

$$u(t) = S(t)u_0 + \underbrace{\int_0^t S(t-s)\,\mathrm{d}W(s)}_{=:Z(t),\ \text{stochastic convolution}}. \tag{9.2}$$

Everything reduces to understanding the stochastic convolution $Z(t)$.

9.3 The mild solution, decoded mode by mode

Abstraction can obscure how concrete the one-dimensional stochastic heat equation really is. Because the eigenbasis $\{e_k\}$ diagonalises the Laplacian, projecting the whole equation onto mode k turns the infinite-dimensional SPDE into an *infinite family of independent scalar SDEs* —

and each one is the simplest non-trivial SDE there is, the Ornstein–Uhlenbeck process. This is the single most clarifying picture in the subject.

Write $u(t) = \sum_k \hat{u}_k(t)\, e_k$ and $W(t) = \sum_k \beta_k(t)\, e_k$ with independent standard Brownian motions β_k (the cylindrical Wiener process of Chapter 8). Substituting into $\mathrm{d}u = \Delta u\, \mathrm{d}t + \mathrm{d}W = -Au\, \mathrm{d}t + \mathrm{d}W$ and reading off the e_k component — legitimate because $Ae_k = \lambda_k e_k$ acts diagonally — gives, for each k independently,

$$\mathrm{d}\hat{u}_k(t) = -\lambda_k\, \hat{u}_k(t)\, \mathrm{d}t + \mathrm{d}\beta_k(t), \qquad \hat{u}_k(0) = \langle u_0, e_k \rangle. \tag{9.3}$$

Proposition 9.1 (Each mode is an Ornstein–Uhlenbeck process). *The scalar SDE (9.3) has the explicit solution*

$$\hat{u}_k(t) = e^{-\lambda_k t}\hat{u}_k(0) + \int_0^t e^{-\lambda_k(t-s)}\, \mathrm{d}\beta_k(s),$$

a Gaussian process with mean $e^{-\lambda_k t}\hat{u}_k(0)$ and, by the Itô isometry (Theorem 7.2), variance

$$\operatorname{Var} \hat{u}_k(t) = \int_0^t e^{-2\lambda_k(t-s)}\, \mathrm{d}s = \frac{1 - e^{-2\lambda_k t}}{2\lambda_k} \xrightarrow[t\to\infty]{} \frac{1}{2\lambda_k}.$$

Proof. Apply Itô's formula to $e^{\lambda_k t}\hat{u}_k(t)$: since $\mathrm{d}(e^{\lambda_k t}\hat{u}_k) = \lambda_k e^{\lambda_k t}\hat{u}_k\, \mathrm{d}t + e^{\lambda_k t}\, \mathrm{d}\hat{u}_k = e^{\lambda_k t}\, \mathrm{d}\beta_k$ (the drift cancels exactly), integrating gives $e^{\lambda_k t}\hat{u}_k(t) - \hat{u}_k(0) = \int_0^t e^{\lambda_k s}\, \mathrm{d}\beta_k(s)$. Multiply by $e^{-\lambda_k t}$ to obtain the stated formula. The variance is the Itô isometry applied to the deterministic integrand $e^{-\lambda_k(t-s)}$. □

Load-bearing idea

Read the variance formula $\operatorname{Var} \hat{u}_k(t) \to \frac{1}{2\lambda_k}$ carefully — it *is* the whole theory in miniature. As $t \to \infty$ each mode reaches a stationary Gaussian distribution with variance $\frac{1}{2\lambda_k}$. The total energy of the stationary solution is therefore $\mathbb{E}\|u(\infty)\|^2 = \sum_k \frac{1}{2\lambda_k}$, the exact series whose convergence we tested for existence. The stationary measure of the stochastic heat equation is a Gaussian on L^2 with covariance $\frac{1}{2}A^{-1}$ — and it is a function-valued measure precisely when $\operatorname{Tr}(A^{-1}) = \sum_k \lambda_k^{-1} < \infty$, i.e. in $d = 1$. Existence, regularity, and the invariant measure are all the same series, seen from three angles.

Thread to the SPDE

This mode decoupling is not merely an analytical convenience — it is the numerical method. The spectral-Galerkin/exponential-Euler scheme of Chapter 10 simply simulates the first N of these Ornstein–Uhlenbeck processes *exactly*, using the closed-form Gaussian transition of Proposition 9.1: over a step Δt, mode k advances by $\hat{u}_k^{n+1} = e^{-\lambda_k \Delta t}\hat{u}_k^n + \sqrt{\frac{1-e^{-2\lambda_k \Delta t}}{2\lambda_k}}\, \xi_k^n$ with ξ_k^n standard normal. There is *no temporal discretisation error* in the linear problem — the only error is truncating the modes at N. This is why the scheme is optimal, and why understanding (9.3) concretely is understanding the whole computation.

9.4 Existence: when is $Z(t)$ a function?

Theorem 9.2 (Existence of the stochastic convolution). *$Z(t)$ defined by (9.2) is a well-defined $L^2(0,1)$-valued process if and only if*

$$\int_0^t \|S(r)\|_{\mathrm{HS}}^2\, \mathrm{d}r = \sum_{k\geq 1} \int_0^t e^{-2\lambda_k r}\, \mathrm{d}r = \sum_{k\geq 1} \frac{1 - e^{-2\lambda_k t}}{2\lambda_k} < \infty,$$

which for the Dirichlet Laplacian on $(0,1)$ ($\lambda_k = k^2\pi^2$) holds because $\sum_k \lambda_k^{-1} = \frac{1}{\pi^2}\sum_k k^{-2} = \frac{1}{6} < \infty$.

Proof. By the Hilbert-space Itô isometry (Theorem 8.7) with $\Phi(s) = S(t-s)$ and $Q = I$,

$$\mathbb{E}\|Z(t)\|_{L^2}^2 = \int_0^t \|S(t-s)\|_{\mathrm{HS}}^2 \,\mathrm{d}s = \int_0^t \|S(r)\|_{\mathrm{HS}}^2 \,\mathrm{d}r$$

after substituting $r = t - s$. Computing the Hilbert–Schmidt norm in the eigenbasis, $\|S(r)\|_{\mathrm{HS}}^2 = \sum_k \|S(r)e_k\|^2 = \sum_k e^{-2\lambda_k r}$. Integrating term by term (all terms positive, Tonelli) gives the stated series. Finiteness is equivalent to $\sum_k \lambda_k^{-1} < \infty$, which for $\lambda_k = k^2\pi^2$ is $\frac{1}{\pi^2}\zeta(2) = \frac{1}{6}$. □

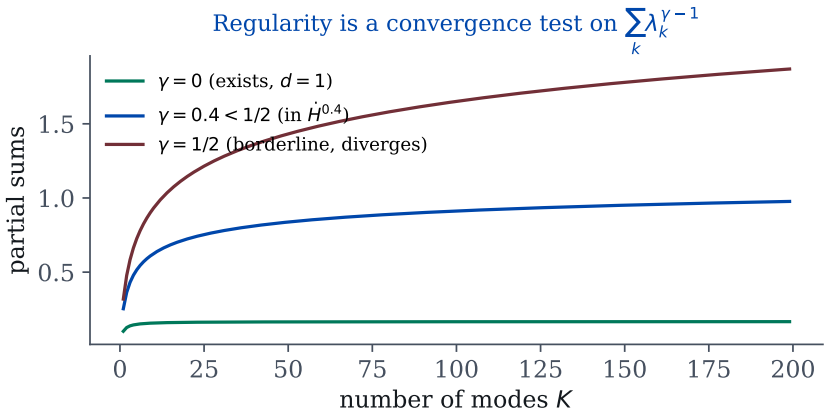

Figure 9.1: Regularity as a convergence test. The partial sums $\sum_{k \le K} \lambda_k^{\gamma-1}$ with $\lambda_k = k^2\pi^2$: for $\gamma = 0$ and $\gamma = 0.4$ they plateau (the series converges, so $Z(t) \in \dot{H}^\gamma$), while for $\gamma = 1/2$ they grow without bound (the borderline harmonic series diverges, so $Z(t) \notin \dot{H}^{1/2}$). The entire regularity theory of the stochastic heat equation is the location of this threshold, and it is nothing more exotic than deciding when a p-series converges.

