Chapter 35

Large Deviations for Stochastic Differential Equations

This last chapter revisits large deviations for stochastic differential equations in the small-noise limit, first raised in Chapter 22.

Section 35.1 establishes the LDP for the Wiener process (Schilder's Theorem).

Section 35.2 proves the LDP for stochastic differential equations where the driving noise is independent of the state of the process.

Section 35.3 states the corresponding result for SDEs when the noise is state-dependent, and gestures in the direction of the proof.

In Chapter 22, we looked at how the diffusions X_ϵ which solve the SDE

$$dX_\epsilon = a(X_\epsilon)dt + \epsilon dW, \; X_\epsilon(0) = x_0 \tag{35.1}$$

converge on the trajectory $x_0(t)$ solving the ODE

$$\frac{dx}{dt} = a(x(t)), \; x(0) = x_0 \tag{35.2}$$

in the "small noise" limit, $\epsilon \rightarrow 0$. Specifically, Theorem 256 gave a (fairly crude) upper bound on the probability of deviations:

$$\lim_{\epsilon \rightarrow 0} \epsilon^2 \log \mathbb{P}\left(\sup_{0 \leq t \leq T} \Delta_\epsilon(t) > \delta \right) \leq -\delta^2 e^{-2K_a T} \tag{35.3}$$

where K_a depends on the Lipschitz coefficient of the drift function a. The theory of large deviations for stochastic differential equations, known as Freidlin-Wentzell theory for its original developers, shows that, using the metric implicit in the left-hand side of Eq. 35.3, the family of processes X_ϵ obey a large deviations principle with rate ϵ^{-2}, and a good rate function.

(The full Freidlin-Wentzell theory actually goes somewhat further than just SDEs, to consider small-noise perturbations of dynamical systems of many sorts, perturbations by Markov processes (rather than just white noise), etc. Time does not allow us to consider the full theory (Freidlin and Wentzell, 1998), or its many applications to nonparametric estimation (Ibragimov and Has'minskii, 1979/1981), systems analysis and signal processing (Kushner, 1984), statistical mechanics (Olivieri and Vares, 2005), etc.)

As in Chapter 31, the strategy is to first prove a large deviations principle for a comparatively simple case, and then transfer it to more subtle processes which can be represented as appropriate functionals of the basic case. Here, the basic case is the Wiener process $W(t)$, with t restricted to the unit interval $[0, 1]$.

35.1 Large Deviations of the Wiener Process

We start with a standard d-dimensional Wiener process W, and consider its dilation by a factor ϵ, $X_\epsilon(t) = \epsilon W(t)$. There are a number of ways of establishing that X_ϵ obeys a large deviation principle as $\epsilon \to 0$. One approach (see Dembo and Zeitouni (1998, ch. 5) starts with establishing an LDP for continuous-time random walks, ultimately based on the Gärtner-Ellis Theorem, and then showing that the convergence of such processes to the Wiener process (the Functional Central Limit Theorem, Theorem 174 of Chapter 16) is sufficiently fast that the LDP carries over. However, this approach involves a number of surprisingly tricky topological issues, so I will avoid it, in favor of a more probabilistic path, marked out by Freidlin and Wentzell (Freidlin and Wentzell, 1998, sec. 3.2).

Until further notice, $\|w\|_\infty$ will denote the supremum norm in the space of continuous curves over the unit interval, $\mathbf{C}([0, 1], \mathbb{R}^d)$.

Definition 459 (Cameron-Martin Spaces) *The* Cameron-Martin space H_T *consists of all continuous sample paths $x \in \mathbf{C}([0, T], \mathbb{R}^d)$ where $x(0) = 0$, x is absolutely continuous, and its Radon-Nikodym derivative $\dot{x}$ is square-integrable.*

Lemma 460 *Cameron-Martin spaces are Hilbert spaces, with norm $\|x\|_{CM} = \int_0^T |\dot{x}(t)|^2 dt$.*

Proof: An exercise (35.1) in verifying that the axioms of a Hilbert space are satisfied. □

Definition 461 *The* effective Wiener action *of an continuous function $x \in \mathbf{C}([0, t], \mathbb{R}^d)$ is*

$$J_T(x) \equiv \frac{1}{2}\|x\|_{CM}^2 \tag{35.4}$$

if $x \in H_T$, and ∞ otherwise. In particular,

$$J_1(x) \equiv \frac{1}{2}\int_0^1 |\dot{x}(t)|^2 dt \tag{35.5}$$

For every $j > 0$, let $L_T(j) = \{x : \; J_T(x) \leq j\}$.

