1. Strassen's Law of the Iterated Logarithm.

Let P be the Wiener measure on the space $\Omega = C[0,\infty)$ of continuos functions on $[0,\infty)$ that starts at time 0 from the point 0. For $\lambda \geq 3$ we define the rescaled process

$$x_\lambda(t) = \frac{1}{\sqrt{\lambda \log\log \lambda}} x(\lambda t).$$

As $\lambda \to \infty$, $x_\lambda(t)$ will go to 0 in probability with respect to P, but the convergence will not be almost sure. Strassen's theorem states:

Theorem 1.1. *On any fixed time interval, say $[0,1]$, for almost all $\omega = x(\cdot)$, the family $\{x_\lambda(\cdot) : \lambda \geq 3\}$ is compact and has as its set of limit points the compact set*

$$K = \{f : I(f) \leq 1\}$$

where

$$I(f) = \frac{1}{2}\int_0^1 [f'(t)]^2 dt.$$

Proof. We will divide the proof into several steps. The first part is to prove that If K_ϵ is a neighborhhod of size ϵ around K then, for almost all ω, $x_\lambda(\cdot) \in K_\epsilon$ for sufficiently large λ. This is proved in two steps.

Fisrt we sample $x_\lambda(\cdot)$ along a discrete sequence $\lambda = \rho^n$ for some $\rho > 1$ and show that almost surely, for any such ρ, $x_{\rho^n}(\cdot) \in K_{\frac{\epsilon}{2}}$ for sufficiently large n. This requires just the Borel-Cantelli lemma. We need to show that

$$\sum_n \Pr[x_{\rho^n}(\cdot) \in K^c_{\frac{\epsilon}{2}}] < \infty.$$

We use the results of the LDP proved in the last section to estimate for any closed set C,

$$\Pr[x_{\rho^n}(\cdot) \in C] = P_{a_n}[C]$$

where P_{a_n} is the Wiener measure scaled by $\frac{1}{\sqrt{\log\log\rho^n}} \sim \frac{1}{\sqrt{\log n}}$. From the results proved in the last section

$$\log P_{a_n}[C] \leq -[\ell - \delta]\log n$$

for sufficiently large n, and to complete this step we need to prove only that

$$\ell = \inf_{f \in C} I(f) = \inf_{f \in K^c_{\frac{\epsilon}{2}}} I(f) > 1$$

which is obvious because we have removed a neighborhood of the set K consisting of all f with $I(f) \leq 1$.

The second step is to show that the price for sampling is not too much. More precisely we will show that, almost surely,

$$\limsup_{n\to\infty} \sup_{\rho^n \leq \lambda \leq \rho^{n+1}} \|x_\lambda(\cdot) - x_{\rho^{n+1}}(\cdot)\| \leq \theta(\rho)$$

where $\theta(\rho)$ is nonrandom and $\to 0$ as $\rho \downarrow 1$. We then pick our ρ so that $\theta(\rho) < \frac{\epsilon}{2}$ to complete the proof. If $\lambda_2 > \lambda_1$ we can estimate

$$
\begin{aligned}
|x_{\lambda_2}(t) - x_{\lambda_1}(t)| &\le \Big|\frac{1}{\sqrt{\lambda_2 \log\log \lambda_2}} - \frac{1}{\sqrt{\lambda_1 \log\log \lambda_1}}\Big|\ |x(\lambda_1 t)| + \frac{1}{\sqrt{\lambda_2 \log\log \lambda_2}}|x(\lambda_2 t) - x(\lambda_1 t)| \\
&\le \left|\sqrt{\frac{\lambda_1 \log\log \lambda_1}{\lambda_2 \log\log \lambda_2}} - 1\right| \|x_{\lambda_1}(\cdot)\| + |x_{\lambda_2}(t) - x_{\lambda_2}(\frac{\lambda_1}{\lambda_2}t)|
\end{aligned}
$$

Taking the supremum over $0 \le t \le 1$,

$$
\|x_{\lambda_2}(\cdot) - x_{\lambda_1}(\cdot)\| \le \left|\sqrt{\frac{\lambda_2 \log\log \lambda_2}{\lambda_1 \log\log \lambda_1}} - 1\right| \|x_{\lambda_2}(\cdot)\| \ + \sup_{|t-s| \le |\frac{\lambda_1}{\lambda_2} - 1|} |x_{\lambda_2}(t) - x_{\lambda_2}(s)|.
$$

If we now take $\lambda_2 = \rho^{n+1}$ and $\lambda_1 = \lambda$ with $\rho^n \le \lambda \le \rho^{n+1}$

$$
\limsup_{n\to\infty} \sup_{\rho^n \le \lambda \le \rho^{n+1}} \|x_\lambda(\cdot) - x_{\rho^{n+1}}(\cdot)\| \le |\sqrt{\rho} - 1| \limsup_{n\to\infty} \|x_{\rho^{n+1}}(\cdot)\| + \limsup_{n\to\infty} \sup_{|t-s| \le |1-\frac{1}{\rho}|} |x_{\rho^{n+1}}(t) - x_{\rho^{n+1}}(s)|.
$$