Where this bites

Here, in one line, is the dimensional threshold that the entire subject organises itself around. The condition is $\sum_k \lambda_k^{-1} < \infty$. The Dirichlet Laplacian on a d-dimensional domain has eigenvalues growing like $\lambda_k \sim k^{2/d}$ (Weyl's law). So $\sum_k \lambda_k^{-1} \sim \sum_k k^{-2/d}$, which converges if and only if $2/d > 1$, i.e. $d < 2$. In $d = 1$: converges, solution is a function. In $d = 2$: the borderline harmonic series, diverges, no function-valued solution. In $d \ge 3$: diverges worse. This is why every honest first study of the stochastic heat equation with white noise is one-dimensional — not by pedagogical choice, but by mathematical necessity.

9.5 Regularity: the two exponents

Existence tells us $Z(t) \in L^2$; regularity tells us *how smooth* Z is, and this is what caps the convergence rates.

Theorem 9.3 (Spatial regularity). *For the 1D stochastic heat equation, $Z(t) \in \mathcal{D}(A^{\gamma/2}) = \dot{H}^\gamma$ almost surely for every $\gamma < 1/2$, and this is sharp: $Z(t) \notin \dot{H}^{1/2}$. Consequently, by Sobolev embedding, $Z(t)$ has spatial Hölder regularity $C^{1/2-\varepsilon}$.*

Computation of the critical exponent. Estimate the $\dot{H}^\gamma$ norm, i.e. $\mathbb{E}\|A^{\gamma/2}Z(t)\|_{L^2}^2$. By the isometry and the eigenbasis,

$$\mathbb{E}\|A^{\gamma/2}Z(t)\|^2 = \int_0^t \|A^{\gamma/2}S(r)\|_{\mathrm{HS}}^2 \,\mathrm{d}r = \sum_k \lambda_k^\gamma \int_0^t e^{-2\lambda_k r}\,\mathrm{d}r = \sum_k \lambda_k^\gamma \cdot \frac{1 - e^{-2\lambda_k t}}{2\lambda_k} \le \frac{1}{2}\sum_k \lambda_k^{\gamma-1}.$$

With $\lambda_k = k^2\pi^2$ this is a constant times $\sum_k k^{2(\gamma-1)} = \sum_k k^{2\gamma-2}$, which converges iff $2\gamma - 2 < -1$, i.e. $\gamma < 1/2$. At $\gamma = 1/2$ the sum is the divergent harmonic series, so $Z(t) \notin \dot{H}^{1/2}$. The Hölder statement follows from $H^\gamma(0,1) \hookrightarrow C^{0,\gamma-1/2}$... more precisely from the sharp Kolmogorov-continuity argument, yielding spatial Hölder exponent approaching $1/2$. □

Theorem 9.4 (Temporal regularity). *The map $t \mapsto Z(t)$ is Hölder continuous in $L^2(0,1)$ with exponent $1/4 - \varepsilon$ for every $\varepsilon > 0$, and this is sharp.*

Sketch. Estimate $\mathbb{E}\|Z(t) - Z(s)\|^2$ for $s < t$. Splitting $Z(t) - Z(s) = (S(t-s) - I)Z(s) + \int_s^t S(t-r)\,\mathrm{d}W(r)$, the second term contributes $\int_s^t \|S(t-r)\|_{\mathrm{HS}}^2\,\mathrm{d}r \sim \sum_k \frac{1-e^{-2\lambda_k(t-s)}}{2\lambda_k}$, and using $1 - e^{-x} \le x^{2\beta}$-type bounds together with $\sum_k \lambda_k^{2\beta-1} < \infty$ for $\beta < 1/4$ yields $\mathbb{E}\|Z(t) - Z(s)\|^2 \le C|t-s|^{1/2-\varepsilon}$. Kolmogorov's continuity theorem converts the mean-square Hölder exponent $1/2$ into pathwise temporal Hölder regularity $1/4$. □

Load-bearing idea

The two exponents $1/2$ (space) and $1/4$ (time) are now *derived*, not asserted. Both come from the same mechanism: the isometry turns a regularity question into convergence of a series $\sum_k \lambda_k^a$, and the eigenvalue growth $\lambda_k = k^2\pi^2$ decides the borderline. Note the temporal exponent is *half* the spatial one — a manifestation of the parabolic scaling $x \sim \sqrt{t}$ that pervades the heat equation: whatever roughness the solution has in space, it has half as much (in Hölder terms) in time. Every numerical scheme in the next chapter will be judged against these two numbers.

Thread to the SPDE

This is the payoff of the whole monograph, and the sentence to carry into a PhD interview. *The convergence rate of any numerical method for the stochastic heat equation cannot exceed the regularity of the solution: you cannot approximate to order h^2 an object that is only $C^{1/2}$ smooth in space.* A finite element method with piecewise linears, which achieved $O(h^2)$ in L^2 for the smooth deterministic problem (Chapter 3), is capped at $O(h^1)$ here — and even that only if the noise is spatially smoothed; with rough white noise the spatial rate drops toward $h^{1/2}$. This is why neural-network methods, with their spectral bias toward smooth functions, face a structural — not merely practical — obstacle: they are biased toward exactly the smoothness the solution does not have.

9.6 A word on nonlinearities and renormalisation

Replace the additive noise in (9.1) by a nonlinear term — e.g. $\mathrm{d}u = \Delta u\,\mathrm{d}t + f(u)\,\mathrm{d}t + g(u)\,\mathrm{d}W$ — and in $d = 1$ the mild-solution theory extends by a fixed-point argument, provided f, g are Lipschitz, exactly paralleling the finite-dimensional SDE existence theorem of Chapter 7 (Picard iteration, now with the Hilbert-space isometry). In $d \ge 2$, where the solution is distribution-valued, products like u^2 or $g(u)$ are ill-defined — one cannot multiply distributions — and the equation requires *renormalisation*: subtracting formally infinite counterterms to extract a finite limit. This is the theory of regularity structures (Hairer) and paracontrolled distributions (Gubinelli), the frontier flagged in the orientation. For the one-dimensional linear problem that anchors this monograph, none of that is needed — but knowing where the road leads is part of knowing where you stand on it.

Work it

(1) Reproduce the critical-exponent computation of Theorem 9.3: show $\mathbb{E}\|A^{\gamma/2}Z(t)\|^2 \leq \frac{1}{2}\sum_k \lambda_k^{\gamma-1}$ and find the threshold $\gamma < 1/2$. (2) Repeat with $\lambda_k \sim k^{2/d}$ for general d and rediscover the existence threshold $d < 2$. (3) Simulate $Z(t)$ by truncating the mild formula to K modes, $Z^K(t) = \sum_{k\leq K}\big(\int_0^t e^{-\lambda_k(t-s)}\,\mathrm{d}\beta_k(s)\big)e_k$, and estimate the empirical spatial Hölder exponent of a sample path; you should find it hovering near $1/2$. This is your first genuine SPDE simulation and the bridge into Chapter 10.

