COURANT

27

LECTURE NOTES

S. R. S. VARADHAN

Large Deviations

American Mathematical Society
Courant Institute of Mathematical Sciences

AMS
AMERICAN MATHEMATICAL SOCIETY

https://doi.org/10.1090//cln/027

Large Deviations

S. R. S. Varadhan
Courant Institute of Mathematical Sciences

27 Large Deviations

Courant Institute of Mathematical Sciences
New York University
New York, New York

American Mathematical Society
Providence, Rhode Island

2010 *Mathematics Subject Classification*. Primary 60-02, 60F10.

For additional information and updates on this book, visit
www.ams.org/bookpages/cln-27

Library of Congress Cataloging-in-Publication Data

Names: Varadhan, S. R. S. | Courant Institute of Mathematical Sciences.
Title: Large deviations / S.R.S. Varadhan.
Description: Providence, RI : American Mathematical Society, [2016] | Series: Courant lecture notes in mathematics ; 27 | "Courant Institute of Mathematical Sciences." | Includes bibliographical references.
Identifiers: LCCN 2016036484 | ISBN 9780821840863 (alk. paper)
Subjects: LCSH: Large deviations. | AMS: Probability theory and stochastic processes – Research exposition (monographs, survey articles). msc | Probability theory and stochastic processes – Limit theorems – Large deviations. msc
Classification: LCC QA273.67 .V365 2016 | DDC 519.2–dc23 LC record available at https://lccn.loc.gov/2016036484

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1 21 20 19 18 17 16

Contents

Preface

These notes are based on a graduate course on large deviations given at the Courant Institute in 2012. While a version of these notes appeared on the web at that time, it took considerable time for me to prepare a revision. The lectures focused on three sets of examples as do these notes:

- diffusions with small noise and the exit problem,
- large time behavior of Markov processes and their connection to the Feynman-Kac formula and the related large-deviation behavior of the number of distinct sites visited by a random walk,
- interacting particle systems, their scaling limits, and large deviations from their expected limits.

We will look at simple exclusion processes in d dimensions. Some of the material is quite intricate and towards the end instead of providing complete proofs, we will give the ideas behind the proofs and provide references.

I want to thank the students who attended the course and motivated me to write these notes. It took much longer than I expected for me to finish the revision and I want to thank the AMS for waiting patiently. I want to thank Ina Mette, whose regular but gentle reminders prevented the delay from being even longer.

https://doi.org/10.1090//cln/027/01

CHAPTER 1

Introduction

1.1. Outline

We will examine the theory of large deviations through three concrete examples. We will work them out in some detail and in the process develop the subject.

The first example is the exit problem. Let $G \subset \mathbb{R}^d$ be a bounded, open domain with smooth boundary ∂G. We consider the solution $u = u_\epsilon$ of the Dirichlet problem

$$\frac{\epsilon}{2}\Delta u + b(x) \cdot \nabla u = 0, \quad x \in G,$$

with boundary condition $u = f$ on ∂G. The vector field b will be of the form $b = -\nabla V$ for some smooth function V. As $\epsilon \to 0$, the limiting behavior of the solution u_ϵ will depend on the behavior of the solutions of the ODE

$$\frac{dx}{dt} = b(x(t)). \tag{1.1}$$

The solution $u_\epsilon(x)$ has the representation

$$u_\epsilon(x) = E_x[f(x(\tau_G))]$$

in terms of the expectation with respect to the distribution $P_{\epsilon,x}$ of the solution $x(t) = x_\epsilon(t)$ of the stochastic integral equation

$$x(t) = x + \int_0^t b(x(s))ds + \sqrt{\epsilon}\beta(t),$$

where $\beta(t)$ is the standard Brownian motion in d dimensions, $\tau_G = \inf\{t : x(t) \notin G\}$ is the exit time from G, and $x(\tau_G)$ is the exit place on ∂G. One expects the limit $u(x) = \lim_{\epsilon\to 0} u_\epsilon(x)$ to exist and be given by

$$u(x) = f(x(\tau_G))$$

where $x(t)$ is the solution of the ODE (1.1).

The difficult case is when the solution of the ODE does not exit from G and therefore $\tau(G) = \infty$. Then large-deviation theory can provide the answer. Assuming that there is a unique stable equilibrium point inside G and that all trajectories starting from $x \in G$ converge to it without leaving G, one can show that

$$\lim_{\epsilon\to 0} u_\epsilon(x) = f(y)$$

provided the minimum of $V(\cdot)$ on the boundary ∂G is attained uniquely at $y \in \partial G$; i.e., for all $y' \in \partial G$ and $y' \neq y$, we have $V(y') > V(y)$.

The second example is about the simple random walk in d dimensions. Let $e_1, \dots, e_d$ be the unit vectors in the d coordinate directions of $\mathbb{Z}^d$ and let $X_1, \dots, X_n, \dots$ be a sequence of independent identically distributed random variables with $P[X_i = \pm e_j] = \frac{1}{2d}$ for $j = 1, \dots, d$. We denote by $S_n = X_1 + \cdots + X_n$ the resulting random walk and by D_n

the range of $S_1, \dots, S_n$. Then $|D_n|$ is the number of distinct sites visited by the random walk. The question is the behavior of

$$E[e^{-\nu|D_n|}]$$

for large n. Contribution comes mainly from paths that do not visit too many sites. We can insist that the random walk be confined to a ball of radius $R = R(n)$. Then the number of sites visited is at most the number of lattice points inside the ball, which is approximately $v(d)R^d$ for large R where $v(d)$ is the volume of the unit ball in $\mathbb{R}^d$. On the other hand, confining a random walk to the region for n steps has exponentially small probability $p(n) \simeq \exp[-n\lambda_d(R)] = \exp[-n\frac{\lambda_d}{R^2}]$. Here $-\lambda_d$ is the ground state eigenvalue of the Laplace operator

$$\frac{1}{2d}\sum_{i=1}^{d}\frac{\partial^2}{\partial x_i^2}$$

in the unit ball of $\mathbb{R}^d$ with the Dirichlet boundary condition. The contribution from these paths is at least

$$\exp\left[-\nu v(d)R^d - n\frac{\lambda(d)}{R^2}\right],$$

and $R = R(n)$ can be chosen to maximize this contribution. One can fashion a proof that establishes this as a lower bound. But to show that the optimal lower bound obtained in this manner is actually a true upper bound requires a theory.