One of the consequences of the result proved in the earlier step is that for any continuos functional $F : \Omega \to R$, almost surely,

$$
\limsup_{n\to\infty} F(x_{\rho^n}(\cdot)) \le \sup_{f\in K} F(f).
$$

Therefore, almost surely

$$
\begin{aligned}
\limsup_{n\to\infty} \sup_{\rho^n \le \lambda \le \rho^{n+1}} \|x_\lambda(\cdot) - x_{\rho^{n+1}}(\cdot)\| &\le |\sqrt{\rho} - 1| \sup_{f\in K} \|f\| + \sup_{f\in K} \sup_{|t-s| \le 1-\frac{1}{\rho}} |f(t) - f(s)| \\
&= \theta(\rho)
\end{aligned}
$$

and it is easily seen that $\theta(\rho) \to 0$ as $\rho \downarrow 1$.

Now we turn to the second part where we need to prove that $x_\lambda(\cdot)$ returns infinitely often to any neighborhood of any point $f \in K$. We can assume without loss of generality that $I(f) = \ell < 1$, for such points are dense in K. Again we apply the Borel-Cantelli lemma but now we need independence. Let us define $a_n = \rho^n - \rho^{n-1}$, and for $0 \le t \le 1$

$$
y_n(t) = \frac{1}{\sqrt{a_n \log\log a_n}}[x(\rho^{n-1}t + a_n t) - x(\rho^{n-1})].
$$

The distribution of $y_n(\cdot)$ is the same as that of Brownian motion scaled by $\frac{1}{\sqrt{\log\log a_n}}$ and from the LDP results of the last section for any $\eta > 0$

$$
\log \Pr\Big[\ \|y_n(\cdot) - f\| < \frac{\delta}{2}\ \Big] \ge -(\ell + \eta)\log n
$$

for sufficiently large n and this shows that

$$
\sum_n \Pr\Big[\ \|y_n(\cdot) - f\| < \frac{\delta}{2}\ \Big] = +\infty.
$$

Because $y_n(\cdot)$ are independent, by Borel-cantelli lemma, $y_n(\cdot)$ returns infinitely often to the $\frac{\delta}{2}$ neighborhood of f.

The last piece of the proof then is to show that, almost surely ,

$$\limsup_{n\to\infty} \|x_{\rho^n}(\cdot) - y_n(\cdot)\| \le \theta(\rho)$$

where $\theta(\rho) \to 0$ as $\rho \uparrow \infty$. We could then complete the proof by picking ρ large enough that $\theta(\rho) < \frac{\delta}{2}$. Then $x_{\rho^n}(\cdot)$ would return infinitely often to the δ neighborhood of f. Hence so does $x_\lambda(\cdot)$ for every $\delta > 0$.

$$\begin{aligned}
|x_{\rho^n}(t) - y_n(t)| &= \left| \frac{x(\rho^n t)}{\sqrt{\rho^n \log\log \rho^n}} - \frac{(x(\rho^n + a_n t) - x(\rho^n))}{\sqrt{a_n \log\log a_n}} \right| \\
&\le \Big| \frac{1}{\sqrt{\rho^n \log\log \rho^n}} - \frac{1}{\sqrt{a_n \log\log a_n}} \Big| \; |x(\rho^n t)| + \frac{1}{\sqrt{a_n \log\log a_n}} |x(\rho^n t) - x(\rho^{n-1} + a_n t)| \\
&\qquad + \frac{1}{\sqrt{a_n \log\log a_n}} |x(\rho^n)| \\
&\le \Big| \frac{\sqrt{\rho^n \log\log \rho^n}}{\sqrt{a_n \log\log a_n}} - 1 \Big| \; |x_{\rho^n}(t)| + \frac{\sqrt{\rho^n \log\log \rho^n}}{\sqrt{a_n \log\log a_n}} |x_{\rho^n}(t) - x_{\rho^n}(\frac{1}{\rho} + [1 - \frac{1}{\rho}]t)| \\
&\qquad + \frac{\sqrt{\rho^n \log\log \rho^n}}{\sqrt{a_n \log\log a_n}} \; |x_{\rho^n}(\frac{1}{\rho})|
\end{aligned}$$

Taking the supremum over $0 \le t \le 1$,

$$\|x_{\rho^n}(\cdot) - y_n(\cdot)\| \le \Big| \frac{\sqrt{\rho^n \log\log \rho^n}}{\sqrt{a_n \log\log a_n}} - 1 \Big| \; \|x_{\rho^n}(\cdot)\| + \frac{\sqrt{\rho^n \log\log \rho^n}}{\sqrt{a_n \log\log a_n}} \Big[\sup_{|t-s| \le \frac{1}{\rho}} |x_{\rho^n}(t) - x_{\rho^n}(s)| + x_{\rho^n}(\frac{1}{\rho}) \Big].$$