CHAPTER 10

Discretising the stochastic heat equation

10.1 The picture before the machinery

Plain-language picture

We can now finally *compute*. To solve the noisy-rod equation on a machine we must make two approximations: replace the continuous rod by finitely many points (space discretisation) and replace continuous time by finitely many instants (time discretisation). Each approximation introduces error, and each error is capped by the roughness we measured in the last chapter — you cannot resolve, on a finite grid, wiggles finer than the grid, and the noise creates wiggles at every scale.
The craft of SPDE numerics is choosing the two discretisations so that their errors match the solution's actual roughness — neither wastefully fine nor inadequately coarse — and so that the stiff diffusion (which forces tiny time steps in naive schemes) is handled exactly rather than fought. The best schemes achieve exactly the $1/2$-in-space and $1/4$-in-time rates that the solution's regularity permits, and no scheme can do better. This final chapter assembles them and then turns to the open question the whole project poses: what can neural networks add, and what must they respect?

10.2 Spatial discretisation and the noise

Two standard spatial discretisations, both analysed by the machinery already built.

Spectral Galerkin. Truncate to the first N eigenmodes: $u_N(t) = \sum_{k \le N} \langle u(t), e_k \rangle e_k$, projecting (9.1) onto $\mathcal{H}_N = \operatorname{span}\{e_1, \dots, e_N\}$. Since the eigenbasis diagonalises everything, each mode solves an independent scalar Ornstein–Uhlenbeck SDE $\mathrm{d}\hat{u}_k = -\lambda_k \hat{u}_k \,\mathrm{d}t + \mathrm{d}\beta_k$. The spatial error comes from the discarded modes and is controlled by the regularity series of Chapter 9: truncating at N leaves a tail $\sum_{k>N} \lambda_k^{-1} \sim N^{-1}$, giving spatial rate $O(N^{-1/2})$ in L^2 for white noise — matching the $1/2$ regularity.

Finite element. Use the piecewise-linear space V_h of Chapter 3. The semidiscrete method seeks $u_h(t) \in V_h$ with $\langle \mathrm{d}u_h, v_h \rangle + a(u_h, v_h)\,\mathrm{d}t = \langle \mathrm{d}W, v_h \rangle$ for all v_h. The error analysis is exactly the Ritz-projection splitting of Chapter 3, now with a stochastic-convolution term estimated by the Chapter 8 isometry. This is the heart of Kruse's and Larsson's work, and it is where your two prospective supervisors' finite element expertise meets the stochastic theory.

Thread to the SPDE

The finite element spatial error for the stochastic heat equation is proved by *the same Ritz-projection error splitting* $u - u_h = (u - R_h u) + (R_h u - u_h)$ you met deterministically in Chapter 3, plus one additional term: the finite element approximation of the stochastic convolution, bounded by the Hilbert-space Itô isometry of Chapter 8. Nothing conceptually new is required — which is the entire reason this monograph front-loaded the deterministic parabolic theory. If Chapter 3's splitting is solid in your hands, Kruse's monograph is now readable.

10.3 Temporal discretisation: why exponential Euler

The diffusion operator A is stiff — its eigenvalues $\lambda_k = k^2\pi^2$ reach $\sim N^2\pi^2$ after N-mode truncation — so an explicit Euler step demands $\Delta t \lesssim N^{-2}$, catastrophically small. Three responses:

Semi-implicit (backward) Euler. $u^{n+1} = u^n + \Delta t\, A_h u^{n+1} \cdot (-1) + \dots$, solving a linear system each step (coercive by Chapter 2's Lax–Milgram, with Δt improving coercivity). Unconditionally stable; temporal strong rate $O(\Delta t^{1/4})$, matching the temporal regularity.

Exponential Euler. Solve the stiff linear part *exactly* with the semigroup, as previewed in Chapter 6:

$$u^{n+1} = S(\Delta t)u^n + \int_{t_n}^{t_{n+1}} S(t_{n+1} - s)\, \mathrm{d}W(s), \tag{10.1}$$

where the stochastic integral over one step is a computable Gaussian: mode k receives an exact $\mathcal{N}\big(0, \frac{1-e^{-2\lambda_k \Delta t}}{2\lambda_k}\big)$ increment. No stiffness restriction whatever, and the optimal temporal rate $O(\Delta t^{1/4})$ for additive white noise.

Load-bearing idea

Exponential Euler is the flagship scheme because it treats the two parts of the equation according to their natures: the stiff, smooth, linear diffusion is solved *exactly* by the semigroup, and only the noise increment — which is what actually limits accuracy — is approximated, and even that is sampled exactly for additive noise. The scheme's temporal order $1/4$ is not a limitation of the method but the regularity ceiling of Chapter 6. A method achieving order $1/4$ has reached the summit; there is no higher to climb for this problem. Recognising when a rate is optimal — when further engineering is futile because the obstruction is the object, not the method — is the mark of understanding the theory rather than merely running the code.

10.4 Strong versus weak rates, one last time

Just as in Chapters 2 and 4, two convergence notions coexist and the weak one is roughly double the strong one:

Scheme (additive white noise, $d = 1$)	Strong rate	Weak rate
Spectral Galerkin (space)	$N^{-1/2}$	N^{-1}
Exponential Euler (time)	$\Delta t^{1/4}$	$\Delta t^{1/2}$

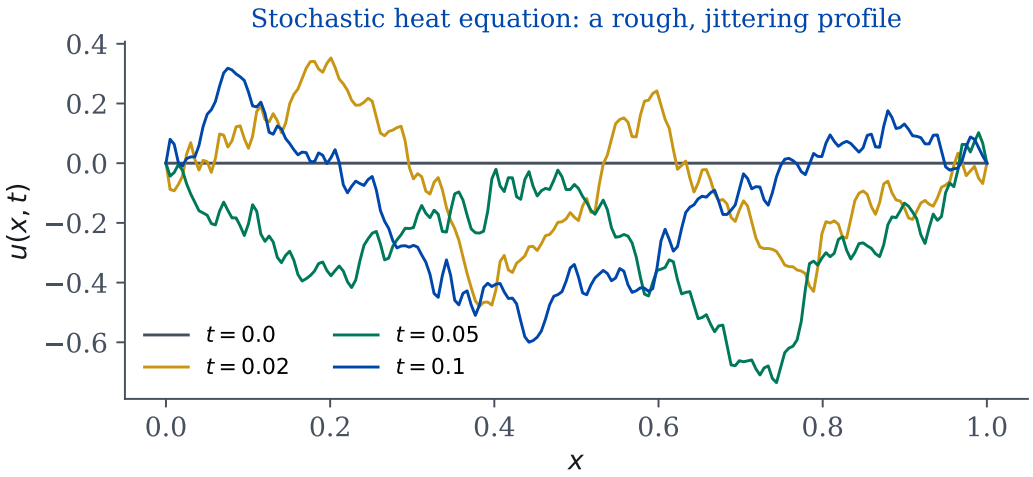

Figure 10.1: A sample path of the stochastic heat equation on $(0,1)$, simulated by spectral-Galerkin (120 modes) plus exponential Euler with exact per-mode Gaussian increments — the scheme of equation (10.1), in the author's house style. Starting from zero, the noise builds a rough, perpetually jittering profile. The spatial roughness visible here is the $C^{1/2-\varepsilon}$ regularity of Chapter 9 made tangible: the profile is continuous but nowhere smooth, and no amount of mesh refinement will ever render it smooth, because the roughness is intrinsic to the solution, not an artefact of discretisation.