The third example that we will consider is the symmetric simple exclusion process. On the periodic d-dimensional integer lattice $\mathbf{Z}_N^d$ of size N^d we have $k(N) = \rho N^d$ particles (with at most one particle per site) doing simple random walks independently with rate 1. However jumps to occupied sites are forbidden. The Markov process is defined through the generator

$$(\mathcal{A}_N u)(x_1, \dots, x_{k(N)}) = \\ \frac{1}{2d}\sum_{i=1}^{k(N)}\sum_{e}[1-\eta(x_i+e)][u(x_1,\dots,x_i+e,\dots,x_{k(N)}) - u(x_1,\dots,x_{k(N)})]$$

acting on functions u defined on $(\mathbf{Z}_N^d)^{k(N)}$. Here e runs over the units in the $2d$ directions $\{\pm e_j\}$ and $\eta(x) = \sum_i \mathbb{1}_{x_i = x}$ is the particle count at x, which is either 0 or 1. We do a diffusive rescaling of space and time and consider the random measure γ_N on the path space $D[[0,T];\mathcal{T}^d]$,

$$\gamma_N = \frac{1}{N^d}\sum_{1\le i\le k(N)} \delta_{\frac{x_i(N^2\cdot)}{N}}.$$

We want to study the behavior of γ_n as $N \to \infty$. The theory of large deviations is needed even to prove a law of large numbers for γ_N.

1.2. Supplementary Material

Large-deviation theorems in some generality were first established by Cramér in [**2**]. He considered deviations from the law of large numbers for sums of independent identically distributed random variables and showed that the rate function is given by the convex conjugate of the logarithm of the moment-generating function of the underlying common distribution. The subject has evolved considerably over time, and several texts are now available offering different perspectives. The exit problem was studied by Wentzell and

Freidlin in their work [**12**]. They go on to study in [**13**] the long-time behavior of small random perturbations of dynamical systems when there are several equilibrium points.

The problem of counting the number of distinct sites comes up in the discussion of a random walk on $\mathbb{Z}^d$ in the presence of randomly located traps. The estimation of the probability of avoiding traps for a long time reduces to the calculation described in the second example. This problem was proposed by Mark Kac [**16**], along with a similar problem for a Brownian path avoiding traps in $\mathbb{R}^d$. The traps are balls of some fixed radius δ with their centers located randomly as a Poisson point process of intensity ρ. Now the role of the number of distinct sites of the random walk is replaced by the volume $|\bigcup_{0\le s\le t} B(x(s),\delta)|$ of the "Wiener sausage," i.e., the δ-neighborhood of the range of the Brownian path $x(\cdot)$ in $[0,t]$.

The use of large-deviation techniques in the study of hydrodynamic scaling limits began with the work of Guo, Papanicolaou, and Varadhan in [**15**], and the results presented here started with the work of Kipnis, Olla, and Varadhan in [**18**] followed by the study of nongradient systems in [**31**], the Ph.D. thesis of Quastel [**19**], and the work of Quastel, Rezakhanlou, and Varadhan in [**20, 21, 23**].

https://doi.org/10.1090//cln/027/02

CHAPTER 2

Basic Formulation

2.1. What Are Large Deviations?

In large deviation theory, the basic object of study is a family $\{P_n\}$ of probability distributions defined on the Borel σ-field of a complete, separable metric space X. Typically, P_n will converge weakly to δ_{x_0}, the degenerate distribution with the entire mass at some point $x_0 \in X$. Probabilities for sets away from x_0 will decay exponentially in n and we expect the limit

$$\lim_{n\to\infty} \frac{1}{n} \log P_n(A) = -c(A) \tag{2.1}$$

to exist for a large class of sets A. The goal is to determine $c(A)$. Because

$$\max\{P_n(A), P_n(B)\} \le P_n(A \cup B) \le 2 \max\{P_n(A), P_n(B)\},$$

it is easy to see that

$$c(A \cup B) = \min\{c(A), c(B)\},$$

and one can expect to have a nonnegative function $I(x)$ such that

$$c(A) = \inf_{x\in A} I(x).$$

One cannot expect the limit (2.1) to hold for all Borel sets. For example, for the single-point set $\{x\}$ consisting of just x, $P[\{x\}]$ can often be 0, while $I(x)$ is finite. The starting point of the theory is the following definition:

DEFINITION 2.1. A sequence $\{P_n\}$ satisfies the large deviation principle (LDP) with rate function $I(x)$ provided

- $I(x) : X \to [0,\infty]$ is lower semicontinuous.
- The sets $K_\ell = \{x : I(x) \le \ell\}$ are compact in X.
- For every closed set $C \subset X$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n[C] \le -\inf_{x\in C} I(x). \tag{2.2}$$

- For every open set $G \subset X$, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n[G] \ge -\inf_{x\in G} I(x). \tag{2.3}$$

We have an alternate definition as well.

DEFINITION 2.2. The following is an alternate set of conditions equivalent to Definition 2.1 above.

- **Exponential tightness property**. Given any $\ell < \infty$, there exists a compact set K_ℓ such that, for any closed set C disjoint from K_ℓ; i.e., $C \cap K_\ell = \varnothing$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n[C] \le -\ell.$$

- **Local upper bound.**

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P_n[B(x,\delta)] \le -I(x).$$

- **Local lower bound.**

$$\liminf_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log P_n[B(x,\delta)] \ge -I(x).$$

Here $B(x,\delta)$ is the ball around x of radius δ. We can take it to be either the open ball $\{y : d(x,y) < \delta\}$ or the closed ball $\{y : d(x,y) \le \delta\}$.