Proposition 462 *Fix a function $f \in H_1$, and let $Y_\epsilon = X_\epsilon - f$. Then $\mathcal{L}(Y_\epsilon) = \nu_\epsilon$ is absolutely continuous with respect to $\mathcal{L}(X_\epsilon) = \mu_\epsilon$, and the Radon-Nikodym derivative is*

$$\frac{d\nu_\epsilon}{d\mu_\epsilon}(\epsilon w) = \exp\left\{-\frac{1}{\epsilon}\int_0^1 \dot{w}(t)\cdot dW - \frac{1}{2\epsilon^2}\int_0^1 |\dot{w}(t)|^2 dt\right\} \tag{35.6}$$

PROOF: This is a special case of Girsanov's Theorem. See Corollary 18.25 on p. 365 of Kallenberg, or, more transparently perhaps, the relevant parts of Liptser and Shiryaev (2001, vol. I).□

Lemma 463 *For any $\delta, \gamma, K > 0$, there exists an $\epsilon_0 > 0$ such that, if $\epsilon < \epsilon_0$,*

$$\mathbb{P}\left(\|X_\epsilon - x\|_\infty \leq \delta\right) \geq e^{-\frac{J_1(x)+\gamma}{\epsilon^2}} \tag{35.7}$$

provided $x(0) = 0$ and $J_1(x) < K$.

PROOF: Using Proposition 462,

$$\begin{aligned} \mathbb{P}\left(\|X_\epsilon - x\|_\infty \leq \delta\right) &= \mathbb{P}\left(\|Y_\epsilon - 0\|_\infty \leq \delta\right) & (35.8)\\ &= \int_{\|\epsilon w\|_\infty < \delta} \frac{d\nu_\epsilon}{d\mu_\epsilon}(\epsilon w) d\mu_\epsilon(\epsilon w) & (35.9)\\ &= e^{-\frac{J_1(x)}{\epsilon^2}} \int_{\|\epsilon w\|_\infty < \delta} e^{-\frac{1}{\epsilon}\int_0^1 \dot{x}\cdot dW} d\mu_\epsilon(w) & (35.10) \end{aligned}$$

From Lemma 254 in Chapter 22, we can see that $\mathbb{P}\left(\|\epsilon W\|_\infty < \delta\right) \to 1$ as $\epsilon \to 0$. So, if ϵ is sufficiently small, $\mathbb{P}\left(\|\epsilon W\|_\infty < \delta\right) \geq 3/4$. Now, applying Chebyshev's inequality to the integrand,

$$\begin{aligned} &\mathbb{P}\left(-\frac{1}{\epsilon}\int_0^1 \dot{x}\cdot dW \leq -\frac{2\sqrt{2}}{\epsilon}\sqrt{J_1(x)}\right) & (35.11)\\ \leq\;& \mathbb{P}\left(\left|\frac{1}{\epsilon}\int_0^1 \dot{x}\cdot dW\right| \leq \frac{2\sqrt{2}}{\epsilon}\sqrt{J_1(x)}\right) & (35.12)\\ \leq\;& \frac{\epsilon^2 \mathbf{E}\left[\left(\int_0^1 \dot{x}\cdot dW\right)^2\right]}{8\epsilon^2 J_1(x)} & (35.13)\\ =\;& \frac{\int_0^1 |\dot{x}|^2 dt}{8J_1(x)} = \frac{1}{4} & (35.14) \end{aligned}$$

using the Itô isometry (Corollary 196). Thus

$$\mathbb{P}\left(e^{-\frac{1}{\epsilon}\int_0^1 \dot{x}\cdot dW} \geq e^{-\frac{2\sqrt{2}}{\epsilon}\sqrt{J_1(x)}}\right) \geq \frac{3}{4} \tag{35.15}$$

