Chapter 9
Stochastic differential equations

We identify the solution to a rough differential equation driven by the Itô or Stratonovich lift of Brownian motion with the solution to the corresponding stochastic differential equation. In combination with continuity of the Itô–Lyons maps, a quick proof of the Wong–Zakai theorem is given. Applications to Stroock–Varadhan support theory and Freidlin–Wentzell large deviations are briefly discussed.

9.1 Itô and Stratonovich equations

We saw in Section 3 that d-dimensional Brownian motion lifts in an essentially canonical way to $\mathbf{B} = (B, \mathbb{B}) \in \mathscr{C}^\alpha\big([0,T], \mathbf{R}^d\big)$ almost surely, for any $\alpha \in \big(\frac{1}{3}, \frac{1}{2}\big)$. In particular, we may use almost every realisation of $(B, \mathbb{B})$ as the driving signal of a rough differential equation. This RDE is then solved "pathwise" i.e. for a *fixed* realisation of $(B(\omega), \mathbb{B}(\omega))$. Recall that the choice of $\mathbb{B}$ is never unique: two important choices are the Itô and the Stratonovich lift, we write $\mathbf{B}^{\text{Itô}}$ and $\mathbf{B}^{\text{Strat}}$, where $\mathbb{B}$ is defined as $\int B \otimes dB$ and $\int B \otimes \circ dB$ respectively. We now discuss the interplay with classical stochastic differential equations (SDEs).

Theorem 9.1. *Let $f \in \mathcal{C}_b^3\big(\mathbf{R}^e, \mathcal{L}\big(\mathbf{R}^d, \mathbf{R}^e\big)\big)$, let $f_0 : \mathbf{R}^e \to \mathbf{R}^e$ be Lipschitz continuous, and let $\xi \in \mathbf{R}^e$. Then,*

i) With probability one, $\mathbf{B}^{\text{Itô}}(\omega) \in \mathscr{C}^\alpha$, any $\alpha \in (1/3, 1/2)$ and there is a unique RDE solution $(Y(\omega), f(Y(\omega))) \in \mathscr{D}^{2\alpha}_{B(\omega)}$ to

$$dY = f_0(Y)dt + f(Y)\, d\mathbf{B}^{\text{Itô}} , \quad Y_0 = \xi .$$

Moreover, $Y = (Y_t(\omega))$ is a strong solution to the Itô SDE $dY = f_0(Y)dt + f(Y)dB$ started at $Y_0 = \xi$.

ii) Similarly, the RDE solution driven by $\mathbf{B}^{\text{Strat}}$ yields a strong solution to the Stratonovich SDE $dY = f_0(Y)dt + f(Y) \circ dB$ started at $Y_0 = \xi$.

Proof. We assume zero drift f_0, but see Exercise 8.5. The map

P. K. Friz, M. Hairer, *A Course on Rough Paths*, Universitext,
https://doi.org/10.1007/978-3-030-41556-3_9

$$B|_{[0,t]} \mapsto (B, \mathbb{B}^{\text{Strat}})|_{[0,t]} \in \mathscr{C}_g^{0,\alpha}\big([0,t], \mathbf{R}^d\big)$$

is measurable, where $\mathscr{C}_g^{0,\alpha}$ denotes the (separable, hence Polish) subspace of $\mathscr{C}^\alpha$ obtained by taking the closure, in α-Hölder rough path metric, of piecewise smooth paths. This follows, for instance, from Proposition 3.6. By the continuity of the Itô–Lyons map (adding a drift vector field is left as an easy exercise) the RDE solution $Y_t \in \mathbf{R}^e$ is the continuous image of the driving signal $(B, \mathbb{B}^{\text{Strat}})|_{[0,t]} \in \mathscr{C}_g^{0,\alpha}\big([0,t], \mathbf{R}^d\big)$. It follows that Y_t is adapted to

$$\sigma\{B_{r,s}, \mathbb{B}_{r,s} : 0 \le r \le s \le t\} = \sigma\{B_s : 0 \le s \le t\}\ ,$$

and it suffices to apply Corollary 5.2. Since $\mathbb{B}^{\text{Itô}}_{s,t} = \mathbb{B}^{\text{Strat}}_{s,t} - \frac{1}{2}(t-s)I$, measurability is also guaranteed and we conclude with the same argument, using Proposition 5.1. □

Remark 9.2. In contrast to standard SDE theory, the present solution constructed via RDEs is immediately well-defined as a *flow*, i.e. for all ξ on a common set of probability one. The price to pay is that of $\mathcal{C}^3$ regularity of f, as opposed to the mere Lipschitz regularity required for the standard theory.

9.2 The Wong–Zakai theorem

A classical result (e.g. [IW89, p.392]) asserts that SDE approximations based on piecewise linear approximations to the driving Brownian motions converge to the solution of the Stratonovich equation. Using the machinery built in the previous sections, we can now give a simple proof of this by combining Proposition 3.6, Theorem 8.5 and the understanding that RDEs driven by $\mathbf{B}^{\text{Strat}}$ yield solutions to the Stratonovich equation (Theorem 9.1).