Again from the first part we conclude that, almost surely,

$$\begin{aligned}
\limsup_{n\to\infty} \|x_{\rho^n}(\cdot) - y_n(\cdot)\| &\le \left| \sqrt{\frac{\rho}{\rho - 1}} - 1 \right| \sup_{f\in K} \|f\| + \sqrt{\frac{\rho}{\rho - 1}} \sup_{f \in K} \big[f(\frac{1}{\rho}) + \sup_{|t-s|\le \frac{1}{\rho}} |f(t) - f(s)| \big] \\
&= \theta(\rho)
\end{aligned}$$

It is easily checked that $\theta(\rho) \to 0$ as $\rho \uparrow \infty$. This concludes the proof. $\square$

Remark 1.2. As we commented earlier we can calculate for any continuous $F : \Omega \to R$,

$$\limsup_{\lambda\to\infty} F(x_\lambda(\cdot)) = \sup_{f\in K} F(f)$$

almost surely. Some simple examples are: if $F(f) = f(1)$ we get, almost surely,

$$\limsup_{t\to\infty} \frac{x(t)}{\sqrt{t \log\log t}} = \sqrt{2} = \sup_{f\in K} f(1)$$

or if $F(f) = \sup_{0\le s\le 1} |f(s)|$, we get for almost all Wiener paths,

$$\limsup_{t\to\infty} \frac{\sup_{0\le s\le t} |x(s)|}{\sqrt{t \log\log t}} = \sqrt{2} = \sup_{f\in K} \sup_{0\le s\le 1} |f(s)|$$

Remark 1.3. There is a way of recovering laws of iterated logarithms for sums of independent random variables from Strassen's theorem for Brownian Motion. This requires a concept known as Skorohod imbedding. If X is a random variable with mean zero and variance σ^2, we find a stopping time τ (perhaps randomized) such that $E\{\tau\} = \sigma^2$and $x(\tau)$ has the same distribution as X. Then the random walk gets imbedded in the Brownian Motion and the LIL for random walk is deduced from the LIL for the Brownian Motion. For instance if $X = \pm 1$ with probability $\frac{1}{2}$ each, τ is the hitting time of ± 1. As an excercise look up the reference and study the details.

2. Behavior of Diffusions with a small parameter.

In this section we will investigate the Large Deviation behavior of the family of diffusion processes P_x^ϵ corresponding to the generator

$$\mathcal{L}_\epsilon = \frac{\epsilon^2}{2} \sum_{i,j} a_{i,j}(x) \frac{\partial^2}{\partial x_i\, \partial x_j} + \sum_j b_j(x) \frac{\partial}{\partial x_j}$$

that start from the point $x \in R^d$. The family P_x^ϵ will be viewed as a family of measures on the space $C[[0,T]; R^d]$ of continuous functions on $[0,T]$ with values in R^d. As we let $\epsilon \to 0$, the generator converges to the first order operator

$$\mathcal{L}_0 = \sum_j b_j(x) \frac{\partial}{\partial x_j}.$$

If we impose enough regularity on $b(x) = \{b_j(\cdot)\}$ so that the trajectories of the ODE

$$\frac{dx(t)}{dt} = b(x)$$

are unique, then the processes P_x^ϵ will converge as $\epsilon \to 0$ to the degenerate distribution concentrated at the unique solution of the above ODE that starts from the initial point $x(0) = x$. We are interested in the validity of LDP for these measures P_x^ϵ.

If we use the theory of Stochastic Differential Equations, we would take a square root σ such that $\sigma\sigma* = a = \{a_{ij}\}$ and solve the SDE

$$dx(t) = \epsilon\sigma(x(t))d\beta(t) + b(x(t))dt$$

with $x(0) = x$. We would then view the solution $x(t)$ as a map $\Phi_{x,\epsilon}$ of the Wiener space $C[[0.T]; R^d]$ with the Wiener measure Q back to $C[[0.T]; R^d]$ and the map will induce P_x^ϵ as the image of Q_ϵ. In fact we can absorb ϵ in β and rewrite the SDE as

$$dx(t) = \sigma(x(t))d[\epsilon\beta(t)] + b(x(t))dt$$

and think of the map Φ_x as independent of ϵ and mapping the scaled Wiener measures Q_ϵ into P_x^ϵ. The advantage now is that we may try to appeal to the contraction principle and deduce the LDP of P_x^ϵ from that of Q_ϵ that we established in Section 3. The map $\Phi_x : f(\cdot) \to x(\cdot)$ is defined by

$$dx(t) = \sigma(x(t))df(t) + b(x(t))dt \ ; \ x(0) = x.$$

Let us ignore for the moment that this map is not defined everywhere and is far from continuous. At the level of rate functions we see the map as

$$g'(t) = \sigma(g(t))f'(t) + b(g(t)) \ ; \ g(0) = x$$

and if we assume that σ or equivalently a is invertible,

$$f'(t) = \sigma^{-1}(g(t))[g'(t) - b(g(t))]$$

and

$$\frac{1}{2}\int_0^T \|f'(t)\|^2 dt = \frac{1}{2}\int_0^T \langle [g'(t) - b(g(t))], a^{-1}(g(t))[g'(t) - b(g(t))]\rangle dt$$