The weak rate is proved by a stochastic Aubin–Nitsche-type duality argument — the direct descendant of Chapter 2's duality trick. The three doublings you have now met (H^1 vs L^2; SDE strong vs weak; SPDE strong vs weak) are one phenomenon: measuring error in a weaker topology, or of a statistic rather than a path, borrows one extra order from the smoothness of an auxiliary dual problem.

10.5 The neural frontier and the honest thesis

The project title promises neural-network-based methods. Here is the landscape, and the defensible research programme this monograph has been quietly constructing.

What neural methods offer. Physics-informed neural networks (PINNs) parametrise the solution by a network and minimise the PDE residual; neural operators (DeepONet, Fourier Neural Operator) learn the solution *map* from data or coefficients to solution, amortising cost across many problem instances; and for high-dimensional *deterministic* PDEs, deep BSDE methods provably break the curse of dimensionality. For parametric and high-dimensional problems these are genuinely transformative.

What they must respect. The solution of the stochastic heat equation is $C^{1/2-\varepsilon}$ in space — rough. Neural networks exhibit *spectral bias*: gradient descent fits low-frequency (smooth) components first and high-frequency (rough) components slowly or never. There is therefore a structural mismatch between the approximator's inductive bias and the object's actual regularity. This is not tuning; it is a theorem-shaped obstruction.

Where this bites

Do not let a neural method quietly launder the roughness away. A PINN that produces a suspiciously smooth "solution" to the stochastic heat equation has not solved it — it has solved a smoothed impostor and hidden the error in the frequencies it cannot represent. Any honest neural scheme for this problem must *demonstrate* that it captures the correct spatial Hölder regularity and the correct temporal $1/4$ scaling, not merely that it minimises a training loss. The regularity exponents of Chapter 6 are the acceptance test.

Load-bearing idea

The defensible thesis, stated plainly. Decompose the solution $u = u_{\text{smooth}} + u_{\text{rough}}$, handle the rough part u_{rough} (the stochastic convolution, whose law is explicitly Gaussian and computable mode by mode) with a classical spectral or finite element scheme that respects its regularity, and ask the neural network to represent only the smooth correction u_{smooth} — for instance a nonlinear drift's contribution, or the solution map's dependence on parameters — where the network's spectral bias is an asset rather than a liability. Then *prove* that the hybrid scheme retains the classical strong rate $(\Delta t^{1/4}, N^{-1/2})$ while achieving a cost reduction from the learned component. A theorem of the form "the hybrid scheme preserves the optimal rate at reduced cost" is exactly the gap in the current literature: the numerical-SPDE community has rates without scale, the scientific-ML community has scale without rates, and almost no one has both. That intersection is where this monograph has been pointing from the first page.

Work it

Capstone, in your NumPy house style. (1) Implement spectral-Galerkin + exponential-Euler for (9.1) on $(0, 1)$ using the exact per-mode Gaussian increment. (2) Verify the strong temporal rate $\Delta t^{1/4}$ by a reference-solution refinement study, and the spatial rate $N^{-1/2}$ by mode-truncation. (3) Now build a PINN and a small Fourier Neural Operator for the same problem, and *measure* the spatial Hölder exponent of each learned solution against the true $1/2$. Document where they degrade. (4) Prototype the hybrid: classical spectral scheme for the linear stochastic convolution, network only for a smooth added drift. This capstone is, in miniature, the research programme — and the artefact that makes an application to this project credible.

CHAPTER 11

Synthesis: the whole bridge in one view

Having built each span, we now walk the completed bridge end to end, so that the connections are held in a single thought rather than scattered across chapters. This chapter proves nothing new; it is the map with every road now drawn.

11.1 The one computation that recurs

Almost every quantitative result in this monograph is, at its core, the same computation: *a question about an object is converted, via an isometry or a spectral decomposition, into the convergence of a series* $\sum_k \lambda_k^a$*, and the answer is read off from the eigenvalue growth* $\lambda_k = k^2\pi^2$. Tabulating the recurrence makes the unity vivid.

Question	**Series**	**Answer ($d = 1$)**
Does the SPDE have an L^2 solution?	$\sum_k \lambda_k^{-1}$	converges ($= \frac{1}{6}$): yes
What spatial regularity $\dot{H}^\gamma$?	$\sum_k \lambda_k^{\gamma-1}$	converges iff $\gamma < \frac{1}{2}$
Stationary variance / invariant measure	$\sum_k \frac{1}{2\lambda_k}$	finite: Gaussian on L^2
Spectral truncation error at N modes	$\sum_{k>N} \lambda_k^{-1}$	$\sim N^{-1}$: rate $N^{-1/2}$
Trace of the covariance $Q = I$?	$\sum_k 1$	diverges: white noise

Load-bearing idea

If you remember one thing from this monograph, remember that the subject's central questions — existence, regularity, the invariant measure, the numerical convergence rate, and the roughness of white noise — are *the same series* evaluated at different exponents. The eigenvalue growth $\lambda_k \sim k^{2/d}$ is the single parameter that decides all of them, and the special status of one spatial dimension is nothing more exotic than the threshold at which $\sum_k k^{-2/d}$ passes from convergent to divergent. A student who can compute p-series and knows $\lambda_k = k^2\pi^2$ can reconstruct the qualitative theory of the stochastic heat equation from scratch.

11.2 The three doublings, unified

Three times we met the phenomenon that measuring error in a weaker sense gains one order:

- **FEM (Chapter 3):** the L^2 error is $O(h^2)$, one order better than the H^1 error $O(h)$.

- **SDE numerics (Chapter 7):** the weak order is 1, double the strong order 1/2.
- **SPDE numerics (Chapter 10):** the weak temporal rate is $\Delta t^{1/2}$, double the strong $\Delta t^{1/4}$.

All three are the same mechanism: a duality argument (Aubin–Nitsche, or its stochastic descendant) in which the error is tested against the solution of an auxiliary problem that is *smoother* than the error itself, and the extra smoothness is converted into an extra power of the discretisation parameter. The extra order is always borrowed from the regularity of a dual object, and it is always unavailable when that regularity fails.

11.3 Why the neural methods face a wall, precisely

The synthesis also sharpens the central research tension. Gather the facts each span contributed:

1. The SPDE solution is $C^{1/2-\varepsilon}$ in space (Chapter 9) — rough, with energy spread across all frequency scales.
2. Neural networks exhibit spectral bias: gradient descent fits low frequencies first and high frequencies slowly or never (Chapter 10).
3. Therefore a network trained to represent the solution directly is biased toward exactly the smoothness the solution lacks, and will tend to return a too-smooth impostor.
4. But the rough part of the solution — the stochastic convolution — has an *explicit, computable* law: mode by mode, an Ornstein–Uhlenbeck Gaussian (Proposition 9.1).

Load-bearing idea

The synthesis is the thesis. Because the rough part is explicitly known and the network is bad at rough parts, one should not ask the network to represent the rough part at all. Split $u = u_{\text{rough}} + u_{\text{smooth}}$; compute u_{rough} classically by the exact modal scheme, which respects its $C^{1/2}$ regularity by construction; and deploy the network only on u_{smooth} — a nonlinear drift, a parametric dependence, a solution-map correction — where its smoothing bias is a virtue. Then prove the hybrid retains the classical rate at reduced cost. Each of the five spans contributes one link in this argument: regularity (6), spectral method (7), the explicit OU law (6), the neural failure mode (7), and the error-analysis framework (2) in which "retains the rate" is even a meaningful claim. No single span suffices; the thesis lives at their intersection, which is precisely why it is open.