Theorem 9.3 (Wong–Zakai, Clark, Stroock–Varadhan). *Let f, f_0, ξ be as in Theorem 9.1 above. Let $\alpha < 1/2$. Consider dyadic piecewise linear approximations (B^n) to B on $[0,T]$, as defined in Proposition 3.6. Write Y^n for the (random) ODE solutions to $dY^n = f_0(Y^n)dt + f(Y^n)dB^n$ and Y for the Stratonovich SDE solution to $dY = f_0(Y)dt + f(Y) \circ dB$, all started at ξ. Then the Wong–Zakai approximations converge a.s. to the Stratonovich solution. More precisely, with probability one,*

$$\|Y - Y^n\|_{\alpha;[0,T]} \to 0.$$

The only reason for *dyadic* piecewise linear approximations in the above statement is the formulation of the martingale-based Proposition 3.6. In Section 10 we shall present a direct analysis (going far beyond the setting of Brownian drivers) which easily entails quantitative convergence (in probability and L^q, any $q < \infty$) for all piecewise linear approximations towards a (Gaussian) rough path.

In the forthcoming Exercise 10.2 it will be seen that (non-dyadic) piecewise linear approximations of mesh size $\sim 1/n$, viewed canonically as rough paths, converge a.s.

in $\mathscr{C}^\alpha$ with rate anything less than $1/2-\alpha$. As long as $\alpha > 1/3$, it then follows from (local) Lipschitzness of the Itô–Lyons map that Wong–Zakai approximations also converge with rate $(1/2-\alpha)^-$. Note that the "best" rate one obtains in this way is $(1/2-1/3)^- = 1/6^-$; the reason being that rate is measured in some Hölder space with exponent $1/3^+$, rather than the uniform norm. The well-known almost sure "strong" rate $1/2^-$ can be obtained from rough path theory at the price of working in rough path spaces of much lower regularity, see [FR14].

9.3 Support theorem and large deviations

We briefly discuss two fundamental results in diffusion theory and explain how the theory of rough paths provides elegant proofs, reducing a question for general diffusion to one for Brownian motion and its Lévy area.

The results discussed in this section were among the very first applications of rough path theory to stochastic analysis, see Ledoux et al. [LQZ02]. Much more on these topics is found in [FV10b], so we shall be brief. The first result, due to Stroock–Varadhan [SV72] concerns the support of diffusion processes.

Theorem 9.4 (Stroock–Varadhan support theorem). *Let f, f_0, ξ be as in Theorem 9.1 above. Let $\alpha < 1/2$, B be a d-dimensional Brownian motion and consider the unique Stratonovich SDE solution Y on $[0,T]$ to*

$$dY = f_0(Y)dt + \sum_{i=1}^{d} f_i(Y) \circ dB^i \tag{9.1}$$

started at $Y_0 = \xi \in \mathbf{R}^e$. Write y^h for the ODE solution obtained by replacing $\circ dB$ with $dh \equiv \dot{h}\,dt$, whenever $h \in \mathcal{H} = W_0^{1,2}$, i.e. absolutely continuous, $h(0) = 0$ and $\dot{h} \in L^2([0,T], \mathbf{R}^d)$. Then, for every $\delta > 0$,

$$\lim_{\varepsilon \to 0} \mathbf{P}\Big(\|Y - Y^h\|_{\alpha;[0,T]} < \delta \;\Big|\; \|B - h\|_{\infty;[0,T]} < \varepsilon \Big) = 1 \tag{9.2}$$

(where Euclidean norm is used for the conditioning $\|B - h\|_{\infty,[0,T]} < \varepsilon$). As a consequence, the support of the law of Y, viewed as measure on the pathspace $\mathcal{C}^{0,\alpha}([0,T], \mathbf{R}^e)$, is precisely the α-Hölder closure of $\{y^h : \dot{h} \in L^2([0,T], \mathbf{R}^d)\}$.

Proof. Using Theorem 9.1 we can and will take Y as RDE solution driven by $\mathbf{B}^{\text{Strat}}(\omega)$. For $h \in \mathcal{H}$ and some fixed $\alpha \in (\frac{1}{3}, \frac{1}{2})$, we furthermore denote by $S^{(2)}(h) = (h, \int h \otimes dh) \in \mathscr{C}_g^{0,\alpha}$ the canonical lift given by computing the iterated integrals using usual Riemann–Stieltjes integration. It was then shown in [FLS06][1] that for every $\delta > 0$,

[1] Strictly speaking, this was shown for $h \in \mathcal{C}^2$; the extension to $h \in \mathcal{H}$ is non-trivial and found in [FV10b].

$$\lim_{\varepsilon \to 0} \mathbf{P}\Big(\varrho_{\alpha;[0,T]}\big(\mathbf{B}^{\text{Strat}}, S^{(2)}(h)\big) < \delta \;\Big|\; \|B - h\|_{\infty;[0,T]} < \varepsilon \Big) = 1. \tag{9.3}$$

The conditional statement then follows easily from continuity of the Itô–Lyons map and so yields the "difficult" support inclusion: every y^h is in the support of Y. The easy inclusion, support of Y contained in the closure of $\{y^h\}$, follows from the Wong–Zakai theorem, Theorem 9.3. If one is only interested in the support statement, but without the conditional statement (9.2), there are "softer" proofs; see Exercise 9.1 below. □

The second result to be discussed here, due to Freidlin–Wentzell, concerns the behaviour of diffusion in the singular ($\varepsilon \to 0$) limit when B is replaced by εB. We assume the reader is familar with large deviation theory.