The reasoning above is not quite valid because the maps Φ_x are not continuous and the contraction principle is not directly applicable. We will have to replace Φ_x by continuous maps $\Phi_{n,x}$ and try to interchange limits. Although we can do it in one step, inorder to illustrate our methods in a better manner, we will perform this in two steps. Let us suppose first that $b \equiv 0$. The map $\Phi_{n,x}$ is defined by

$$x_n(t) = x + \int_0^t \sigma(x_n(\pi_n(s)))d\beta(s)$$

where

$$\pi_n(s) = \frac{[ns]}{n}$$

Although $x_n(\cdot)$ appears to be defined implicitly it is in fact defined explicitly, by induction on j, using the updating rule

$$x_n(t) = x_n(\frac{j}{n}) + \int_{\frac{j}{n}}^t \sigma(x_n(\frac{j}{n}))d\beta(s)$$

for $\frac{j}{n} \le t \le \frac{j+1}{n}$. The map $\Phi_{n,x}$ are clearly continuous and the contraction principle applies to yield an LDP for the distribution $P_{n,x}^\epsilon$ of $\Phi_{n,x}$ under Q_ϵ with a rate function that is easily seen to equal

$$I_n(g) = \frac{1}{2}\int_0^T \langle g'(t), a^{-1}(g(\pi_n(t)))g'(t)\rangle dt$$

for functions g with $g(0) = x$ that have a square integrable derivative in t. Otherwise $I_n(g) = +\infty$. We will prove the following superexponential approximation theorem.

Theorem 2.1. *For any $\delta > 0$, and compact set $K \subset R^d$,*

$$\limsup_{n\to\infty} \limsup_{\epsilon\to 0} \sup_{x\in K} \epsilon^2 \log Q_\epsilon\big[\|\Phi_{n,x}(\cdot) - \Phi_x(\cdot)\| \ge \delta\big] = -\infty$$

If we have the approximation theorem then it is straight forward to interchange the ϵ and n limits.

Theorem 2.2. *For the measures $P^{\epsilon}_{n,x}$ an LDP holds with the rate function*

$$I_x(f) = \frac{1}{2}\int_0^T \langle f'(t), a^{-1}(f(t))f'(t)\rangle dt$$

for functions $f(t)$ with a square integrable derivative that satisfy $f(0) = x$ and equal to $+\infty$ otherwise.

Proof. Let $C \in C[0.T]$ be closed and $\delta > 0$ be positive. Then

$$P^{\epsilon}_x[C] = Q_\epsilon\big[\Phi_x(\cdot) \in C\big] \leq Q_\epsilon\big[\Phi_{n,x}(\cdot) \in C^\delta\big] + Q_\epsilon\big[\|\Phi_{n,x}(\cdot) - \Phi_x(\cdot)\| \geq \delta\big]$$

Taking logarithms, multiplying by ϵ^2 and taking limsups

$$\limsup_{\epsilon\to 0} \epsilon^2 \log P^{\epsilon}_x[C] \leq -\max\{a_{n,x}(\delta), b_{n,x}(\delta)\}$$

where

$$a_{n,x}(\delta) = \inf_{g\in \bar{C}^\delta} I_{n.x}(g)$$

and in view of the superexponential estimate,

$$\limsup_{n\to\infty} b_{n,x}(\delta) = -\infty.$$

We therefore obtain for every $\delta > 0$,

$$\limsup_{\epsilon\to 0} \epsilon^2 \log P^{\epsilon}_x[C] \leq -\limsup_{n\to\infty} a_{n,x}(\delta)$$

and finally letting $\delta \to 0$, it is easily seen that

$$\limsup_{\delta\to 0}\limsup_{n\to\infty} a_{n,x}(\delta) \geq -\inf_{g\in C} I_x(g).$$

To prove the lower bound, we take a small ball $B(g,\delta)$ around g and

$$P^{\epsilon}_x[B(g,\delta)] = Q_\epsilon[\Phi_x \in B(g,\delta)] \geq Q_\epsilon[\Phi_{n,x} \in B(g,\frac{\delta}{2})] - Q_\epsilon[\ \|\Phi_{n,x}(\cdot) - \Phi_x(\cdot)\| \geq \frac{\delta}{2}]$$

For large n the second term decays a lot faster than the first term and so we get the lower bound

$$\liminf_{\epsilon\to 0} \epsilon^2 \log P^{\epsilon}_x[B(g,\delta)] \geq -\liminf_{n\to\infty} I_{n,x}(g) = -I_x(g)$$

and we are done. $\square$

We now return to the proof of the Theorem on super-exponential estimates. We will carry it out in several steps each of which will be formulated as a lemma.

Lemma 2.3. *Let $\xi(t)$ be a stochastic integral of the form*

$$\xi(t) = \int_0^t \sigma(s,\omega) d\beta(s)$$

with values in R^d, with a σ satisfying the bound

$$\sigma(s,\omega)\sigma^*(s,\omega) \le C(\|\xi(s)\|^2 + \delta)\ I$$

for some constant C. Let $\alpha > 0$ be a positive number and let τ_α be the stopping time

$$\tau_\alpha = \inf\{t : \|x(t)\| \ge \alpha\}.$$

Then for any $T > 0$

$$\Pr[\tau_\alpha \le T] \le \exp\Big[- \frac{[\log(1+\frac{\alpha^2}{\delta^2})]^2}{4k_d CT}\Big].$$

where k_d is a constant depending only on the dimension.