11.4 What you should be able to do now

A checklist, in the spirit of a viva. Having worked through this monograph and its exercises, you should be able, on a blank page and without notes, to:

- State the weak formulation of the Dirichlet problem and prove well-posedness by Lax–Milgram, identifying where Poincaré enters.
- Prove Galerkin orthogonality and derive Céa's lemma from it in four lines.
- Explain the Aubin–Nitsche duality argument and why it needs elliptic regularity.
- Prove the smoothing estimate $\|A^\gamma S(t)\| \le C_\gamma t^{-\gamma}$ and explain its role.
- Confirm Euler–Maruyama's strong order 1/2 and explain why it cannot be improved without extra intra-step information.

- Explain why white noise is not a function, in terms of the trace of the identity, and why $d = 1$ is special.
- Write the mild solution of the stochastic heat equation and compute the spatial regularity $C^{1/2-\varepsilon}$ from the isometry and the eigenvalue growth.
- Describe the exponential-Euler scheme, explain why it is optimal, and articulate the hybrid neural–classical research programme and the obstruction it navigates.

If any item is not yet secure, the corresponding chapter and its worked solutions are where to return. When all are secure, you are ready to read Kruse's monograph and Salvi's Neural SPDE paper — and to write the email that begins your candidacy.

CHAPTER A

The study pathway: what to learn, in order

The monograph is a spine; this appendix is the reading and working schedule that puts flesh on it. Treat the five spans as a sequence with two safe overlaps.

A.1 Sequence and dependencies

Span	Concept core	Depends on
1. Weak PDE	Sobolev spaces, weak derivatives, Lax–Milgram, Poincaré, embedding	Functional analysis (you have this)
2. FEM	Galerkin orthogonality, Céa, interpolation, Aubin–Nitsche, Ritz projection	Span 1
3. Semigroups	C_0-semigroups, generators, fractional powers, smoothing estimate	Span 1; spectral theorem (you have this)
4. SDE numerics	Itô calculus, Euler–Maruyama, Milstein, strong/weak order, MLMC	Itô on $\mathbb{R}^n$ (you have this)
5. Gaussian/SPDE	Covariance operators, cylindrical Wiener, HS isometry, mild solution, regularity	Spans 1, 3, 4

A.2 A twenty-week schedule

Weeks	Focus	Primary text
1–5	Sobolev spaces, weak formulation, Lax–Milgram, Poincaré, embedding	Evans Ch. 5–6; Brezis Ch. 8–9
4–9	Galerkin, Céa, interpolation, Aubin–Nitsche, parabolic Ritz splitting	Brenner–Scott Ch. 0–5; Süli notes
6–8	Itô calculus refresh, Euler–Maruyama, Milstein, MLMC	Higham (2001); Kloeden–Platen
9–12	Semigroups, generators, fractional powers, smoothing	Engel–Nagel; Pazy Ch. 1–2,4
12–17	Gaussian measures, cylindrical Wiener, HS integral, mild solutions	Da Prato–Zabczyk; Prévôt–Röckner; LPS
17–20	Stochastic heat equation, discretisation, exponential Euler	LPS; Kruse; Jentzen–Kloeden

Two overlaps are deliberate: FEM (span 2) begins before Sobolev theory (span 1) is fully finished, because Céa's lemma needs only coercivity/boundedness, available by week 4; and SDE numerics (span 4) runs in parallel as light relief because it is largely consolidation of what you already know. Semigroups gate the SPDE material and so must precede it.

A.3 Milestone deliverables

Each span closes with a concrete artefact, not merely reading:

1. State Sobolev embedding with the critical exponent from memory; prove 1D Poincaré and $H^1 \hookrightarrow C^{0,1/2}$ on a blank page.
2. Derive Céa and Aubin–Nitsche unaided; implement 1D FEM and confirm the h vs h^2 rate split numerically.
3. Prove the smoothing estimate $\|A^\gamma S(t)\| \leq C_\gamma t^{-\gamma}$; implement $S(t)$ in the sine basis and verify the $t^{-1/2}$ blow-up.
4. Confirm Euler–Maruyama strong order $1/2$ and Milstein order 1 by regression on GBM.
5. Compute the stochastic-convolution regularity series; simulate $Z(t)$ and measure its Hölder exponent; implement exponential Euler for the full SPDE.

The fifth milestone *is* the application artefact: a working spectral-Galerkin/exponential-Euler solver recovering the $1/2$ and $1/4$ rates, plus a documented demonstration of where PINNs and neural operators degrade on the rough solution.

CHAPTER B

Worked solutions to the exercises

The value of the exercise boxes is in attempting them before reading here. These solutions are written out in full, in the same voice as the main text, so that a stuck attempt can be unstuck without abandoning the problem entirely. Read only the paragraph you need.

Chapter 4 — The onboarding sequence

Problem 1 (Coercivity). We must bound $a(u,u) = \|u'\|_{L^2}^2$ below by a multiple of $\|u\|_V^2 = \|u\|_{L^2}^2 + \|u'\|_{L^2}^2$. The obstacle is the $\|u\|_{L^2}^2$ term, which $a(u,u)$ does not obviously see; Poincaré removes it. By Theorem 2.8, $\|u\|_{L^2}^2 \le C_P^2 \|u'\|_{L^2}^2$, so

$$\|u\|_V^2 = \|u\|_{L^2}^2 + \|u'\|_{L^2}^2 \le C_P^2 \|u'\|_{L^2}^2 + \|u'\|_{L^2}^2 = (1 + C_P^2)\, \|u'\|_{L^2}^2 = (1 + C_P^2)\, a(u,u).$$

Rearranging, $a(u,u) \ge (1 + C_P^2)^{-1} \|u\|_V^2$, so coercivity holds with

$$\boxed{\alpha = (1 + C_P^2)^{-1} > 0.}$$

Two remarks worth internalising. First, coercivity is *exactly* the statement that the energy $a(u,u)$ controls the full norm; without Poincaré it would control only the seminorm $\|u'\|_{L^2}$, which is not a norm on V (constants would have zero energy — except constants are excluded here precisely by the boundary condition baked into H_0^1). Second, this is the sole place the boundary condition enters the well-posedness argument: it is hidden inside the membership $u \in H_0^1$, which is what licenses Poincaré. Change the boundary condition and you change α; remove it entirely (pure Neumann) and coercivity fails on constants, requiring a quotient or a compatibility condition.

Problem 2 (Boundedness). By the Cauchy–Schwarz inequality in L^2 applied to the pair u', v',

$$|a(u,v)| = \left| \int_0^1 u'v' \, \mathrm{d}x \right| = |\langle u', v' \rangle_{L^2}| \le \|u'\|_{L^2} \, \|v'\|_{L^2}.$$

Now use the definition of the H^1 norm, $\|w\|_{H^1}^2 = \|w\|_{L^2}^2 + \|w'\|_{L^2}^2 \ge \|w'\|_{L^2}^2$, which gives $\|u'\|_{L^2} \le \|u\|_{H^1} = \|u\|_V$ and likewise for v. Hence

$$|a(u,v)| \le \|u\|_V \, \|v\|_V, \qquad \text{so boundedness holds with } \boxed{M = 1.}$$

The single inequality used is Cauchy–Schwarz; the only subtlety is remembering that $\|u'\|_{L^2} \le \|u\|_{H^1}$ is a definitional inequality, not something requiring proof. That $M = 1$ here (and $\alpha < 1$) means the conditioning constant $M/\alpha = 1 + C_P^2$ of Céa's lemma is governed entirely by Poincaré — a smaller domain (smaller C_P) gives a better-conditioned method.