Theorem 9.5 (Freidlin–Wentzell large deviations). *Let f, f_0, ξ be as in Theorem 9.1 above. Let $\alpha < 1/2$, B be a d-dimensional Brownian motion and consider the unique Stratonovich SDE solution $Y = Y^\varepsilon$ on $[0,T]$ to*

$$dY = f_0(Y)dt + \sum_{i=1}^{d} f_i(Y) \circ \varepsilon dB^i \tag{9.4}$$

started at $Y_0 = \xi \in \mathbf{R}^e$. Write Y^h for the ODE solution obtained by replacing $\circ\varepsilon dB$ with dh where $h \in \mathcal{H} = W_0^{1,2}$. Then $(Y_t^\varepsilon : 0 \le t \le T)$ satisfies a large deviation principle (in α-Hölder topology) with good rate function on pathspace given by

$$J(y) = \inf\big\{ I(h) : Y^h = y \big\}.$$

Here I is Schilder's rate function for Brownian motion, i.e. $I(h) = \frac{1}{2}\|\dot{h}\|^2_{L^2([0,T],\mathbf{R}^d)}$ for $h \in \mathcal{H}$ and $I(h) = +\infty$ otherwise.

Proof. The key remark is that large deviation principles are robust under continuous maps, a simple fact known as *contraction principle*. The problem is then reduced to establishing a suitable large deviation principle for the Stratonovich lift of εB (which is exacly $\delta_\varepsilon \mathbf{B}^{\text{Strat}}$) in the α-Hölder rough path topology. Readers familiar with general facts of large deviation theory, in particular the *inverse* and *generalised* contraction principles, are invited to complete the proof along Exercise 9.2 below. □

9.4 Laplace method

We have seen that $(Y_t^\varepsilon : 0 \le t \le T)$, given as continuous images of the rescaled Brownian rough path, $Y^\varepsilon = \Phi\big(\delta_\varepsilon \mathbf{B}^{\text{Strat}}\big)$, satisfies a large deviations principle (in Hölder and hence also in uniform topology) with rate function

$$J(y) = \inf\{ I(h) : \Phi(h) = y,\, h \in \mathcal{H} \} \tag{9.5}$$

with the mild abuse of notation $\Phi(h) \equiv \Phi(\mathbf{h})$ where $\mathbf{h} = \big(h, \int h \otimes dh\big)$ is the canonical lift of $h \in \mathcal{H}$. A standard fact of large deviation theory, *Varadhan's lemma*, implies the following Laplace principle: for bounded continuous $F : \mathcal{C}([0,T], \mathbf{R}^e) \to \mathbf{R}$,

$$\lim_{\varepsilon \to 0} \varepsilon^2 \log \mathbf{E}\big[\exp\big(-F(Y^\varepsilon)/\varepsilon^2\big)\big] = -\inf\{F_\Lambda(h) : h \in \mathcal{H}\}\ ,$$

where we set $F_\Lambda = F \circ \Phi + I$, for I as in Theorem 9.5. We are interested in precise asymptotics, hence the following collection of hypotheses.

(H1) The function F is bounded continuous on $\mathcal{C}([0,T], \mathbf{R}^e)$.
(H2) The function F_Λ attains its unique minimum at $\gamma \in \mathcal{H}$.
(H3) The function F is $\mathcal{C}^3$ in the Fréchet sense at $\varphi := \Phi(\gamma)$.
(H4) The element γ is a non-degenerate minimum of F_Λ restricted to $\mathcal{H}$ namely, for all $h \in \mathcal{H}\backslash\{0\}$,

$$D^2 F_\Lambda(\gamma)(h,h) = D^2(F \circ \Phi)\big|_\gamma (h,h) + \|h\|^2_{\mathcal{H}} > 0$$

Theorem 9.6. *Let Y^ε be the unique Stratonovich SDE solution on $[0,T]$ in the small noise regime from Theorem 9.5. Under conditions (H1-H4), the following precise Laplace asymptotic holds*

$$\mathbf{E}\big[\exp\big(-F(Y^\varepsilon)/\varepsilon^2\big)\big] = \exp\left(-\frac{F_\Lambda(\gamma)}{\varepsilon^2}\right)(c_0 + o(1)) \qquad \text{as } \varepsilon \downarrow 0, \tag{9.6}$$

for some constant $c_0 \in (0,\infty)$.