Proof. Consider the function

$$U(x) = (\delta^2 + \|x\|^2)^N$$

with the choice of N to be made later. One can easily estimate

$$\frac{1}{2}\sum_{i,j}(\sigma\sigma^*)_{i,j}(s,\omega)\frac{\partial^2 U}{\partial x_i \partial x_j}(\xi(s)) \le k_d C N^2 U(\xi(s))$$

where k_d is a constant depending only on the dimension d. If we pick N so that, for some $\lambda > 0$,

$$k_d C N^2 = \lambda$$

then $e^{-\lambda t}U(\xi(t))$ is a supermartingale and

$$E\{\exp[-\lambda\tau_\alpha]\} \le \left[\frac{\delta^2}{\delta^2+\alpha^2}\right]^N$$

and by Tchebechev's inequality

$$\Pr[\tau_\alpha \le T] \le \exp[\lambda T]\left[\frac{\delta^2}{\delta^2+\alpha^2}\right]^N$$

Making the optimal choice of

$$\lambda = \frac{[\log(1+\frac{\alpha^2}{\delta^2})]^2}{4k_d CT}$$

we obtain

$$\Pr[\tau_\alpha \le T] \le \exp\Big[- \frac{[\log(1+\frac{\alpha^2}{\delta^2})]^2}{4k_d CT}\Big].$$

□

Lemma 2.4. *If we drop the assumption that*

$$\sigma(s,\omega)\sigma^*(s,\omega) \le C(\|\xi(s)\|^2 + \delta)\ I$$

we get instead the estimate

$$\Pr\big[\tau_\alpha \le T\big] \le \exp\left[-\frac{[\log(1+\frac{\alpha^2}{\delta^2})]^2}{4k_d CT}\right] + 1 - \Pr\big[\sigma(s,\omega)\sigma^*(s,\omega) \le C(\|\xi(s)\|^2 + \delta)\ I\ \forall s \in [0,T]\big].$$

Proof. The proof is routine. Just have to make sure that the stochastic integral is modified in an admissible manner to give the proof. Try the details as an exercise. □

We now return to prove the super-exponential estimate.

Proof.

$$\begin{aligned} x_n(t) - x(t) &= \epsilon \int_0^t [\sigma(x_n(\pi_n(s))) - \sigma(x(s))] d\beta(s) \\ &= \epsilon \int_0^t [\sigma(x_n(\pi_n(s))) - \sigma(x_n(s))] d\beta(s) + \epsilon \int_0^t [\sigma(x_n(s)) - \sigma(x(s))] d\beta(s) \end{aligned}$$

If we fix n and consider

$$\xi(t) = \int_0^t \sigma(s,\omega) d\beta(s)$$

with

$$\sigma(s,\omega) = \epsilon\big[[\sigma(x_n(\pi_n(s))) - \sigma(x_n(s))] + [\sigma(x_n(s)) - \sigma(x(s))]\big]$$

then

$$\sigma\sigma * (s,\omega) \le C\epsilon^2[\ \|\xi(s)\|^2 + \delta^2]\ I$$

provided

$$\sup_{0\le s\le T} \|x_n(\pi_n(s)) - x_n(s)\| \le \delta.$$

If we apply the earlier lemma we get

$$Q_\epsilon\big[\|\Phi_{n,x}(\cdot) - \Phi_x(\cdot)\| \ge \alpha\big] \le \exp\left[-\frac{[\log(1+\frac{\alpha^2}{\delta^2})]^2}{4k_d C\epsilon^2 T}\right] + Q_\epsilon\big[\sup_{0\le s\le T} \|x_n(\pi_n(s)) - x_n(s)\| \ge \delta\big]$$

If we assume that $\sigma(\cdot)$ is uniformly bounded, the second term is dominated by

$$nT \exp[-\frac{n\delta^2}{A\epsilon^2}].$$

for some constant A. We now conclude the proof of the superexponential estimate by estimating

$$\limsup_{\epsilon\to 0} \epsilon^2 \log Q_\epsilon\big[\|\Phi_{n,x}(\cdot) - \Phi_x(\cdot)\| \ge \alpha\big] \le -\min\left\{\frac{[\log(1+\frac{\alpha^2}{\delta^2})]^2}{4k_d CT}, \frac{n\delta^2}{C\epsilon^2}\right\}$$

For fixed α and δ we can let $n \to \infty$ and then let $\delta \to 0$ keeping α fixed to get $-\infty$ on the right hand side. □

Remark 2.5. If we look at the distribution at time 1 of P_x^ϵ we have the transition probability $p_\epsilon(1, x, dy)$ that is viewed as probability measures on R^d. The LDP for them is now obtained by the contraction principle with the rate function

$$\begin{aligned} I(x, y) &= \inf_{\substack{f(0)=x \\ f(1)=y}} \frac{1}{2} \int_0^1 \langle f'(t), a^{-1}(f(t))\ f'(t)\rangle\ dt \\ &= \frac{[d(x, y)]^2}{2} \end{aligned}$$

where d is the Riemannian distance in the metric $ds^2 = a_{i,j}^{-1}(x) dx_i dx_j$.