$$\int_{\|\epsilon w\|_\infty < \delta} e^{-\frac{1}{\epsilon}\int_0^1 \dot{x}\cdot dW} d\mu_\epsilon(w) \quad > \quad \frac{1}{2}e^{-\frac{2\sqrt{2}}{\epsilon}\sqrt{J_1(x)}} \tag{35.16}$$

$$\mathbb{P}\left(\|X_\epsilon - x\|_\infty \leq \delta\right) \quad > \quad \frac{1}{2}e^{-\frac{J_1(x)}{\epsilon^2} - \frac{2\sqrt{2}}{\epsilon}\sqrt{J_1(x)}} \tag{35.17}$$

where the second term in the exponent can be made less than any desired γ by taking ϵ small enough.
□

Lemma 464 *For every $j > 0$, $\delta > 0$, let $U(j,\delta)$ be the open δ neighborhood of $L_1(j)$, i.e., all the trajectories coming within δ of a trajectory whose action is less than or equal to j. Then for any $\gamma > 0$, there is an $\epsilon_0 > 0$ such that, if $\epsilon < \epsilon_0$ and*

$$\mathbb{P}\left(X_\epsilon \notin U(j,\delta)\right) \leq e^{-\frac{s-\gamma}{\epsilon^2}} \tag{35.18}$$

PROOF: Basically, approximating the Wiener process by a continuous piecewise-linear function, and showing that the approximation is sufficiently fine-grained. Chose a natural number n, and let $Y_{n,\epsilon}(t)$ be the piecewise linear random function which coincides with X_ϵ at times $0, 1/n, 2/n, \ldots 1$, i.e.,

$$Y_{n,\epsilon}(t) = X_\epsilon([tn]/n) + \left(t - \frac{[tn]}{n}\right) X_\epsilon([tn+1]/n) \tag{35.19}$$

We will see that, for large enough n, this is exponentially close to X_ϵ. First, though, let's bound the probability in Eq. 35.18.

$$\begin{aligned}
&\mathbb{P}\left(X_\epsilon \notin U(j,\delta)\right) \\
&\quad = \mathbb{P}\left(X_\epsilon \notin U(j,\delta), \|X_\epsilon - Y_{n,\epsilon}\|_\infty < \delta\right) \\
&\qquad + \mathbb{P}\left(X_\epsilon \notin U(j,\delta), \|X_\epsilon - Y_{n,\epsilon}\|_\infty \geq \delta\right) && (35.20)\\
&\quad \leq \mathbb{P}\left(X_\epsilon \notin U(j,\delta), \|X_\epsilon - Y_{n,\epsilon}\|_\infty < \delta\right) + \mathbb{P}\left(\|X_\epsilon - Y_{n,\epsilon}\|_\infty \geq \delta\right) && (35.21)\\
&\quad \leq \mathbb{P}\left(J_1(Y_{n,\epsilon}) > j\right) + \mathbb{P}\left(\|X_\epsilon - Y_{n,\epsilon}\|_\infty \geq \delta\right) && (35.22)
\end{aligned}$$

$J_1(Y_{n,\epsilon})$ can be gotten at from the increments of the Wiener process:

$$J_1(Y_{n,\epsilon}) \quad = \quad n\frac{\epsilon^2}{2}\sum_{i=1}^{n} \left|W(i/n) - W((i-1)/n)\right|^2 \tag{35.23}$$

$$= \quad \frac{\epsilon^2}{2}\sum_{i=1}^{dn} \xi_i \tag{35.24}$$

where the ξ_i have the χ^2 distribution with one degree of freedom. Using our results on such distributions and their sums in Ch. 22, it is not hard to show that, for sufficiently small ϵ,

$$\mathbb{P}\left(J_1(Y_{n,\epsilon}) > j\right) \leq \frac{1}{2} e^{-\frac{j-\gamma}{\epsilon^2}} \tag{35.25}$$

To estimate the probability that the distance between X_ϵ and $Y_{n,\epsilon}$ reaches or exceeds δ, start with the independent-increments property of X_ϵ, and the fact that the two processes coincide when $t = i/n$.