Proof. (i) Localisation around the minimiser. We regard $\mathbf{B} = \mathbf{B}^{\text{Strat}}$ (and its ε-dilations) as random variables in the (Polish) rough path space $\mathscr{C} := \mathscr{C}^{0,\alpha}_g([0,T])$. Write $\boldsymbol{\gamma} := (\gamma, \int \gamma \otimes d\gamma) \in \mathscr{C}$ for the canonical lift of the minimiser $\gamma \in \mathcal{H}$. Take now an arbitrary neighbourhood O of $\boldsymbol{\gamma} \in \mathscr{C}$ and decompose

$$\begin{aligned}\mathbf{E}\big[\exp\big(-F(Y^\varepsilon)/\varepsilon^2\big)\big] &= \mathbf{E}\big[\exp\big(-F \circ \Phi(\delta_\varepsilon \mathbf{B})/\varepsilon^2\big)\big] \\ &= \mathbf{E}[\dots; \{\delta_\varepsilon \mathbf{B} \in O\}] + \mathbf{E}[\dots; \{\delta_\varepsilon \mathbf{B} \in O\}^c]\ .\end{aligned}$$

Since $(\delta_\varepsilon \mathbf{B})$ satisfies an LDP with *good* rate function, (H1) implies that there exists $d > a := F_\Lambda(\gamma)$ and $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$

$$\mathbf{E}\big[\exp\big(-F \circ \Phi(\delta_\varepsilon \mathbf{B})/\varepsilon^2\big); \{\delta_\varepsilon \mathbf{B} \in O\}^c\big] \le \exp\big(-d/\varepsilon^2\big). \tag{9.7}$$

Hence this term does not contribute to the asymptotics (9.6). In the sequel, we shall take, for some $\varrho > 0$,

$$O := O_\varrho := \{T_\gamma \mathbf{X} : \mathbf{X} \in \mathscr{C}, \|\!|\mathbf{X}|\!\| < \varrho\} = \{\mathbf{X} \in \mathscr{C} : \|\!|T_{-\gamma}\mathbf{X}|\!\| < \varrho\}\ .$$

(By continuity of the translation operator, this is indeed an open neighbourhood of $T_\gamma \mathbf{0} = \boldsymbol{\gamma}$.) We are thus left to analyse

$$J_\varrho := \mathbf{E}\big[\exp\big(-F \circ \Phi(\delta_\varepsilon \mathbf{B})/\varepsilon^2\big); \|\!|T_{-\gamma}\delta_\varepsilon \mathbf{B}|\!\| < \varrho\big]\ .$$

(ii) Cameron–Martin shift. It is easy to see that, for Wiener a.e. ω, one has $\mathbf{B}(\omega+h) = T_h\mathbf{B}(\omega)$. In particular, the Cameron–Martin shift $\varepsilon B \rightsquigarrow \varepsilon B + \gamma$ (or $\omega \rightsquigarrow \omega + \gamma/\varepsilon$) induces a translation of $\delta_\varepsilon \mathbf{B}$ in the sense that

$$\delta_\varepsilon \mathbf{B} = \Big(\varepsilon B, \int \varepsilon B \otimes d(\varepsilon B)\Big) \rightsquigarrow \Big(\varepsilon B + \gamma, \int (\varepsilon B + \gamma) \otimes d(\varepsilon B + \gamma)\Big) = T_\gamma \delta_\varepsilon \mathbf{B} \,.$$

From the Cameron–Martin theorem, with all integrals below understood over $[0,T]$,

$$\begin{aligned} J_\varrho(\varepsilon) &= \mathbf{E}\Big(\exp\Big(-\frac{\|\gamma\|_{\mathcal{H}}^2}{2\varepsilon^2} - \frac{\int \dot\gamma d(\varepsilon B)}{\varepsilon^2}\Big) \exp\Big(-\frac{F\circ\Phi(T_\gamma\delta_\varepsilon \mathbf{B})}{\varepsilon^2}\Big); \|\!|\delta_\varepsilon \mathbf{B}\|\!| < \varrho\Big) \\ &= \exp\Big(-\frac{\|\gamma\|_{\mathcal{H}}^2 + F\circ\Phi(\gamma)}{2\varepsilon^2}\Big) \mathbf{E}\Big(\exp\Big(-\frac{(*)}{\varepsilon^2}\Big); \|\!|\delta_\varepsilon \mathbf{B}\|\!| < \varrho\Big); \end{aligned}$$

where we recognise $F_\Lambda(\gamma)$ in the first exponential and also set

$$(*) = F\circ\Phi(T_\gamma\delta_\varepsilon \mathbf{B}) - F\circ\Phi(\gamma) + \varepsilon\int \dot\gamma dB \,.$$