Remark 2.6. The distribution $p_\epsilon(1, x, dy)$ is the same as $p_1(\epsilon^2, x, dy)$ and so for the transition probability distributions $p(t, x, dy)$ of the diffusion process with generator

$$\mathcal{L} = \frac{1}{2} \sum_{i,j} a_{i,j}(x) \frac{\partial^2}{\partial x_i \partial x_j}$$

we hace an LDP as $t \to 0$ with rate function $\frac{[d(x,y)]^2}{2}$.

Remark 2.7. All the results are valid locally uniformly in the starting point x. What we mean by that is if we take $P_\epsilon = P_{x_\epsilon,\epsilon}$, so long as $x_\epsilon \to x$ as $\epsilon \to 0$, the LDP is valid for P_ϵ with the same rate function as the one for $P_{x,\epsilon}$.

Remark 2.8. Just for the record the assumptions on a are boundedness, uniform ellipticity or the boundedness of a^{-1} and a uniform Lipschitz condition on a or equivalently on σ.

Remark 2.9. Under the above regularity conditions one can look around in PDE books and find that for the equation

$$\frac{\partial u}{\partial t} = \mathcal{L} u$$

there is a fundamental solution which is nothing else but the density $p(t, x, y)$ of the transition probability distribution. There are some reasonable estimates for it that imply

$$\limsup_{|x-y|\to 0} \limsup_{t\to 0} t \log p(t, x, y) = \liminf_{|x-y|\to 0} \liminf_{t\to 0} t \log p(t, x, y) = 0.$$

One can use the Chapman-Kolmogorov equations to write

$$p(t, x, y) = \int_{R^d} p(\alpha t, x, dz) p(1 - \alpha t, z, y)$$

and bootstrap from the LDP of $p(t, x, dy)$ to the behavior of $p(t, x, y)$ itself

$$\lim_{t\to 0} t \log p(t, x, y) = -\frac{[d(x, y)]^2}{2}.$$

We now turn our attention to the operator

$$\mathcal{L}_\epsilon = \frac{\epsilon^2}{2} \sum_{i,j} a_{i,j}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_j b_j(x) \frac{\partial}{\partial x_j}$$

If we denote by $P^\epsilon_{x,b}$ the probability measure corresponding to the above operator, and by P^ϵ_x the measure corresponding to the operator with $b \equiv 0$, Girsanov formula provides the Radon-Nikodym derivative

$$\frac{dP^\epsilon_{x.b}}{dP^\epsilon_x} = \exp\left\{\frac{1}{\epsilon^2}\int_0^T \langle a^{-1}(x(s)b(x(s)), dx(s)\rangle - \frac{1}{2\epsilon^2}\int_0^T \langle b(x(s)), a^{-1}(x(s))b(x(s))\rangle ds\right\}$$

If we pretend that the exponent

$$F(\omega) = \int_0^T \langle a^{-1}(x(s)b(x(s)), dx(s)\rangle - \frac{1}{2}\int_0^T \langle b(x(s)), a^{-1}(x(s))b(x(s))\rangle ds$$

is a bounded, continuous function on $C[0,T]$, then the LDP for $P^\epsilon_{x,b}$ would follow from the LDP for P^ϵ_x with the rate function

$$I_{x.b}(g) = I_x(g) + F(g) = \frac{1}{2}\int_0^T \langle [g'(t) - b(g(t))], a^{-1}(g(t))[g'(t) - b(g(t))]\rangle\ dt$$

which is what we wanted to prove. The problem really is taking care of the irregular nature of the function $F(\cdot)$. We leave it as an exercise to show that the following lemma will do the trick.

Lemma 2.10. *There exist bounded continuous functions $F_{n,\ell}(\cdot)$ such that for every real λ and $C < \infty$,*

$$\limsup_{\ell\to\infty}\limsup_{n\to\infty} \epsilon^2 \log E\left\{\exp\left[\frac{\lambda}{\epsilon^2}[F_{n,\ell}(\omega) - F(\omega)]\right]\right\} = 0$$

and

$$\lim_{\ell\to\infty}\lim_{n\to\infty} F_{n,\ell}(g) = F(g)$$

uniformly on $\{g : I_x(g) \le C\}$.

Proof. Since only the Stochastic Integral part is troublesome we need to approximate

$$G(\omega) = \int_0^T \langle a^{-1}(x(s)b(x(s)), dx(s)\rangle.$$

It is routine to see that

$$\begin{aligned} G_{n,\ell}(\omega) &= G_n(\omega) \text{ if } |\mathrm{G_n}(\omega)| \le \ell \\ &= 0 \text{ otherwise} \end{aligned}$$

with

$$G_n(\omega) = \int_0^T \langle a^{-1}\left(x\left(\frac{[ns]}{n}\right)\right) b\left(x\left(\frac{[ns]}{n}\right)\right), dx(s)\rangle$$

will do the trick by repeating the steps in the similar approximation we used earlier during the course of the proof of the case with $b \equiv 0$. We leave the details to the reader. □

Remark 2.11. The local uniformity in the starting point continues to hold without any problems.