LEMMA 2.3. *The two Definitions* 2.1 *and* 2.2 *are equivalent.*

PROOF. We first prove that Definition 2.1 implies Definition 2.2. To prove the exponential tightness property the choice of $K_\ell = \{x : I(x) \le \ell\}$ works. For any closed set C disjoint from K_ℓ, $I(x) \ge \ell$ on C and

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n[C] \le -\inf_{x\in C} I(x) \le -\ell.$$

For any $\delta > 0$,

$$\liminf_{n\to\infty}\frac{1}{n}\log P_n[B(x,\delta)] \ge -\inf_{x\in B(x,\delta)} I(y) \ge -I(x),$$

and for the closed ball $B(x,\delta)$,

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n[B(x,\delta)] \le -\inf_{y\in B(x,\delta)} I(y),$$

and, by lower semicontinuity,

$$\lim_{\delta\to 0}\inf_{y\in B(x,\delta)} I(y) = I(x).$$

The other direction is just as easy. First, we note that if G is open, for any $x \in G$, there is a ball $B(x,\delta_x) \subset G$. Therefore, for every $x \in G$,

$$\liminf_{n\to\infty}\frac{1}{n}\log P_n[G] \ge \liminf_{n\to\infty}\frac{1}{n}\log P_n\big[B(x,\delta_x)\big] \ge -I(x),$$

proving (2.3). We next show that exponential tightness together with the lower bound implies compact level sets for $I(\cdot)$. Let $\ell < \infty$ be arbitrary, and the compact set K_ℓ be such that for any C disjoint from it

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n[C] \le -\ell.$$

In particular, if $x \notin K_\ell$, since $B(x,\delta) \in K_\ell^c$ for some δ, we must have $I(x) \ge \ell$. Therefore,

$$\{x : I(x) \le \ell\} \subset K_\ell.$$

Next, we prove that $I(x)$ is lower semicontinuous. Suppose $I(x) \ge \ell$. Then, given any $\epsilon > 0$, there is a $\delta > 0$ such that

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n[B(x,\delta)] \le -\ell + \epsilon.$$

If $x_k \to x$, eventually $B(x_k,\delta/2) \subset B(x,\delta)$ and

$$-I(x_k) \le \liminf_{n\to\infty}\frac{1}{n}\log P_n\Big[B\Big(x_k,\frac{\delta}{2}\Big)\Big] \le \limsup_{n\to\infty}\frac{1}{n}\log P_n[B(x,\delta)] \le -\ell + \epsilon.$$

Finally, we come to the upper bound (2.2). Let C be any closed set, and $\ell = \inf_{x\in C} I(x)$. Take K_ℓ to be the compact set provided by the exponential tightness. Let $\epsilon > 0$ be given. For any $x \in C$, there is a ball $B(x, \delta_x)$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n[B(x,\delta)] \le -\ell + \epsilon.$$

These balls cover C and therefore $C \cap K_\ell$, which is compact. Let us choose a finite subcover, with their union G covering $C \cap K_\ell$. Then

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n[G] \le -\ell + \epsilon.$$

We have

$$P_n[C] \le P_n[G] + P_n[C \cap G^{\mathrm{c}}],$$

and, because $C \cap G^{\mathrm{c}} \subset K_\ell^{\mathrm{c}}$ and is closed, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n[C \cap G^{\mathrm{c}}] \le -\ell.$$

This concludes the proof. $\square$

Although it appears to be a bit stronger, the following version of the exponential tightness property is a consequence of the Definitions 2.1 or 2.2.

LEMMA 2.4. *Given any ℓ, there is compact set K_ℓ such that for all $n \ge 1$,*

$$P_n\big[K_\ell^{\mathrm{c}}\big] \le e^{-n\ell}. \tag{2.4}$$

PROOF. Let ℓ be given. There is a compact set $F = F_\ell$ such that, for any positive integer k, if $G_k = \bigcup_{x\in F} B(x, 1/k)$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n\big[G_k^{\mathrm{c}}\big] \le -\ell - 1.$$

For every k, we can find n_k such that for $n \ge n_k$,

$$P_n\big[G_k^{\mathrm{c}}\big] \le 2^{-k} e^{-n\ell}.$$

We can find a compact set C_k such that $P_n\big[C_k^{\mathrm{c}}\big] \le 2^{-k} e^{-n\ell}$, for $n \le n_k$. We consider

$$D_\ell = \bigcap_k [C_k \cup G_k].$$

Since $K_\ell = \overline{D}_\ell$ is compact and

$$D_\ell^{\mathrm{c}} = \bigcup_k \big(C_k^{\mathrm{c}} \cap G_k^{\mathrm{c}}\big)$$

we obtain

$$P_n\big[K_\ell^{\mathrm{c}}\big] \le P_n[D^{\mathrm{c}}] \le \sum_k P_n\big[C_k^{\mathrm{c}} \cap G_k^{\mathrm{c}}\big] \le \sum_k 2^{-k} e^{-n\ell} = e^{-n\ell}.$$

Finally, we need to check that $\overline{D}$ is compact. Let x_r be a sequence from D. If an infinite number of them are in some C_k, then since each C_k is compact, we can have a convergent subsequence. Otherwise, there is r_k such that $x_r \in G_k$ for $r \ge r_k$. The sequence x_r can then be shadowed by $y_r \in K$ with $d(x_r, y_r) \le 1/k$ for $r \ge r_k$. Since K is compact, there is again a convergent subsequence from $\{y_r\}$ and therefore from $\{x_r\}$ as well. This proves the compactness of $\overline{D}$. $\square$

2.2. Evaluation of Integrals

The following result is an easy consequence of the definition of LDP. If we have functions that grow exponentially at different rates at different points and probability that decays at different rates around different points, then the rate of exponential growth of the integral is easy to obtain.