(iii) Local analysis around the minimiser. We argue on a fixed rough path realisation $\mathbf{X} := \mathbf{B}(\omega)$. One checks that $\varepsilon \mapsto \Phi(T_\gamma\delta_\varepsilon \mathbf{X})$ is sufficiently smooth so that

$$\Phi(T_\gamma\delta_\varepsilon \mathbf{X}) = \Phi(\gamma) + \varepsilon G^1(\mathbf{X}) + \tfrac{\varepsilon^2}{2} G^2(\mathbf{X}) + \varepsilon^3 R_\varepsilon(\mathbf{X})$$

with remainder $R_\varepsilon(\mathbf{X})$, uniformly bounded in $\varepsilon \in (0,1]$. We now use (H3) to obtain the expansion

$$\begin{aligned} (F\circ\Phi)(T_\gamma\delta_\varepsilon \mathbf{X}) &= (F\circ\Phi)(\gamma) + \varepsilon\, DF|_\varphi\big(G^1(\mathbf{X})\big) \\ &\quad + \frac{\varepsilon^2}{2}\underbrace{\Big[DF|_\varphi\big(G^2(\mathbf{X})\big) + D^2F|_\varphi\big(G^1(\mathbf{X}), G^1(\mathbf{X})\big)\Big]}_{=:Q(\mathbf{X})} + \varepsilon^3 R_\varepsilon^F(\mathbf{X}) \,, \end{aligned}$$

where (H3) requires us to take ε less than some $\varepsilon_1(\mathbf{X})$, with remainder $R_\varepsilon^F(\mathbf{X})$, uniformly bounded in $\varepsilon \in (0,\varepsilon_1)$. Write $G^1 = G^1(h)$, and similar for G^2, Q, when evaluated at the canonical lift of an element $h \in \mathcal{H}$. We note for later

$$Q(h) = \frac{\partial^2}{\partial\varepsilon^2}\Big|_{\varepsilon=0} (F\circ\Phi)(\gamma+\varepsilon h) \,.$$

Since γ minimises $F_\Lambda = F\circ\Phi + I$, first order optimality leads precisely to

$$DF|_\varphi\big(G^1(h)\big) + \int \dot\gamma dh = 0\,, \tag{9.8}$$

for any $h \in \mathcal{H}$. By continuous extension we have $DF|_\varphi\big(G^1(\mathbf{B}(\omega))\big) + \int \dot\gamma dB = 0$, see Exercise 9.3 (ii), and so

$$J_\varrho(\varepsilon) = \exp\left(-\frac{F_\Lambda(\gamma)}{2\varepsilon^2}\right)\mathbf{E}\big(\exp\left(-Q(\mathbf{B})/2 + \varepsilon R_\varepsilon^F(\mathbf{B})\right); \|\delta_\varepsilon \mathbf{B}\| < \varrho\big) .$$

We claim that, as one would expect from exchanging $\varepsilon \to 0$ with expectation,

$$\lim_{\varepsilon \to 0} \mathbf{E}\big[\exp\left(-Q(\mathbf{B})/2 + \varepsilon R_\varepsilon^F(\mathbf{B})\right); \|\delta_\varepsilon \mathbf{B}\| < \varrho\big] = \mathbf{E}[\exp(-Q(\mathbf{B}))/2] < \infty.$$

To see why this is so, we first show integrability and even $\exp\left[-Q(\mathbf{B})/2\right] \in L^{1+\beta}$, for some $\beta > 0$, as consequence of the non-degeneracy assumption on the minimizer. The claimed integrability follows from the tail estimate $\mathbf{P}(-Q(\mathbf{B})/2 \geq r) \leq e^{-Cr}$, with $C > 1$ and for sufficiently large r. Now Q is "quadratic" in the precise sense $Q(\delta_\lambda \mathbf{X}) = \lambda^2 Q(\mathbf{X}),\ \lambda > 0$, so that upon setting $r \equiv 1/\varepsilon^2$, we are left to show

$$\mathbf{P}(-Q(\delta_\varepsilon \mathbf{B}) \geq 2) \leq e^{-C/\varepsilon^2}.$$

Since Q is seen to be continuous on rough path space, we have a good Large Deviations Principle for $\{-Q(\delta_\varepsilon \mathbf{B}) : \varepsilon > 0\}$, and using the upper LDP bound

$$\mathbf{P}(-Q(\delta_\varepsilon \mathbf{B})/2 \geq 1) \leq e^{-(C^* + o(1))/\varepsilon^2} ,$$

it remains to see $1 < C^*$, where, using goodness of the rate function,

$$C^* = \inf\left\{\tfrac{1}{2}\|h\|_{\mathcal{H}}^2 : h \in \mathcal{H}, -Q(h)/2 \geq 1\right\} = \tfrac{1}{2}\|h^*\|_{\mathcal{H}}^2 \quad \text{for some } h^* \in \mathcal{H} .$$

But this follows exactly from "$D^2(F \circ \Phi + I)(\gamma) > 0$" in direction h^*,

$$1 \leq -Q(h^*)/2 = \frac{1}{2}\frac{\partial^2}{\partial \varepsilon^2}\Big|_{\varepsilon=0}(-F \circ \Phi)(\gamma + \varepsilon h^*) < \frac{1}{2}\|h^*\|_{\mathcal{H}}^2 .$$