$$\begin{aligned}
& \mathbb{P}\left(\|X_\epsilon - Y_{n,\epsilon}\|_\infty \geq \delta\right) \\
\leq & \sum_{i=1}^{n} \mathbb{P}\left(\max_{(i-1)/n \leq t \leq i/n} |X_\epsilon(t) - Y_{n,\epsilon}(t)| \geq \delta\right) & (35.26)\\
= & \; n\mathbb{P}\left(\max_{0 \leq t \leq 1/n} |X_\epsilon(t) - Y_{n,\epsilon}(t)| \geq \delta\right) & (35.27)\\
= & \; n\mathbb{P}\left(\max_{0 \leq t \leq 1/n} |\epsilon W(t) - n\epsilon W(1/n)| \geq \delta\right) & (35.28)\\
\leq & \; n\mathbb{P}\left(\max_{0 \leq t \leq 1/n} |\epsilon W(t)| \geq \frac{\delta}{2}\right) & (35.29)\\
\leq & \; 4dn\mathbb{P}\left(W_1(1/n) \geq \frac{\delta}{2d\epsilon}\right) & (35.30)\\
\leq & \; 4dn\frac{2d\epsilon}{\delta\sqrt{2\pi n}} e^{-\frac{n\delta^2}{8d^2\epsilon^2}} & (35.31)
\end{aligned}$$

again freely using our calculations from Ch. 22. If $n > 4d^2 j/\delta^2$, then $\mathbb{P}\left(\|X_\epsilon - Y_{n,\epsilon}\|_\infty\right) \leq \frac{1}{2} e^{-\frac{j-\gamma}{\epsilon^2}}$, and we have overall

$$\mathbb{P}\left(X_\epsilon \notin U(j,\delta)\right) \leq e^{-\frac{j-\gamma}{\epsilon^2}} \tag{35.32}$$

as required. □

Proposition 465 *The Cameron-Martin norm has compact level sets.*

PROOF: See Kallenberg, Lemma 27.7, p. 543. □

Theorem 466 (Schilder's Theorem) *If W is a d-dimensional Wiener process on the unit interval, then $X_\epsilon = \epsilon W$ obeys an LDP on $\mathbf{C}([0,1], \mathbb{R}^d)$, with rate ϵ^{-2} and good rate function $J_1(x)$, the effective Wiener action over $[0,1]$.*

PROOF: It is easy to show that Lemma 463 implies the large deviation lower bound for open sets. (Exercise 35.2.) The tricky part is the upper bound. Pick any closed set C and any $\gamma > 0$. Let $s = J_1(C) - \gamma$. By Lemma 465, the set $K = L_1(s) = \{x : J_1(x) \leq s\}$ is compact. By construction, $C \cap K = \emptyset$. So

$\delta = \inf_{x \in C, y \in K} \|x - y\|_{\infty} > 0$. Let U be the closed δ-neighborhood of K. Then use Lemma 464

$$\begin{aligned}
\mathbb{P}\left(X_{\epsilon} \in C\right) &\leq \mathbb{P}\left(X_{\epsilon} \notin U\right) && (35.33)\\
&\leq e^{-\frac{s-\gamma}{\epsilon^2}} && (35.34)\\
&\leq e^{-\frac{J_1(C)-2\gamma}{\epsilon^2}} && (35.35)\\
\log \mathbb{P}\left(X_{\epsilon} \in C\right) &\leq -\frac{J_1(C)-2\gamma}{\epsilon^2} && (35.36)\\
\epsilon^2 \log \mathbb{P}\left(X_{\epsilon} \in C\right) &\leq -J_1(C) - 2\gamma && (35.37)\\
\limsup_{\epsilon \rightarrow 0} \epsilon^2 \log \mathbb{P}\left(X_{\epsilon} \in C\right) &\leq -J_1(C) - 2\gamma && (35.38)
\end{aligned}$$

Since γ was arbitrary, this completes the proof. □

Remark: The trick used here, about establishing results like Lemmas 464 and 463, and then using compact level sets to prove large deviations, works more generally. See Theorem 3.3 in Freidlin and Wentzell (1998, sec. 3.3).