THEOREM 2.5. *Let $\{P_n\}$ satisfy LDP with rate function $I(x)$. Let $F(x)$ be a bounded continuous function on X. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_X \exp[nF(x)]dP_n(x) = \sup_x [F(x) - I(x)].$$

PROOF. The lower bound is easy. For any x and any ϵ, there is a $\delta > 0$ such that $F(y) \geq F(x) - \epsilon$ if $y \in B(x, \delta)$. Then,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log \int_X \exp[nF(y)]dP_n(y) &\geq \lim_{n\to\infty} \frac{1}{n} \log \int_{B(x,\delta)} \exp[nF(y)]dP_n(y) \\
&\geq F(x) - \epsilon + \liminf_{n\to\infty} \frac{1}{n} \log P_n[B(x,\delta)] \\
&\geq F(x) - \epsilon - I(x).
\end{aligned}$$

Since $\epsilon > 0$ and $x \in X$ are arbitrary, the lower bound follows. The upper bound is not any harder. We do an approximation with simple functions. Since $F(x)$ is bounded, given any $\epsilon > 0$, let us cover the space X with a finite collection of closed subsets $\{C_j\}$ where $C_j = \{x : a_i \leq F(x) \leq a_{i+1}\}$ and $|a_{j+1} - a_j| \leq \epsilon$. Then

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \int_X \exp[nF(x)]dP &\leq \limsup_{n\to\infty} \frac{1}{n} \log \sum_j \int_{C_j} \exp[nF(x)]dP \\
&\leq \sup_j \limsup_{n\to\infty} \frac{1}{n} \log \int_{C_j} \exp[nF(x)]dP \\
&\leq \sup_j \left[a_{j+1} + \limsup_{n\to\infty} \frac{1}{n} \log P_n[C_j] \right] \\
&\leq \sup_j \left[a_{j+1} - \inf_{x\in C_j} I(x) \right] \\
&\leq \sup_j \sup_{x\in C_j} [F(x) - I(x) + \epsilon] \\
&= \sup_{x\in X} [F(x) - I(x)] + \epsilon. \qquad \square
\end{aligned}$$

Sometimes F is only upper semicontinuous and that is sufficient to provide the upper bound. The proof is slightly more involved.

THEOREM 2.6. *Let F be an upper semicontinuous function on X that is bounded above, and $\{P_n\}$ a sequence that satisfies LDP with rate function $I(x)$. Then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_X \exp[nF(x)]dP_n(x) \leq \sup_{x\in X} [F(x) - I(x)]. \tag{2.5}$$

PROOF. We cannot use uniform approximation by simple functions, because now the sets $\{x : a \le F(x) \le b\}$ may not be closed. In fact, $F(x)$ can be allowed to take the value $-\infty$. It can even be that $\sup_x[F(x) - I(x)] = -\infty$. If for some x either $F(x) = -\infty$ or $I(x) = \infty$, by upper semicontinuity of F and the lower semicontinuity of I, given any $\ell < \infty$, we can find $\delta = \delta_x > 0$ such that

$$\sup_{y \in B(x,\delta)} F(y) - \inf_{y \in B(x,\delta)} I(y) \le -\ell,$$

implying

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{B(x,\delta)} \exp[nF(y)]dP_n(y) \le -\ell.$$

If $F(x) - I(x) = a$ is finite, for any $\epsilon > 0$ we can find $\delta = \delta_x$ such that

$$\sup_{y \in B(x,\delta)} F(y) - \inf_{y \in B(x,\delta)} I(y) \le a + \epsilon,$$

implying

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{B(x,\delta)} \exp[nF(y)]dP_n(y) \le a + \epsilon.$$

By Lemma 2.4, given any ℓ, we can find a compact set K_ℓ such that

$$P_n[K_\ell] \le e^{-n\ell},$$

implying

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{K_\ell^c} \exp[nF(y)]dP_n(y) \le M - \ell.$$

We can choose a finite subset from $\{B(x, \delta_x)\}$ that covers K_ℓ and conclude that

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_X \exp[nF(y)]dP_n(y) \le \max\left\{\sup_x[F(x) - I(x)] + \epsilon, -\ell, M - \ell\right\}$$

We can now let $\epsilon \to 0$ and $\ell \to \infty$, and complete the proof. $\square$

2.3. Contraction Principle

When we have the joint distribution of two components but are interested only in the marginal distribution of one component, we project by integrating out the other variable. In large deviation theory, the integration is replaced by minimization.

THEOREM 2.7. *If $\{P_n\}$ satisfies LDP on X with rate function $I(x)$ and $f : X \to Y$ is a continuous map into another complete, separable metric space, then $Q_n = P_n f^{-1}$ satisfies an LDP on Y with a rate function $J(y)$ given by*

$$J(y) = \inf_{x: f(x)=y} I(x).$$

PROOF. We just note that if $A \subset Y$ is a Borel set, $Q_n(A) = P_n[f^{-1}A]$. If A is closed, so is $f^{-1}A \subset X$,

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n(A) = \limsup_{n\to\infty} \frac{1}{n} \log P_n[f^{-1}A] = -\inf_{x \in f^{-1}A} I(x) = -\inf_{y \in A} J(y);$$

and, if A is open,

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(A) = \liminf_{n\to\infty} \frac{1}{n} \log P_n[f^{-1}A] = -\inf_{x\in f^{-1}A} I(x) = -\inf_{y\in A} J(y).$$

It is straightforward to show that the level sets $A_\ell = \{y : J(y) \le \ell\}$ are compact (and closed), implying lower semicontinuity of $J(\cdot)$. They are the images under f of the sets $B_\ell = \{x : I(x) \le \ell\}$ that are compact in X. □

The following slight modification of this result is sometimes needed.