This establishes $\exp\left[-Q(\mathbf{B})/2\right] \in L^{1+\beta}$. This additional amount of integrability, $\beta > 0$, is now used to give a uniform L^1-bound on $\exp\left(-Q(\mathbf{B})/2 + \varepsilon R_\varepsilon^F(\mathbf{B})\right)$ over $\|\delta_\varepsilon \mathbf{B}\| < \varrho$, after which one can conclude by dominated convergence. To this end, we revert to a pathwise consideration, $\mathbf{X} := \mathbf{B}(\omega)$. We need the remainder estimate, Exercise 9.4,

$$\sup_{\varepsilon \in (0,\varepsilon_1]} \left|R_\varepsilon^F(\mathbf{X})\right| \lesssim 1 + \|\mathbf{X}\|^3 , \tag{9.9}$$

valid whenever $\varepsilon\|\mathbf{X}\| = \|\delta_\varepsilon \mathbf{X}\|$ remains bounded. It follows that, on $\|\delta_\varepsilon \mathbf{B}\| < \varrho$, we have the (uniform in small ε) estimate

$$\varepsilon\left|R_\varepsilon^F(\mathbf{B})\right| \lesssim 1 + \varepsilon\|\mathbf{B}\|^3 \lesssim 1 + \varrho\|\mathbf{B}\|^2 \tag{9.10}$$

and this estimate is uniform over $\varepsilon \in (0, 1]$. By Fernique's estimate for the (homogeneous!) rough path norm $\|\mathbf{B}\|$ of $\mathbf{B} = \mathbf{B}(\omega)$ and by choosing $\varrho = \varrho(\beta)$ small enough, we can guarantee that

$$e^{\varepsilon R_\varepsilon^F(\mathbf{B})} 1_{\{\|\delta_\varepsilon \mathbf{B}\| < \varrho\}} \lesssim \exp\left(C\varrho\|\mathbf{B}\|^2\right) \in L^{\beta'} ,$$

where $\beta' < \infty$ is the Hölder conjugate of $\beta > 1$. Hence $\exp[-Q(\mathbf{B})/2 + \varrho \|\mathbf{B}\|^2] \in L^1$ serves as the uniform L^1-bound we were looking for and the proof is complete. □

9.5 Exercises

Exercise 9.1 (Support of Brownian rough path [FV10b]) *Fix $\alpha \in (\frac{1}{3}, \frac{1}{2})$ and view the law μ of $\mathbf{B}^{\text{Strat}}$ as probability measure on the Polish space $\mathscr{C}_{g,0}^{0,\alpha}$, the (closed) subspace of $\mathscr{C}_g^{0,\alpha}$ of rough paths $\mathbf{X}$ started at $X_0 = 0$. Show that $\mathbf{B}^{\text{Strat}}$ has full support. The "easy" inclusion, supp $\mu \subset \mathscr{C}_g^{0,\alpha}$ is clear from Proposition 3.6. For the other inclusion, recall the translation operator from Exercise 2.15 and follow the steps below.*

a) **(Cameron–Martin theorem for Brownian rough path)** *Let $h \in [0,T] \in \mathcal{H} = W_0^{1,2}$. Show that $\mathbf{X} \in \operatorname{supp} \mu$ implies $T_h(\mathbf{X}) \in \operatorname{supp} \mu$.*

b) Show that the support of μ contains at least one point, say $\hat{\mathbf{X}} \in \mathscr{C}_g^{0,\alpha}$ with the property that there exists a sequence of Lipschitz paths $(h^{(n)})$ so that $T_{h^{(n)}}(\hat{\mathbf{X}}) \to (0,0)$ in α-Hölder rough path metric.

Hint: *Almost every realisation of $\mathbf{B}^{\text{Strat}}(\omega)$ will do, with $-h^{(n)} = B^{(n)}$, the dyadic piecewise linear approximations from Proposition 3.6.*

c) Conclude that $(0,0) = \lim_{n\to\infty} T_{h^{(n)}}(\hat{\mathbf{X}}) \in \operatorname{supp} \mu$.

d) As a consequence, any $(h, \int h \otimes dh) = T_h(0,0) \in \operatorname{supp} \mu$, for any $h \in \mathcal{H}$ and taking the closure yields the "difficult" inclusion.

e) Appeal to continuity of the Itô–Lyons map to obtain the "difficult" support inclusion ("every y^h is in the support of Y") in the context of Theorem 9.4.