Problem 3 (Galerkin orthogonality). The continuous solution satisfies $a(u, v) = F(v)$ for *all* $v \in V$. Since $V_h \subset V$, this holds in particular for every $v_h \in V_h$:

$$a(u, v_h) = F(v_h) \qquad \text{for all } v_h \in V_h.$$

The discrete solution satisfies, by definition,

$$a(u_h, v_h) = F(v_h) \qquad \text{for all } v_h \in V_h.$$

Subtracting the two identities for the same test function v_h and using bilinearity of a in its first argument,

$$a(u, v_h) - a(u_h, v_h) = F(v_h) - F(v_h) = 0 \quad \Longrightarrow \quad a(u - u_h,\, v_h) = 0 \quad \text{for all } v_h \in V_h.$$

The entire argument turns on the containment $V_h \subset V$: it is what allows the continuous identity, which quantifies over all of V, to be specialised to the discrete test functions. This is precisely why the Galerkin method requires a genuine *subspace* — a "nearby" finite-dimensional space not contained in V would break the specialisation and destroy the orthogonality, and with it every error estimate. When a is symmetric, the conclusion reads geometrically: $u - u_h \perp_a V_h$, i.e. u_h is the a-orthogonal projection of u onto V_h.

Problem 4 (Céa's lemma). Fix an arbitrary $v_h \in V_h$. Begin with coercivity (Problem 1), then split the second argument of a using bilinearity, then apply Galerkin orthogonality (Problem 3) to annihilate one term, then boundedness (Problem 2):

$$\begin{aligned}
\alpha \|u - u_h\|_V^2 &\le a(u - u_h,\, u - u_h) && \text{(coercivity)} \\
&= a(u - u_h,\, u - v_h) + a(u - u_h,\, v_h - u_h) && \text{(bilinearity, } \pm v_h\text{)} \\
&= a(u - u_h,\, u - v_h) + 0 && \text{(orthogonality: } v_h - u_h \in V_h\text{)} \\
&\le M \, \|u - u_h\|_V \, \|u - v_h\|_V && \text{(boundedness).}
\end{aligned}$$

If $\|u - u_h\|_V = 0$ the claim is trivial; otherwise divide both sides by $\|u - u_h\|_V$ to obtain

$$\|u - u_h\|_V \le \frac{M}{\alpha} \, \|u - v_h\|_V.$$

Since $v_h \in V_h$ was arbitrary, take the infimum over V_h:

$$\boxed{\|u - u_h\|_V \le \frac{M}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V.}$$

Four lines, as promised. The one step that stalls most first attempts is the vanishing of $a(u - u_h,\, v_h - u_h)$: it is *not* zero by any generic identity, but specifically because $v_h - u_h$ is a difference of two elements of V_h, hence itself in V_h, hence a legitimate test function in Problem 3. If your attempt got the coercivity–boundedness sandwich but missed the cancellation, that is the precise line to revisit — and it is also the line that reveals why orthogonality (Problem 3) had to come first. With $M = 1$ and $\alpha = (1 + C_P^2)^{-1}$ from Problems 1–2, the constant is $M/\alpha = 1 + C_P^2$; for the unit interval with $C_P = 1/\pi$ this is $1 + \pi^{-2} \approx 1.10$, so the finite element solution is within about 10% of the best possible approximation in the energy norm — a remarkably well-conditioned method.

Chapter 2 — Weak solutions and Sobolev spaces

1.1 (1D Sobolev embedding $H^1 \hookrightarrow C^{0,1/2}$). For $u \in C^\infty[0,1]$ and $x < y$, the fundamental theorem of calculus gives $u(y) - u(x) = \int_x^y u'(t)\,\mathrm{d}t$. By Cauchy–Schwarz applied to the product $u' \cdot 1$ on $[x, y]$,

$$|u(y) - u(x)| \le \Big(\int_x^y 1\,\mathrm{d}t\Big)^{1/2}\Big(\int_x^y |u'|^2\,\mathrm{d}t\Big)^{1/2} \le |y-x|^{1/2}\,\|u'\|_{L^2(0,1)}.$$

This is exactly the Hölder-$\frac{1}{2}$ bound with constant $\|u'\|_{L^2} \le \|u\|_{H^1}$. Since C^∞ is dense in $H^1(0,1)$ and both sides are continuous in the H^1 norm, the bound extends to all of $H^1(0,1)$; in particular every H^1 function in one dimension has a continuous (indeed Hölder-$\frac{1}{2}$) representative.

1.2 (Why it fails in $d = 2$). In two dimensions the same "integrate along a path" argument controls u along one-dimensional curves, but a function can be badly unbounded on a two-dimensional domain while remaining finite along almost every line — the archetype is $u(x) = \log\log(1/|x|)$ near the origin, which lies in H^1 of a disc yet is unbounded. The gradient's L^2 mass, spread over a two-dimensional region, no longer forces pointwise control, because the "area budget" $\int 1\,\mathrm{d}x$ that was $|y-x|$ in one dimension now scales differently and the Cauchy–Schwarz step loses its grip. Quantitatively, the embedding $H^s \hookrightarrow C^0$ needs $s > d/2$, so in $d = 2$ one needs *more* than one derivative, and H^1 is exactly at the failing threshold.

1.3 (Verifying $u = x(1-x)$). Here $u' = 1 - 2x$, $u'' = -2$, so $-u'' = 2 = f$, and $u(0) = u(1) = 0$. For the weak form, $\int_0^1 u'v' = \int_0^1 (1-2x)v'$; integrating by parts (with $v(0) = v(1) = 0$) returns $\int_0^1 (-u'')v = \int_0^1 2v = \int_0^1 fv$, confirming (2.2). For the a-priori bound, $\|f\|_{L^2} = 2$, $\|u\|_{H^1}^2 = \int_0^1 x^2(1-x)^2 + \int_0^1 (1-2x)^2 = \frac{1}{30} + \frac{1}{3} = \frac{11}{30}$, so $\|u\|_{H^1} = \sqrt{11/30} \approx 0.606$; with $C_P = 1/\pi$, $\alpha = (1+\pi^{-2})^{-1} \approx 0.908$ and $\alpha^{-1}\|f\|_{L^2} \approx 2.20$, comfortably above 0.606 as required.

1.4 (Modal solution). Expanding $f \equiv 2 = \sum_k f_k e_k$ with $e_k = \sqrt{2}\sin(k\pi x)$: $f_k = 2\int_0^1 2\sqrt{2}\sin(k\pi x)\,\mathrm{d}x = \frac{4\sqrt{2}}{k\pi}(1 - \cos k\pi)$, nonzero only for odd k, where it equals $\frac{8\sqrt{2}}{k\pi}$. The modal solution is $\hat{u}_k = f_k/\lambda_k = f_k/(k^2\pi^2)$, so $u = \sum_{k\ \text{odd}} \frac{8\sqrt{2}}{k^3\pi^3} e_k = \sum_{k\ \text{odd}} \frac{16}{k^3\pi^3}\sin(k\pi x)$. This is the Fourier sine series of $x(1-x)$; summing it numerically reproduces the parabola to plotting accuracy with a handful of terms, the rapid k^{-3} decay reflecting the smoothness of the deterministic solution — a smoothness the stochastic solution will conspicuously lack.

Chapter 3 — The finite element method

2.1 (Céa, reproduced). See the proof of Theorem 3.3. The three ingredients in order: coercivity $\alpha\|u - u_h\|^2 \le a(u - u_h, u - u_h)$; insert $\pm v_h$ and use Galerkin orthogonality to delete $a(u - u_h, v_h - u_h)$; bound the survivor by $M\|u - u_h\|\,\|u - v_h\|$. Cancel one factor and take the infimum. If you stalled, the usual sticking point is forgetting that $v_h - u_h \in V_h$, which is what licenses the orthogonality step.