3. The Exit Problem.

The exit problem deals with the following question. Suppose we are given a diffusion generator $\mathcal{L}_\epsilon$ of the form

$$\mathcal{L}_\epsilon = \frac{\epsilon^2}{2}\sum_{i,j} a_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_j b_j(x)\frac{\partial}{\partial x_j}$$

We suppose that $a_{i,j}$ and b_j are smooth and that a is uniformly elliptic, i.e. boundedly invertible. As $\epsilon \to 0$ the process will converge to the deterministic solution of the ODE

$$\frac{dx}{dt} = b(x(t)).$$

Let us assume that we have a bounded open set G with the property that for some point x_0 inside G, all trajectories of the above ODE starting from arbitrary points $x \in G$, remain forever with in G and converge to x_0 as $t \to \infty$. In other words x_0 is a stable equlibrium point for the flow and all trajectories starting from G remain in G and approach x_0. We can solve the Dirichlet Problem for $\mathcal{L}_\epsilon$, i.e solve the equation

$$\begin{aligned} \mathcal{L}_\epsilon U_\epsilon(x) &= 0 \text{ for } x \in G \\ U_\epsilon(y) &= f(y) \text{ for } y \in \partial G \end{aligned}$$

where f is the boundary value of U_ϵ specified on the boundary ∂G of G. We will assume that the domain G is regular and the function f is continuous. The exit problem is to understand the behavior of U_ϵ as $\epsilon \to 0$. If we denote by $P_{\epsilon,x}$ the measure on path space corresponding to the generator $\mathcal{L}_\epsilon$ that starts from the point x in G at time 0, then we know that U_ϵ has the representation in terms of the first exit place $x(\tau)$ where

$$\tau = \inf\{t : x(t) \notin G\}.$$

For any $\epsilon > 0$,

$$U_\epsilon(x) = E^{P_{\epsilon,x}}\big[f(x(\tau))\big].$$

It is tempting to let $\epsilon \to 0$ in the above representation. While $P_{\epsilon,x}$ converges weakly to P_x the degenerate measure at the solution to the ODE, since the ODE never exits, $\tau = +\infty$ a.e P_x for every $x \in G$. What happens to $x(\tau)$ is far from clear.

The solution to the exit problem is a good application of large deviations as originally carried out by Wentzell and Freidlin. Let us consider the rate function

$$I(x(\cdot)) = \frac{1}{2}\int_0^T \langle [x'(t) - b(x(t)], a^{-1}(x(t)\ [x'(t) - b(x(t))]\rangle$$

and define for any $x \in G$ and boundary pount $y \in \partial G$

$$V(x,y) = \inf_{0<T<\infty} \inf_{\substack{x(0)=x\\ x(T)=y}} I(x(\cdot))$$

and when $x = x_0$

$$R(y) = V(x_0, y)$$

We have the following theorem

Theorem 3.1. *Suppose for some (necessarily unique) point y_0 on ∂G.*

$$R(y) > R(y_0) \quad \text{for all } y \in \partial G, y \neq y_0$$

then,

$$\lim_{\epsilon \to 0} U_\epsilon(x) = f(y_0)$$

for all $x \in G$.

The proof will be based on the following lemma which will be proved at the end.

Lemma 3.2. *Let N, a eighborhood of y_0 on the boundary, be given. Then there are two neighborhoods B and H of x_0 with $x_0 \in B \subset \bar{B} \subset H \subset \bar{H} \subset G$ such that*

$$\lim_{\epsilon \to 0} \frac{\sup_{x \in \partial H} P_{\epsilon,x}[E_1]}{\inf_{x \in \partial H} P_{\epsilon,x}[E_2]} = 0$$

where E_1 and E_2 are defined by

$$E_1 = \{x(\cdot) : x(\cdot) \text{ exits } G \text{ in } N \text{ before visiting } \bar{B}\}$$
$$E_2 = \{x(\cdot) : x(\cdot) \text{ exits } G \text{ in } N^c \text{ before visiting } \bar{B}\}$$

Proof. Given the lemma the proof of the main theorem is easy. Suppose ϵ is small and the process starts from a point x_0 in G. It will follow the ODE and end up inside B before it exits from G. Then it will hang around x_0 inside H for a very long time. Because $\epsilon > 0$ it will exit from H at some boundary point of H. Then it may exit from G before entering back into B with a very small probability. This small probability is split between exiting in N and in N^c. Denoting by $p_{n,\epsilon}$ and $q_{n,\epsilon}$ the respective probabilities of exiting from N and N^c during the n-th trip back from H into B

$$\begin{aligned} P_{\epsilon,x}[x(\tau) \notin N] &= \sum_n q_{n,\epsilon} \\ &\leq C_\epsilon \; p_{n,\epsilon} \\ &= C_\epsilon P_{\epsilon,x}[x(\tau) \in N] \end{aligned}$$

where $C_\epsilon \to 0$ as $\epsilon \to 0$ according to the lemma. Clearly

$$P_{\epsilon,x}[x(\tau) \notin N] \leq \frac{C_\epsilon}{C_\epsilon + 1}$$

and $\to 0$ as $\epsilon \to 0$.