THEOREM 2.8. *Let $f_n(x)$ be a sequence of continuous maps from $X \to Y$ converging to $f(x)$ uniformly on compact sets and let $Q_n = P_n f_n^{-1}$. Then again, $\{Q_n\}$ satisfies an LDP on Y with a rate function $J(y)$ given by*

$$J(y) = \inf_{x:f(x)=y} I(x).$$

PROOF. The proof has to be a bit carefully done. The lower bound is not hard. Let $y \in Y$ with $J(y) = \ell < \infty$. Because $I(x)$ is lower semicontinuous and the sets $\{x : I(x) \le \ell\}$ are compact, the infimum in $J(y) = \inf_{x:f(x)=y} I(x)$ is attained; i.e., there is an $x \in X$ with $f(x) = y$ and $J(x) = \ell$. Let $B(y, \epsilon)$ be a small ball around y. We need to get a lower bound on $P_n[f_n^{-1} B(y, \epsilon)]$. If we show that for sufficiently large n, $f_n^{-1} B(y, \epsilon)$ contains $B(x, \delta)$ for some $\delta > 0$, then $Q_n[B(y, \epsilon)] = P_n[f_n^{-1} B(y, \epsilon)] \ge P_n[B(x, \delta)]$ and

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n[B(y, \epsilon)] \ge \liminf_{n\to\infty} \frac{1}{n} \log P_n[B(x, \delta)] \ge -I(x) = -J(y).$$

Suppose it is not true. Then, there will be a (sub)sequence such that $x_n \to x$ but $f_n(x_n) \notin B(y, \epsilon)$, contradicting the uniform convergence of $f_n \to f$ on compact subsets of X.

For proving the upper bound, pick a large ℓ and use Lemma 2.4 to get K with $P_n(K^c) \le e^{-n\ell}$. It is enough to estimate $P_n[f_n^{-1} C \cap K]$. For that, it is sufficient to show that for any $\delta > 0$, for sufficiently large n,

$$f_n^{-1} C \cap K \subset \bigcup_{x\in f^{-1}C} B(x, \delta).$$

If not, there is a sequence $x_n \in K$ with $f_n(x_n) \in C$, but $d(f(x_n), C) \ge \delta$. This contradicts the uniform convergence of f_n to f on K. □

2.4. Simple Examples and Remarks

(1) The easiest example is to take the average of n independent and normally distributed random variables with mean 0 and variance 1. By the law of large numbers, their average will be close to the distribution with all its mass at 0. The actual distribution P_n is Gaussian with mean 0 and variance $1/n$ given by

$$P_n(A) = \sqrt{\frac{n}{2\pi}} \int_A \exp\left[-\frac{nx^2}{2}\right] dx.$$

It is not hard to check that LDP holds on $\mathbb{R}$ with $I(x) = x^2/2$.

(2) If we toss a fair coin n times, the probability of r heads is $\binom{n}{r} 2^{-n}$, and using Stirling's approximation, $\log n! = -n + n \log n + o(n)$, we see again that LDP holds for the distribution of r/n on $[0, 1]$ with the rate function

$$I(x) = x \log(2x) + (1 - x) \log(2(1 - x)).$$

(3) More generally, we can look at distributing n objects randomly and independently into k pigeonholes with the probability of any object being placed in the cell i being π_i. If $\{f_i\}$ are the numbers ending up in various cells, their joint distribution is a multinomial

$$P_n[f_1, \ldots, p_k] = \frac{n!}{f_1! f_2! \cdots f_k!} \pi_1^{f_1} \pi_2^{f_2} \cdots \pi_k^{f_k}.$$

Again using Stirling's approximation, we see that LDP holds for the ratios $\mathbf{x} = \{x_i\} = f_i/n$ with rate function

$$I(\mathbf{x}) = \sum_i x_i \log \frac{x_i}{\pi_i}.$$

In some sense, this function, called the relative entropy or Kullback-Leibler information of one distribution $\{x_i\}$ with respect another $\{\pi_i\}$, controls large deviations.

(4) Cramér's theory for sums of independent identically distributed random variables assumes that they come from a distribution α, and that it has a moment generating function

$$M(\theta) = \int e^{\theta x} \, d\alpha(x) < \infty,$$

for all θ. It is convenient to assume that the support of α extends to $\pm\infty$ in both directions. Then, $\log M(\theta)$ is a convex function whose derivative takes all the values on $\mathbb{R}$. Then, the distribution of the mean of n independent observations from α satisfies an LDP with rate function

$$I(x) = \sup_\theta [\theta \, x - \log M(\theta)].$$

The proof is relatively easy. For the upper bound, if $a > m = \int x \, d\alpha(x)$, and if

$$Z_n = \frac{X_1 + \cdots + X_n}{n},$$

then

$$\begin{aligned} P[Z_n \geq a] \leq e^{-n\theta a} E e^{n\theta Z_n} &= e^{-n\theta a} \left[\int e^{\theta x} \, d\alpha(x) \right]^n \\ &= \exp[-n[\theta a - \log M(\theta)]]. \end{aligned}$$

This yields the upper bound. The decay rate for $P_n(A)$ depends only on the points a^+, a^- in A that are closest to the mean m on either side. Moreover, the convex function $I(x)$ has its minimum value of 0 at m and is monotone on either side of it. This reduces the proof of the upper bound to the case of half lines that avoid m.