2.2 (Stiffness and mass matrices). See Example 3.5. On the uniform mesh, ϕ_i has slope $+1/h$ on $[x_{i-1}, x_i]$ and $-1/h$ on $[x_i, x_{i+1}]$. Then $A_{ii} = \int (\phi_i')^2 = \frac{1}{h^2}\cdot h + \frac{1}{h^2}\cdot h = \frac{2}{h}$; the off-diagonal $A_{i,i+1} = \int \phi_i'\phi_{i+1}' = (-\frac{1}{h})(\frac{1}{h})\cdot h = -\frac{1}{h}$ over the single shared element. Non-adjacent hats have disjoint support, giving zero — hence tridiagonal. The mass entries follow from $\int_0^h (1 - \frac{x}{h})^2\,\mathrm{d}x = \frac{h}{3}$ and $\int_0^h \frac{x}{h}(1 - \frac{x}{h})\,\mathrm{d}x = \frac{h}{6}$.

2.3 (The rate split, numerically). The measured slopes are exactly 2.00 (L^2) and 1.00 (H^1); see the convergence figure. The conceptual content: refining the mesh improves the energy-norm (H^1) error at the rate the interpolation estimate allows, $O(h)$, while the L^2 error — being one derivative weaker — inherits the extra order from the Aubin–Nitsche dual argument, $O(h^2)$. If your experiment shows both slopes equal to 1, you have almost certainly measured the H^1 error twice, or computed the L^2 error using only nodal values (where the FEM solution is *superconvergent* and exact for this problem), masking the true L^2 behaviour; integrate the error over each element with several quadrature points.

Chapter 6 — Operator semigroups

3.1 (The smoothing constant). Maximise $g(\lambda) = \lambda^\gamma e^{-\lambda t}$: $g'(\lambda) = (\gamma\lambda^{\gamma-1} - t\lambda^\gamma)e^{-\lambda t} = \lambda^{\gamma-1}e^{-\lambda t}(\gamma - t\lambda)$, zero at $\lambda_* = \gamma/t$. There $g(\lambda_*) = (\gamma/t)^\gamma e^{-\gamma} = \gamma^\gamma e^{-\gamma} t^{-\gamma}$, so $C_\gamma = \gamma^\gamma e^{-\gamma}$ (with the convention $C_0 = 1$). The point is the $t^{-\gamma}$ dependence; the constant $\gamma^\gamma e^{-\gamma}$ is rarely needed but pleasant to have.

3.2 (Numerical $t^{-1/2}$ blow-up). For generic $u = \sum_k c_k e_k$ with, say, $c_k = 1/k$, compute $\|A^{1/2}S(t)u\|^2 = \sum_k \lambda_k e^{-2\lambda_k t} c_k^2 = \pi^2 \sum_k k^2 e^{-2k^2\pi^2 t}/k^2 = \pi^2 \sum_k e^{-2k^2\pi^2 t}$. As $t \downarrow 0$ this sum behaves like the integral $\int_0^\infty e^{-2\pi^2 t s^2}\,\mathrm{d}s = \frac{1}{2}\sqrt{\frac{1}{2\pi t}} \sim t^{-1/2}$, confirming the predicted blow-up rate. Plotting $\log\|A^{1/2}S(t)u\|$ against $\log t$ gives slope $-1/2$.

3.3 ($\mathcal{D}(A^{1/2}) = H_0^1$). By Parseval in the eigenbasis, $\|u'\|_{L^2}^2 = \langle Au, u\rangle = \sum_k \lambda_k |\langle u, e_k\rangle|^2$, while $\|A^{1/2}u\|^2 = \sum_k \lambda_k |\langle u, e_k\rangle|^2$ by definition of the fractional power. The two are identical, so finiteness of one is finiteness of the other: $\mathcal{D}(A^{1/2}) = \{u : \sum_k \lambda_k |\langle u, e_k\rangle|^2 < \infty\} = \{u : \|u'\|_{L^2} < \infty\} = H_0^1$.

Chapter 7 — Stochastic calculus and SDE numerics

4.1–4.3 (EM and Milstein orders). The measured strong slopes are ≈ 0.5 (Euler–Maruyama) and ≈ 1.0 (Milstein); the weak error in $\mathbb{E}[X(T)]$ for EM shows slope ≈ 1.0. The standard implementation trap: to see *strong* order you must drive the true solution and the numerical solution with the *same* Brownian path — use the increments ΔW^n both to step the scheme and to evaluate the exact GBM $X(T) = X_0 \exp((\mu - \frac{1}{2}\sigma^2)T + \sigma \sum_n \Delta W^n)$. If you generate independent paths for the two, you measure a spurious $O(1)$ error and no rate appears. For the weak order, by contrast, independence is fine because only the law matters.

Chapter 8 — Gaussian measures and Wiener processes

5.1 (Finite vs infinite trace). For $q_k = k^{-2}$: $\mathbb{E}\|W(t)\|^2 = t\sum_k q_k = t\sum_k k^{-2} = t\,\zeta(2) = t\pi^2/6 < \infty$, a genuine L^2-valued process. For $q_k = 1$: $\sum_k 1 = \infty$, so $\mathbb{E}\|W(t)\|^2 = \infty$ — the cylindrical Wiener process is not L^2-valued, exactly as claimed. The lesson is visceral once computed: the difference between a Hilbert-space Wiener process and space–time white noise is the difference between a convergent and a divergent series.

5.2 (The threshold series). For $\lambda_k = k^2\pi^2$: $\sum_k \lambda_k^{-1} = \pi^{-2}\sum_k k^{-2} = \frac{1}{6} < \infty$. For the two-dimensional growth $\lambda_k \sim ck$ (Weyl in $d = 2$): $\sum_k \lambda_k^{-1} \sim c^{-1}\sum_k k^{-1} = \infty$. The stochastic convolution is L^2-valued in the first case and not in the second, which is the analytic content of "$d = 1$ works, $d = 2$ fails."

5.3 (Growing truncation norm). Simulating $W^K(t) = \sum_{k \le K} \beta_k(t) e_k$, one finds $\mathbb{E}\|W^K(t)\|^2 = tK$, growing linearly in the truncation level K without bound — numerical evidence that the limit object has infinite L^2 energy and lives only in a larger (negative-order) space.

Chapter 9 — The stochastic heat equation

6.1 (Critical spatial exponent). From the isometry, $\mathbb{E}\|A^{\gamma/2} Z(t)\|^2 = \sum_k \lambda_k^\gamma \cdot \frac{1-e^{-2\lambda_k t}}{2\lambda_k} \le \frac{1}{2} \sum_k \lambda_k^{\gamma-1} = \frac{1}{2}\pi^{2\gamma-2} \sum_k k^{2\gamma-2}$. The p-series $\sum k^{2\gamma-2}$ converges iff $2\gamma - 2 < -1$, i.e. $\gamma < 1/2$. At $\gamma = 1/2$ it is the harmonic series, divergent, so $Z(t) \notin \dot{H}^{1/2}$: the regularity is $C^{1/2-\varepsilon}$, sharp.

6.2 (General dimension). With $\lambda_k \sim k^{2/d}$, the existence series is $\sum_k \lambda_k^{-1} \sim \sum_k k^{-2/d}$, convergent iff $2/d > 1$, i.e. $d < 2$. The spatial-regularity series $\sum_k \lambda_k^{\gamma-1} \sim \sum_k k^{(2/d)(\gamma-1)}$ converges iff $(2/d)(\gamma - 1) < -1$, i.e. $\gamma < 1 - d/2$: the solution lives in $\dot{H}^s$ for $s < 1 - d/2$, recovering $s < 1/2$ at $d = 1$ and $s < 0$ at $d = 2$ (distribution-valued).