Corollary 467 *Schilder's theorem remains true for Wiener processes on* $[0, T]$*, for all* $T > 0$*, with rate function* J_T*, the effective Wiener action on* $[0, T]$*.*

PROOF: If W is a Wiener process on $[0, 1]$, then, for every T, $S(W) = \sqrt{T} W(t/T)$ is a Wiener process on $[0, T]$. (Show this!) Since the mapping S is continuous from $\mathbf{C}([0, 1], \mathbb{R}^d)$ to $\mathbf{C}([0, T], \mathbb{R}^d)$, by the Contraction Principle (Theorem 410) the family $\epsilon S(W)$ obey an LDP with rate ϵ^{-2} and good rate function $J_T(x) = J_1(S^{-1}(x))$. (Notice that S is invertible, so $S^{-1}(x)$ is a function, not a set of functions.) Since $x \in H_1$ iff $y = S(x) \in H_T$, it's easy to check that for such, $\dot{y}(t) = T^{-1/2} \dot{x}(t/T)$, meaning that

$$\|y\|_{CM} = \int_0^T |\dot{y}(t)|^2 dt = \int_0^T |\dot{x}(t/T)|^2 \frac{dt}{T} = \|x\|_{CM} \quad (35.39)$$

which completes the proof. □

Corollary 468 *Schilder's theorem remains true for Wiener processes on* $\mathbb{R}^+$*, with good rate function* J_∞ *given by the effective Wiener action on* $\mathbb{R}^+$*,*

$$J_\infty(x) \equiv \frac{1}{2} \int_0^\infty |\dot{x}(t)|^2 dt \quad (35.40)$$

if $x \in H_\infty$*,* $J_\infty(x) = \infty$ *otherwise.*

PROOF: For each natural number n, let $\pi_n x$ be the restriction of x to the interval $[0, n]$. By Corollary 467, each of them obeys an LDP with rate function $\frac{1}{2} \int_0^n |\dot{x}(t)|^2 dt$. Now apply the projective limit theorem (420) to get that $J_\infty(x) = \sup_n J_n(x)$, which is clearly Eq. 35.40, as the integrand is non-negative. □

35.2 Large Deviations for SDEs with State-Independent Noise

Having established an LDP for the Wiener process, it is fairly straightforward to get an LDP for stochastic differential equations where the driving noise is *independent* of the state of the diffusion process.

Definition 469 (SDE with Small State-Independent Noise) *An* SDE with small state-independent noise *is a stochastic differential equation of the form*

$$dX_\epsilon = a(X_\epsilon)dt + \epsilon dW \quad (35.41)$$

$$X_\epsilon(0) = 0 \quad (35.42)$$

where $a : \mathbb{R}^d \mapsto \mathbb{R}^d$ *is uniformly Lipschitz continuous.*

Notice that any non-random initial condition x_0 can be handled by a simple change of coordinates.

Definition 470 (Effective Action: State-Independent Noise) *The* effective action *of a trajectory* $x \in H_\infty$ *is*

$$J(x) \equiv \frac{1}{2}\int_0^t |\dot{x}(t) - a(x(t))|^2 dt \quad (35.43)$$

and $= \infty$ *if* $x \in \mathbf{C} \setminus H_\infty$.

Lemma 471 *The map* $F : \mathbf{C}(\mathbb{R}^+, \mathbb{R}^d) \mapsto \mathbf{C}(\mathbb{R}^+, \mathbb{R}^d)$ *given by*

$$x(t) = w(t) + \int_0^t a(w(s))ds \quad (35.44)$$

when $x = F(w)$ *is continuous.*

Proof: This goes rather in the same manner as the proof of existence and uniqueness for SDEs (Theorem 216). For any $w_1, w_2 \in \mathbf{C}(\mathbb{R}^+, \mathbb{R}^d)$, set $x_1 = F(w_1)$, $x_2 = F(w_2)$. From the Lipschitz property of a,

$$|x_1(t) - x_2(t)| \leq \|w_1 - w_2\| + K_a \int_0^t |x_1(s) - x_2(s)|\, ds \quad (35.45)$$

(writing $|y(t)|$ for the norm of Euclidean vectors y, and $\|x\|$ for the supremum norm of continuous curves). By Gronwall's Inequality (Lemma 214), then,

$$\|x_1 - x_2\| \leq \|w_1 - w_2\|\, e^{K_a T} \quad (35.46)$$

on every interval $[0, T]$. So we can make sure that $\|x_1 - x_2\|$ is less than any desired amount by making sure that $\|w_1 - w_2\|$ is sufficiently small, and so F is continuous.