Exercise 9.2 ("Schilder" large deviations, see [FV10b]) *Fix $\alpha \in (\frac{1}{3}, \frac{1}{2})$ and consider*

$$\delta_\varepsilon \mathbf{B}^{\text{Strat}} = (\varepsilon B, \varepsilon^2 \mathbb{B}^{\text{Strat}}) ,$$

the laws of which are viewed as probability measures μ^ε on the Polish space $\mathscr{C}_{g,0}^{0,\alpha}$. Show that $(\mu^\varepsilon) : \varepsilon > 0$ satisfies a large deviation principle in α-Hölder rough path topology with good rate function

$$J(\mathbf{X}) = I(X) ,$$

where $\mathbf{X} = (X, \mathbb{X})$ and I is Schilder's rate function for Brownian motion, i.e. $I(h) = \frac{1}{2}\|\dot{h}\|^2_{L^2([0,T],\mathbf{R}^d)}$ for $h \in \mathcal{H} = W_0^{1,2}$ and $I(h) = +\infty$ otherwise.

Hint: *Thanks to Gaussian integrability for the homogeneous rough paths norm of $\mathbf{B}^{\text{Strat}}$ it is actually enough to establish a large deviation principle for $(\delta_\varepsilon \mathbf{B}^{\text{Strat}} : \varepsilon > 0)$ in the (much coarser) uniform topology, which is not very hard to do "by hand", cf. [FV10b].*

Exercise 9.3 *In the context of Laplace asymptotics given in Theorem 9.6:*

a) *Detail the localisation estimate* (9.7).
b) *Derive the first order optimality condition* (9.8) *and justify its "continuous extension", i.e. replacing* h *by* $B(\omega)$.
c) *Show that* $G_2 = G_2(\mathbf{X})$ *is continuous in rough path sense. Conclude that the same holds for* $Q = Q(\mathbf{X})$.

Remark: *Related results appear in [BA88] (on path space) and [Ina06, Lemma 8.2].*

Exercise 9.4 (Stochastic Taylor-like rough path expansion) *We aim to show the remainder estimate* (9.9).

a) *As a warmup, consider* $\Phi : \mathcal{C}([0,1], \mathbf{R}^d) \to \mathbf{R}$ *so that* $\Phi(X) = \varphi(X_1)$, *for some* $\varphi \in \mathcal{C}^3(\mathbf{R}^d)$. *Fix* $\gamma \in \mathcal{C}([0,1], \mathbf{R}^d)$ *and establish the expansion*

$$\Phi(\gamma + \varepsilon X) \equiv g_0 + \varepsilon g_1(X) + \varepsilon^2 g_2(X) + \varepsilon^3 r_\varepsilon(X) \, ,$$

such that $|r_\varepsilon(X)| \lesssim |X_1|^3$, *uniformly in* $\varepsilon \in (0,1]$, *provided* $|\varepsilon X_1|$ *remains bounded.*
b) *Show that an extra* ε*-dependent drift, say* εX *replaced by* $\varepsilon X + \varepsilon \mu$ *for some fixed* $\mu \in \mathcal{C}([0,1], \mathbf{R}^d)$, *alters the remainder estimate to* $|r_\varepsilon(X)| \lesssim 1 + |X_1|^3$.
c) *Generalise a) and b) to the situation when* Φ *is* $\mathcal{C}^3$*-regular in Fréchet sense. (This trivially covers the case* $F \circ \Phi$, *with another* $F \in \mathcal{C}^3$.*)*
d) *Prove the real thing, i.e. the remainder estimate* (9.9) *based on the expansion of* $\varepsilon \mapsto F \circ \Phi(T_\gamma \delta_\varepsilon \mathbf{X})$ *where* Φ *is the Itô–Lyons map. (See e.g. [IK07, Thm 5.1] and the references therein. For a similar estimate in a slightly different setting, see also [FGP18].)*

9.6 Comments

The rough path approach to solving stochastic differential equations (SDEs) driven by d-dimensional noise, can be seen as far-reaching extension of the works of Doss and Sussmann [Dos77, Sus78], and the Wong–Zakai approximation result [WZ65] ($d = 1$) and Clark [Cla66], Stroock-Varadhan [SV72] for $d > 1$. Lyons [Lyo98] used the Wong–Zakai theorem in conjunction with his continuity result to deduce the fact that RDE solutions (driven by the Brownian rough path $\mathbf{B}^{\text{Strat}}$) coincide with solution to (Stratonovich) stochastic differential equations. Similar to Friz–Victoir [FV10b], the logic is reversed in our presentation: thanks to an a priori identification of $\int f(Y)\, d\mathbf{B}^{\text{Strat}}$ as a Stratonovich stochastic integral, the Wong–Zakai results is obtained. Ikeda–Watanabe [IW89] present "twisted" Wong–Zakai approximation, based on McShane [McS72], in which case an additional limiting drift vector field appears; see also [Sus91, FO09]. Wong-Zakai type results for SPDEs (with finite-dimensional noise) is a straight-forward consequence of continuity statements for rough partial differential equations, as discussed in Sections 12.1 and 12.2. A version of the Wong–Zakai theorem for a singular SPDEs with space-time white noise via regularity structures is established by Hairer–Pardoux [HP15].