The lower bound uses tilting. Let us replace α by β defined as $d\beta = [M(\theta)]^{-1} e^{\theta x} \, d\alpha$ with θ picked so that it maximizes $a\theta - \log M(\theta)$; i.e., $a = M'(\theta)/M(\theta) = \int x \, d\beta(x)$. If we denote by Q_n the distribution of the sample mean under β, we have

$$P_n[(a-\delta, a+\delta)] = \int_{(a-\delta, a+\delta)} e^{-nx} [M(\theta)]^n \, dQ_n.$$

Since $Q_n[(a-\delta, a+\delta)] \to 1$ as $n \to \infty$, it is clear that

$$\lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} \log P_n[(a-\delta, a+\delta)] = -a\theta + \log M(\theta) = -I(a).$$

(5) There is a multivariate analogue of Cramér's theorem due to Chernoff. Basically, the moment generating function

$$M(\theta) = \int_{\mathbb{R}^d} \exp[\langle \theta, x\rangle] d\alpha(x) < \infty$$

exists on $\mathbb{R}^d$, and its convex conjugate,

$$I(x) = \sup_{\theta \in \mathbb{R}^d} [\langle \theta, x\rangle - \log M(\theta)],$$

provides the rate function for the LDP of the average of n independent observations from α.

(6) The multinomial is a special case where α on $\mathbb{R}^d$ is the discrete distribution with mass π_i at the unit vector e_i in the direction of the i^{th} coordinate.

$$\log M(\theta) = \log \sum \pi_i e^{\theta_i}.$$

Then, $I(x) = \sum_i x_i \log(x_i/\pi_i)$.

(7) We can look at the empirical distribution $r_n = \frac{1}{n} \sum_i \delta_{X_i}$ of n independent observations taking values in a Polish space X having a common distribution α. The distribution P_n of r_n on the space $\mathcal{M}(X)$ of probability measures on X will satisfy an LDP with rate function given by the relative entropy

$$I(\beta) = h(\beta; \alpha) = \int_X \frac{d\beta}{d\alpha}(x) \log \frac{d\beta}{d\alpha}(x) d\alpha(x),$$

which is defined to be infinite if the integral fails to exist; i.e., unless $\beta \ll \alpha$ and $\log \frac{d\beta}{d\alpha} \in L_1(\beta)$. This was proved originally by Sanov.

(8) Finally, one can recover the Cramér rate function as

$$I(y) = \inf_{\beta : \int x \, d\beta = m} h(\beta; \alpha),$$

and this allows us to view Cramér's result as a contraction of Sanov's theorem.

(9) Or, you can recover Sanov's result from Cramér's. Replace $\mathbb{R}^d$ by the convex set $\mathcal{M}(X)$ of all probability measures on X, a subset of the infinite-dimensional vector space of all measures on X. For the distribution of a random measure on $\mathcal{M}(X)$, we lift $x \to \delta_x$, and we will have $\hat{\alpha}$ on $\mathcal{M}(X)$. The dual is the space of continuous functions $V : X \to \mathbb{R}$, and the rate function takes the form

$$\sup_V \left[\int V(x) d\beta(x) - \log \int e^{V(x)} d\alpha(x) \right] = h(\beta; \alpha).$$

(10) Convex duality plays a role in the theory. We are used to the duality between L_p and L_q spaces. If $1/p + 1/q = 1$, then the identity

$$\frac{x^p}{p} = \sup_y \left[xy - \frac{y^q}{q} \right]$$

leads to duality between L_p and L_q. Similarly, the duality relation

$$\sup_y [xy - e^y] = x \log x - x$$

can be used to obtain the duality relation

$$\log \int e^{V(y)} \, d\alpha = \sup_{\substack{f \geq 0 \\ \|f\|_1 = 1}} \left[\int f(y) V(y) d\alpha - \int f(y) \log f(y) d\alpha \right].$$

https://doi.org/10.1090//cln/027/03

CHAPTER 3

Small Noise

3.1. The Exit Problem

Let $\mathcal{L}$ be the second-order elliptic operator (with smooth coefficients)

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

on $\mathbb{R}^d$. If $G \subset \mathbb{R}^d$ is a bounded, connected open set with smooth boundary ∂G, the solution to the Dirichlet problem

$$\begin{aligned} (\mathcal{L}u)(x) &= 0 \qquad \text{for } x \in G, \\ u(y) &= f(y) \quad \text{for } y \in \partial G, \end{aligned} \tag{3.1}$$

can be represented as

$$u(x) = E_x[f(x(\tau)] \tag{3.2}$$

where E_x is expectation with respect to the diffusion process P_x corresponding to $\mathcal{L}$ starting from $x \in G$, τ is the exit time from the region G, and $x(\tau)$ is the exit place on the boundary ∂G of G. If $\mathcal{L}$ is elliptic and G is bounded, then τ is finite almost surely, and in fact its distribution has an exponentially decaying tail under every P_x. If G has a regular boundary (exterior cone condition is sufficient), then the function $u(x)$ defined by (3.2) solves (3.1) and $u(x) \to f(y)$ as $x \in G \to y \in \partial G$.

We are interested in the situation where $\mathcal{L}$ depends on a parameter ϵ that is small:

$$(\mathcal{L}_\epsilon u)(x) = \frac{\epsilon}{2}\sum_{i,j=1}^{d} a_{i,j}(x)\frac{\partial^2 u}{\partial x_i \partial x_j}(x) + \sum_{j=1}^{d} b_j(x)\frac{\partial u}{\partial x_j}(x).$$

As $\epsilon \to 0$, $\mathcal{L}_\epsilon$ degenerates to a first-order operator; i.e., a vector field

$$(Xu)(x) = \sum_{j=1}^{d} b_j(x)\frac{\partial u}{\partial x_j}(x). \tag{3.3}$$

The behavior of the solution $u_\epsilon(x)$ of

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

will depend on the behavior of the solution of the ODE (3.3),

$$\frac{dx(t)}{dt} = b(x(t)), \quad x(0) = x. \tag{3.4}$$

If $x(t)$ exits cleanly from G at a point $y_0 \in \partial G$, then $u_\epsilon(x) \to f(y_0)$. If the trajectory $x(t)$ touches the boundary and reenters G one or more times before exiting from the closure $\overline{G}$, it is problematic. If the trajectory does not ever exit G, then we have a real problem.