The proof of the theorem is complete. □

We now turn to the proof of the lemma.

Proof. We will break up the proof into two steps. First we show that in estimating various quantities we can limit ourselves to a finite time interval.

Step 1. Let us denote by τ_B and τ_G the hitting times of ∂B and ∂G respectively. We prove that

$$C(T) = \limsup_{\epsilon \to 0} \epsilon^2 \sup_{x \in \partial H} \log P_{\epsilon,x}[(\tau_B \geq T) \cap (\tau_G \geq T)]$$

tends to $-\infty$ as $T \to \infty$. By the large deviation property

$$C(T) \leq -\inf\{I(x(\cdot))\ ; x(\cdot) : x(0) \in \partial H, x(t) \in G \cap B^c \text{ for } 0 \leq t \leq T\}$$

If $C(T)$ does not go to $-\infty$ as $T \to \infty$ by the additivity property of the rate function , there will be a long interval $[T_1, T_2]$ over which

$$\int_{T_1}^{T_2} \langle [x'(t) - b(x(t)]a^{-1}(x(t)[x'(t) - b(x(t))]\rangle dt$$

will be small. This in the limit will give us a trajectory of the ODE that lies in $G \cap B^c$ for all times thereby contradicting stability.

Step 2. One can construct trajectories that connect points close to each other that have rate functions no larger than the distance between them. Just go from one point to the other at speed 1 on a straightline. Suppose $R(y_0) = \ell$ and $R(y) \geq \ell + \delta$ on N^c. We can pick a small ball H around x_0 such that for some $T_0 < \infty$ and for any $x \in \partial H$ the following statements are true: Every path from x that exits from G in the set N^c with in time T_0 has a rate function that exceeds $\ell + \frac{7}{8}\delta$. There is a path from x that exits G in N that has a rate function that is atmost $\ell + \frac{1}{8}\delta$. We can assume that T_0 is large eough that $C(T_0)$ of step 1 is smaller than $-2\ell - \delta$. It is easy to use the LDP and derive the following estimates uniformly for $x \in \partial H$.

$$P_{\epsilon,x}[\tau_G < \tau_B, x(\tau_G) \in N^c] \leq \exp[-\frac{1}{\epsilon^2}(\ell + \frac{7}{8}\delta)] + \exp[-\frac{1}{\epsilon^2}(2\ell + \delta)]$$

and

$$P_{\epsilon,x}[\tau_G < \tau_B, x(\tau_G) \in N] \geq \exp[-\frac{1}{\epsilon^2}(\ell + \frac{1}{8}\delta)] - \exp[-\frac{1}{\epsilon^2}(2\ell + \delta)]$$

The lemma follows from this. $\square$

Remark 3.3. In the special case when $A(x) = I$ and $b(x) = -\frac{1}{2}\nabla F$ for some F we can carry out the calculation of the rate function more or less explicitly.

$$\begin{aligned}
\int_0^T \|x'(t) + \frac{1}{2}(\nabla F)(x(t))\|^2 dt &= \int_0^T \|x'(t) - \frac{1}{2}(\nabla F)(x(t))\|^2 dt + 2\int_0^T \langle(\nabla F)(x(t)), x'(t)\rangle dt \\
&\geq 2\int_0^T \langle(\nabla F)(x(t)), x'(t)\rangle dt \\
&= 2[F(x(T)) - F(x(0))].
\end{aligned}$$

Therefore

$$V(x, y) \geq [F(y) - F(x)]$$

On the other hand if $x(t)$ sojves the ODE

$$x'(t) = (\nabla F)(x(t)$$

with $x(0) = y$ and $x(T) = x$, then the trajectory

$$y(t) = x(T - t) \text{ for } 0 \leq t \leq T$$

connects x to y and its rate function is

$$\begin{aligned}
I[y(\cdot)] &= \frac{1}{2}\int_0^T \|y'(t) + \frac{1}{2}(\nabla F)(y(t))\|^2 dt \\
&= \frac{1}{2}\int_0^T \|x'(t) - \frac{1}{2}(\nabla F)(x(t))\|^2 dt \\
&= \frac{1}{2}\int_0^T \|x'(t) + \frac{1}{2}(\nabla F)(x(t))\|^2 dt - \int_0^T \langle(\nabla F)(x(t)), x'(t)\rangle dt \\
&= [F(x(0)) - F(x(T))] \\
&= [F(y) - F(x)].
\end{aligned}$$

It follows then that for any y in the domain of attraction of x_0

$$R(y) = V(x_0, y) = F(y) - F(x_0)$$

and the exit point is obtained by minimizing the potential $F(y)$ on ∂G.