Almost sure rates for Wong–Zakai approximations in Brownian (and then more general Gaussian) rough path situations, were studied by Hu–Nualart [HN09], Deya–Neuenkirch–Tindel [DNT12] and Friz–Riedel [FR14]; see also Riedel–Xu [RX13]. Let us also note that L^q-rates for the convergence of approximations are not easy to obtain with rough path techniques (in contrast to Itô calculus which is ideally suited for moment calculations). Nonetheless, such rates can be obtained by Gaussian techniques, as discussed in Section 11.2.3 below; applications include multi-level Monte Carlo for SDEs and more generally Gaussian RDEs [BFRS16]. The rough path approach to SDEs (and more generally Gaussian RDEs) leads naturally to random dynamical systems, cf. comment Section 10.5.

The rough path approach to the *Stroock-Varadhan support theorem* [SV72] in Section 9.3 goes back to Ledoux–Qian–Zhang [LQZ02] in p-variation and Friz [Fri05] in Hölder topology, simplified and extended with Victoir in [FV05, FV07, FV10b]; the conditional estimate (9.3) is due to Friz, Lyons and Stroock [FLS06]. We note that this strategy of proof applies whenever one has rough path stability, which includes many stochastic partial differential equations (with finite-dimensional noise) discussed in Chapter 12. In the case of infinite-dimensional noise, a general support theorem for singular SPDEs was obtained via regularity structures by Hairer–Schönbauer [HS19] and extends the paracontrolled work of Chouk–Friz [CF18], as well as classical results such as the work of Bally, Millet and Sanz-Sole [BMSS95].

The rough path approach to *Freidlin–Wentzell* (small noise) *large deviations* in Section 9.3 goes also back to Ledoux, Qian and Zhang [LQZ02]; in p-variation, strengthened to Hölder topology in [FV05]; Inahama studies large deviations for pinned diffusions [Ina15], see also [Ina16a]. Once more, the strategy of proof applies whenever one has rough path stability, and thus applies to many stochastic partial differential equations as discussed in Chapter 12. Large deviations for Banach valued Wiener–Itô chaos proved useful in extensions to Gaussian rough paths and then Gaussian models (in the sense of regularity structures), see [FV07] and [HW15], where Hairer–Weber establish small noise large deviations for large classes of singular SPDEs.

Theorem 9.6 is an elegant application of rough paths, due to Aida [Aid07], to the classical theme of *Laplace method on Wiener space*, in a setting close to Ben Arous [BA88]; see also Inahama [Ina06], his work with Kawabi [IK07] and [Ina13]. Our presentation borrows from Friz, Gassiat and Pigato [FGP18]. See Friz–Klose [FK20] for a recent extension of these works to singular SPDEs via regularity structures. Recent applications to heat kernel expansions include [IT17].

The pathwise approach has also been useful to study *mean field* or *McKean–Vlasov* stochastic differential equations. This goes back to Tanaka [Tan84], with pathwise analysis of additive noise, revisited and extended by Coghi et al. [CDFM18]. The rough path case was pioneered by Cass–Lyons [CL15], with measure dependent drift, followed by Bailleul, Catellier and Delarue [BCD20, BCD19] to a setup that includes the important case of measure dependent noise vector fields. Dawson–Gärtner type large deviations from the McKean-Vlasov limit of weakly interacting diffusions is studied in by [Tan84, CDFM18], and also in Deuschel et al. [DFMS18] via rough paths, always under additive noise. Coghi–Nilssen [CN19] study, from a rough path point of view, McKean-Vlasov diffusion with "common" noise.

The Lions–Sznitman theory of reflecting SDEs [LS84] was revisited from a purely analytic rough path perspective by Aida [Aid15] and Deya et al. [DGHT19a] (existence) Gassiat [Gas20] shows non-uniqueness.

Homogenisation has also seen much impetus from rough path theory. After early works by Lejay–Lyons [LL03], we mention Bailleul–Catellier [BC17] and Kelly–Melbourn [KM16, KM17], who pioneered applications to deterministic homogenisation for fast-slow systems with chaotic noise, work continued by Chevyrev et al. [CFK$^+$19b, CFK$^+$19a, CFKM19].

Stochastic differential equations with jumps, driven by Lévy or general semimartingale noise, noise are well-known [KPP95, Pro05, App09] to require a careful interpretation: forward vs. geometric (a.k.a. Marcus canonical) sense. The pathwise interpretation of such differential equations was started by Williams [Wil01] and essentially completed by Chevyrev, Friz, Shekhar and Zhang [FS17, FZ18, CF19], consistency with the corresponding stochastic theories is also shown.

Rough analysis is "strong" by nature, yet has also proven a powerful tool for "weak" (or martingale) problems. This was pioneered by Delarue–Diehl [DD16], using rough paths to study a one-dimensional *SDE with distributional drift*, with applications to *polymer measures*. The extension to higher dimensions was carried out with paracontrolled methods by Cannizzaro–Chouk [CC18a].

Bruned et al. [BCF18] construct examples of renormalised SDE solutions, partially based on the "Hoff" process [Hof06, FHL16], related to Itô SDE solutions as averaging Stratonovich solutions [LY16].