We will concentrate on the following situation. The operator $\mathcal{L}_\epsilon$ is given by

$$\mathcal{L}_\epsilon u = \frac{\epsilon}{2}\Delta u + Xu.$$

We will assume that every solution of the corresponding ODE (3.3) with $x \in \overline{G}$ stays in G forever, and as $t \to \infty$, they all converge to a limit x_0 that is the unique equilibrium point in G; i.e., the only point with $b(x_0) = 0$. In other words, x_0 is the unique globally stable equilibrium in G, and every solution converges to it without leaving G. Let $P_{\epsilon,x}$ be the distribution of the solution to the stochastic differential equation

$$x_\epsilon(t) = x + \int_0^t b(x_\epsilon(s))ds + \sqrt{\epsilon}\beta(t) \tag{3.5}$$

where $\beta(t)$ is the d-dimensional Brownian motion. It is clear that while the paths will exit from G almost surely under $P_{\epsilon,x}$ as $\epsilon \to 0$. It will take an increasingly longer time and in the limit there will be no exit. The behavior of the solution u_ϵ is far from clear. The problem is to determine when, how, and where $x_\epsilon(\cdot)$ will exit from G when $\epsilon \ll 1$ is very small. We will investigate it when $b(x) = -(\nabla V)(x)$; that is, (3.3) is the gradient flow of a nice function $V(\cdot)$. The infimum of $V(y)$ on ∂G is assumed to be attained uniquely at a point $y_0 \in \partial G$.

The picture that emerges is that a typical path will go quickly near the equilibrium point and stay around it for a long time, making periodic, futile, short-lived attempts to get out. These attempts, although infrequent, are large in number, since the total time it takes to exit is very large. The more serious the attempt, the fewer the number of such attempts. Each individual attempt occurs at a Poisson rate that is tiny. Finally, a successful excursion takes place. The point of exit is close to the minimizer y_0 of $V(y)$ on the boundary. If we assume it is unique, the path followed near the end is the reverse path of the approach to equilibrium of $x(\cdot)$ starting from y_0, and the total time it takes for the exit to take place is of the order $\exp[\frac{2}{\epsilon}(V(y_0) - V(x_0))]$. Compared to the total time, the duration of individual excursions are tiny and can be considered to be almost instantaneous, and so they are almost independent. Various excursions take place more or less independently with various tiny rates. Among the excursions that get out, the one that occurs first is the reverse path that exits at y_0. Its rate is the highest among those that get out.

3.2. Large Deviations of $\{P_{\epsilon,x}\}$

The mapping $x(\cdot) \to g(\cdot)$ of

$$x(t) = x(0) + \int_o^t b(x(s))ds + g(t)$$

is clearly a continuous map of $C[[0,T];\mathbb{R}^d] \to C_0[[0,T];\mathbb{R}^d]$. On the other hand, the difference of two solutions $x(\cdot)$ and $y(\cdot)$, corresponding to $g(\cdot)$ and $h(\cdot)$ respectively, satisfies

$$x(t) - y(t) = \int_0^t [b(x(s)) - b(y(s))]ds + g(t) - h(t),$$

and if $b(x)$ is Lipschitz with constant A, $\Delta(t) = \sup_{0\le s\le t} |x(s) - y(s)|$ satisfies

$$\Delta(t) \le A\int_0^t \Delta(s)ds + \sup_{0\le s\le t} |g(s) - h(s)|.$$

Applying Gronwall's inequality, for any fixed interval $[0, T]$,

$$\Delta(T) \le c(T) \sup_{0 \le s \le T} |g(s) - h(s)|,$$

proving that the map from $g(\cdot) \to x(\cdot)$ is continuous. If we denote this continuous map by $\phi = \phi_x$ and the distribution of the scaled Brownian motion $\sqrt{\epsilon}\beta(\cdot)$ by Q_ϵ, then $P_{x,\epsilon} = Q_\epsilon \phi_x^{-1}$. The probability $P_{\epsilon,x}(B(f,\delta))$ will be estimated by $Q_\epsilon[B(g,\delta')]$. We will prove two theorems.

THEOREM 3.1. *The measures Q_ϵ on $C_0[[0,T];\mathbb{R}^d]$ satisfy the following estimates: For any closed set C that is a subset of $C_0[[0,T];\mathbb{R}^d]$,*

$$\limsup_{\epsilon \to 0} \epsilon \log Q_\epsilon[C] \le - \inf_{g \in C} \frac{1}{2} \int_0^T [g'(t)]^2 \, dt, \tag{3.6}$$

and for any open set $G \subset C_0[[0,T];\mathbb{R}^d]$

$$\liminf_{\epsilon \to 0} \epsilon \log Q_\epsilon[G] \ge - \inf_{g \in G} \frac{1}{2} \int_0^T [g'(t)]^2 \, dt. \tag{3.7}$$

THEOREM 3.2. *The measures $P_{x,\epsilon}$ on $C[[0,T];\mathbb{R}^d]$ satisfy the following estimates: For any closed set $C \subset C[[0,T];\mathbb{R}^d]$,*

$$\limsup_{\epsilon \to 0} \epsilon \log P_{x,\epsilon}[C] \le - \inf_{\substack{f \in C \\ f(0)=x}} \frac{1}{2} \int_0^T [f'(t) - b(f(t))]^2 \, dt, \tag{3.8}$$

and for any open set $G \subset C_0[[0,T];\mathbb{R}^d]$,

$$\liminf_{\epsilon \to 0} \epsilon \log P_{x,\epsilon}[G] \ge - \inf_{\substack{f \in G \\ f(0)=x}} \frac{1}{2} \int_0^T [f'(t) - b(f(t))]^2 \, dt. \tag{3.9}$$

In both theorems the infimum is taken over f and g, which are absolutely continuous in t and have square integrable derivatives.