Lemma 472 *If $w \in H_\infty$, then $x = F(w)$ is in H_∞.*

PROOF: Exercise 35.3. □

Theorem 473 (Freidlin-Wentzell Theorem: State-Independent Noise) *The Itô processes X_ϵ of Definition 469 obey the large deviations principle with rate ϵ^{-2} and good rate function given by the effective action $J(x)$.*

PROOF: For every ϵ, $X_\epsilon = F(\epsilon W)$. Corollary 468 tells us that ϵW obeys the large deviation principle with rate ϵ^{-2} and good rate function J_∞. Since (Lemma 471) F is continuous, by the Contraction Principle (Theorem 410) X_ϵ also obeys the LDP, with rate given by $J_\infty(F^{-1}(x))$. If $F^{-1}(x) \cap H_\infty = \emptyset$, this is ∞. On the other hand, if $F^{-1}(x)$ does contain curves in H_∞, then $J_\infty(F^{-1}(x)) = J_\infty(F^{-1}(x) \cap H_\infty)$. By Lemma 472, this implies that $x \in H_\infty$, too. For any curve $w \in F^{-1}(x) \cap H_\infty$, $\dot{x} = \dot{w} + a(x)$, or $\dot{w} = \dot{x} - a(x)$. $J_\infty(w) = \int_0^\infty |\dot{x} - a(x)|^2 dt$ is however the effective action of the trajectory x (Definition 470). □

35.3 Large Deviations for State-Dependent Noise

If the diffusion term in the SDE does depend on the state of the process, one obtains a very similar LDP to the results in the previous section. However, the approach must be modified: the mapping from W to X_ϵ, while still measurable, is no longer necessarily continuous, so we can't use the contraction principle as before.

Definition 474 (SDE with Small State-Dependent Noise) *An* SDE with small state-dependent noise *is a stochastic differential equation of the form*

$$dX_\epsilon = a(X_\epsilon)dt + \epsilon b(X_\epsilon)dW \quad (35.47)$$

$$X_\epsilon(0) = 0 \quad (35.48)$$

where a and b are uniformly Lipschitz continuous, and b is non-singular.

Definition 475 (Effective Action: State-Dependent Noise) *The* effective action *of a trajectory $x \in H_\infty$ is given by*

$$J(x) \equiv \int_0^\infty L(x(t), \dot{x}(t))dt \quad (35.49)$$

where

$$L(q,p) = \frac{1}{2}\left(p_i - a_i(q)\right) B_{ij}^{-1}(q) \left(p_j - a_j(q)\right) \quad (35.50)$$

and

$$B(q) = b(q)b^T(q) \quad (35.51)$$

with $J(x) = \infty$ if $x \in \mathbf{C} \setminus H_\infty$.

Theorem 476 (Freidlin-Wentzell Theorem: State-Dependent Noise) *The processes X_ϵ obey a large deviations principle with rate ϵ^{-2} and good rate function equal to the effective action.*

PROOF: Considerably more complicated. (See, e.g., Dembo and Zeitouni (1998, sec. 5.6, pp. 213–220).) The essence, however, is to consider an approximating Itô process X_n, where $a(X_t)$ and $b(X_t)$ are replaced in Eq. 35.47 by $a(X_n([tn]/n))$ and $b(X_n([tn]/n))$. Here the mapping from W to X_n is continuous, so it's not too hard to show that the latter obey an LDP with a reasonable rate function, and also that they're exponentially equivalent (in n) to X_ϵ. □

35.4 Exercises

Exercise 35.1 *Prove Lemma 460.*

Exercise 35.2 *Prove that Lemma 463 implies the large deviations lower bound for open sets.*

Exercise 35.3 *Prove Lemma 472.* Hint: *Use Gronwall's Inequality (Lemma 214) again to show that F maps H_T into H_T, and then show that $H_\infty = \bigcap_{n=1}^{\infty} H_n$.*