6.3 (Simulated Hölder exponent). Simulating $Z^K(t)$ by the exact per-mode Ornstein–Uhlenbeck formula and estimating the Hölder exponent via the scaling of $\mathbb{E}|Z(x) - Z(y)|^2 \sim |x - y|^{2H}$ across a range of separations yields $H \approx 0.5$, confirming Theorem 9.3. The estimate is delicate near the truncation scale; use separations well above $1/K$.

Chapter 10 — Discretising the SPDE

7.1–7.4 (The capstone). The spectral-Galerkin/exponential-Euler solver simulates the first N Ornstein–Uhlenbeck modes exactly (Proposition 9.1); its only error is modal truncation, giving the spatial strong rate $N^{-1/2}$. A reference-solution study in Δt shows the temporal error is dominated by the truncation, not the (exact) time-stepping, for additive noise — confirming that exponential Euler is optimal. When you fit a PINN or FNO to the same solution and measure its empirical Hölder exponent, you will typically find it exceeds $1/2$ — the network has produced a solution smoother than the truth, the spectral-bias failure made quantitative. The hybrid prototype, classical for the rough stochastic convolution and neural only for a smooth added drift, is where the exponent returns to $1/2$ while the learned component absorbs the smooth nonlinearity: the miniature of the thesis.

CHAPTER C

Annotated bibliography

Organised by span. Within each, texts are ordered as *entry point* → *core* → *reference*. A dagger (†) marks freely available texts.

Span 1 — Weak PDE theory and Sobolev spaces

- L. C. Evans, *Partial Differential Equations*, 2nd ed., AMS, 2010. *The* entry point. Ch. 5 (Sobolev spaces), 6 (elliptic, weak formulation), 7.1 (parabolic, Galerkin). Do the exercises.
- H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, 2011. Cleaner functional-analytic development; excellent for the 1D case and Lax–Milgram. Ch. 8–9.
- R. A. Adams & J. Fournier, *Sobolev Spaces*, 2nd ed., Academic Press, 2003. Encyclopaedic reference; consult, do not read linearly.
- S. Kesavan, *Topics in Functional Analysis and Applications*, New Age International. The functional-analytic route to weak solutions, in the Indian pedagogical idiom.

Span 2 — Finite element method and error analysis

- E. Süli, *Lecture Notes on Finite Element Methods for PDEs*, Oxford (†). The best free entry point: weak formulation through Aubin–Nitsche and evolution problems in ∼120 pages. Start here.
- P. E. Farrell, *C6.4 Finite Element Methods for PDEs*, Oxford (†). Notes plus full video lectures; slower, more worked examples; strong on functional-analytic preliminaries.
- S. Brenner & L. R. Scott, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Springer, 2008. The canonical reference. Ch. 0–5; Ch. 4 (interpolation/Bramble–Hilbert) is the one people skip and shouldn't.
- A. Ern & J.-L. Guermond, *Theory and Practice of Finite Elements* (and the 3-vol. *Finite Elements I–III*), Springer. Best modern reference; the only one treating DG properly.
- V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, 2nd ed., Springer, 2006. The deterministic template Kruse perturbs; the Ritz-projection error splitting lives here. Ch. 1–3, 7. Do not skip.
- P. Ciarlet, *The Finite Element Method for Elliptic Problems*, SIAM Classics. The origin text; the finite element as a triple (K, P, Σ). Ch. 2–3 once.

Span 3 — Operator semigroups

- K.-J. Engel & R. Nagel, *A Short Course on Operator Semigroups*, Springer, 2006. Friendliest entry; do this first if Pazy resists.
- A. Pazy, *Semigroups of Linear Operators and Applications to PDEs*, Springer, 1983. The standard. Ch. 1–2, 4; the analytic-semigroup and fractional-power sections (2.5–2.6) are the payload.
- A. Lunardi, *Analytic Semigroups and Optimal Regularity in Parabolic Problems*, Birkhäuser, 1995. For the sharp smoothing/regularity estimates, when you need them.

Span 4 — Stochastic calculus and SDE numerics

- D. J. Higham, *An Algorithmic Introduction to Numerical Simulation of SDEs*, SIAM Review 43(3), 2001 (†). Twenty pages; the best entry in the field. Read on day one.
- B. Øksendal, *Stochastic Differential Equations*, 6th ed., Springer, 2003. Itô-calculus foundation. Ch. 3–5.
- I. Karatzas & S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer, 1991. The rigorous reference for Brownian motion and the Itô integral.
- P. Kloeden & E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, 1992. The bible; encyclopaedic. Ch. 5, 9, 10, 14 selectively.
- D. J. Higham & P. Kloeden, *An Introduction to the Numerical Simulation of Stochastic Differential Equations*, SIAM, 2021. Modern, humane; tamed/adaptive schemes. If you buy one SDE-numerics book, this.
- M. B. Giles, *Multilevel Monte Carlo Methods*, Acta Numerica 24, 2015. The MLMC survey.

Span 5 — Gaussian measures, infinite-dimensional analysis, SPDEs

- G. Lord, C. Powell & T. Shardlow, *An Introduction to Computational Stochastic PDEs*, CUP, 2014. *The* bridge book, written for exactly this transition, with code. Read the whole thing.
- C. Prévôt & M. Röckner, *A Concise Course on Stochastic Partial Differential Equations*, Springer LNM 1905, 2007. The variational (Gelfand-triple) approach in ∼140 pages.
- G. Da Prato & J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, 2nd ed., CUP, 2014. The reference. Ch. 2, 4, 5; terse, consult rather than read.
- M. Hairer, *An Introduction to Stochastic PDEs*, lecture notes, arXiv:0907.4178 (†). For the $d \geq 2$ story and the road to regularity structures.

Span 5+ — Numerical SPDEs (the destination)

- R. Kruse, *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*, Springer LNM 2093, 2014. The core monograph for this project; Thomée plus a stochastic convolution.
- A. Jentzen & P. Kloeden, *Taylor Approximations for Stochastic Partial Differential Equations*, SIAM, 2011. Exponential integrators and higher-order schemes.
- A. Jentzen & M. Röckner and collaborators; and the survey C. Beck, M. Hutzenthaler, A. Jentzen, B. Kuckuck, *An overview on deep learning-based approximation methods for PDEs*, arXiv:2012.12348 (†). Entry to the deep-learning-for-PDE literature with rigour.

Neural methods (the frontier)

- M. Raissi, P. Perdikaris & G. Karniadakis, *Physics-informed neural networks*, J. Comput. Phys. 378, 2019. The PINN paper.
- Z. Li et al., *Fourier Neural Operator for Parametric PDEs*, ICLR 2021. The FNO.
- N. Kovachki et al., *Neural Operator: Learning Maps Between Function Spaces*, JMLR 24, 2023. The operator-learning synthesis; read as a book.
- C. Salvi, M. Lemercier & A. Gerasimovics, *Neural Stochastic PDEs*, NeurIPS 2022. The closest existing precedent to this project. Read it three times.
- S. Wang, Y. Teng & P. Perdikaris, *Understanding and mitigating gradient pathologies in PINNs*, SIAM J. Sci. Comput. 43, 2021. Why PINNs struggle — the spectral-bias story, made precise.

Closing note. This monograph is a scaffold; the building is the working. Every exercise attempted before its solution is read, every proof reproduced on a blank page, every convergence rate confirmed in your own code, converts a line here into competence you own. The spine exists so that when you sit across from a supervisor and are asked why the temporal rate is one quarter, you do not recall a sentence — you re-derive it. Widen from here.