We note that Theorem 3.2 follows immediately from Theorem 3.1. Since $P_{x,\epsilon}(A) = Q_\epsilon(\phi_x^{-1} A)$ and ϕ_x is a continuous one-to-one map of $C_0[0,T]$ on to $C_x[[0,T];\mathbb{R}^d]$, we only need to observe that

$$\inf_{\phi_x g \in C} \frac{1}{2} \int_0^T [g'(t)]^2 \, dt = \inf_{\substack{f \in C, \\ f(0)=x}} \frac{1}{2} \int_0^T [f'(t) - b(f(t))]^2 \, dt,$$

which is an immediate consequence of the following relation: if $f = \phi_x g$, then

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

We now turn to the proof of Theorem 3.1. This was independently observed in some form by Strassen [**27**] and Schilder [**25**].

PROOF OF THEOREM 3.1. Let us take an integer N and divide the interval $[0,T]$ into N equal parts. For any $f \in C[[0,T];\mathbb{R}^d]$ we denote by $f_N = \pi_N f$ the piecewise linear approximation of f obtained by interpolating linearly over $[\frac{(j-1)T}{N}, \frac{jT}{N}]$, for $j = 1, \dots, N$. In particular, $f_N\left(\frac{jT}{N}\right) = f\left(\frac{Tj}{N}\right)$ for $j = 0, \dots, N$. To prove the upper bound, let $\delta > 0$ be arbitrary and N be an integer. Then,

$$Q_\epsilon[C] \le Q_\epsilon\left[f_N \in C^\delta\right] + Q_\epsilon[\|\pi_N f - f\| \ge \delta]$$

where $C^\delta = \bigcup_{f\in C} B(f,\delta)$. Under Q_ϵ, $\{f\left(\frac{jT}{N}\right)\}$ has a multivariate Gaussian distribution with density

$$\psi_N(\epsilon, z) = \left[\sqrt{\frac{N}{2\pi\epsilon T}}\right]^{Nd} \exp\left[-\frac{N}{2\epsilon T}\sum_{j=1}^{N}[z_j - z_{j-1}]^2\right].$$

Moreover, with $z_j = f(\frac{jT}{N})$,

$$\frac{N}{T}\sum_{j=1}^{N}[z_j - z_{j-1}]^2 = \int_0^T [f_N'(t)]^2\, dt$$

If we define by $D_N \subset (\mathbb{R}^d)^N$ the range of $\{f_N(\frac{jT}{N})\}$ as f varies over C_δ,

$$Q_\epsilon[f_N \in C^\delta] \le \int_{D_N} \psi_N(\epsilon, z)dz, \tag{3.10}$$

and it is now not difficult to show that

$$\limsup_{\epsilon\to 0} \epsilon \log Q_\epsilon[f_N \in C^\delta] \le -\frac{1}{2}\inf_{f\in C^\delta}\int_0^T [f'(t)]^2\, dt.$$

A simple estimate on the maximum oscillation of Brownian motion with variance ϵ in an interval of size $\frac{T}{N}$ provides the bound

$$Q_\epsilon[\|f_N - f\| \ge \delta] \le N\, Q_\epsilon\left[\sup_{0\le t\le \frac{T}{N}} |f(t)| \ge \frac{\delta}{2}\right] \le C(N,d)\exp\left[-\frac{N\delta^2}{8d\epsilon T}\right]$$

yielding

$$\limsup_{\epsilon\to 0} \epsilon \log Q_\epsilon[\|f_N - f\| \ge \delta] \le -\frac{N\delta^2}{8dT}.$$

If we now let $N \to \infty$ and then let $\delta \to 0$, we obtain (3.6). We note that the function

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

is lower semicontinuous on $C[[0,T];\mathbb{R}^d]$ and the level sets $\{f : I(f) \le \ell\}$ are all compact. This allows us to conclude that, for any closed set C,

$$\lim_{\delta\to 0}\inf_{f\in C^\delta} I(f) = \inf_{f\in C} I(f).$$

Another elementary but important fact is that the sum of two nonnegative quantities behaves like the maximum if we are only interested in the exponential rate of decay (or growth).

Now we turn to the lower bound. It is sufficient for us to show that for any $f \in C_0[[0,T];\mathbb{R}^d]$ with $\ell = \frac{1}{2}\int_0^T |f'(t)|^2\, dt < \infty$ and $\delta > 0$

$$\liminf_{\epsilon\to 0} \epsilon \log Q_\epsilon[B(f,\delta)] \ge -\ell.$$

Since f can be approximated by more regular functions f_k with the corresponding ℓ_k approximating ℓ, we can assume without loss of generality that f is smooth. If we denote by $Q_{f,\epsilon}$ the distribution of $\sqrt{\epsilon}\beta(t) - f(t)$, we have

$$\begin{aligned}
Q_\epsilon[B(f,\delta)] &= Q_{f,\epsilon}[B(0,\delta)] \\
&= \int\limits_{B(0,\delta)} \frac{dQ_{f,\epsilon}}{dQ_\epsilon} dQ_\epsilon \\
&= \int\limits_{B(0,\delta)} \exp\left[\frac{1}{\epsilon}\int_0^T f'(s)\cdot dx(s) - \frac{1}{2\epsilon}\int_0^T |f'(t)|^2\,dt\right] dQ_\epsilon \\
&\geq e^{-\frac{\ell}{\epsilon}} Q_\epsilon[B(0,\delta)] \frac{1}{Q_\epsilon[B(0,\delta)]} \int\limits_{B(0,\delta)} \exp\left[\frac{1}{\epsilon}\int_0^T f'(s)\cdot dx(s)\right] dQ_\epsilon \\
&\geq e^{-\frac{\ell}{\epsilon}} Q_\epsilon[B(0,\delta)] \exp\left[\frac{1}{Q_\epsilon[B(0,\delta)]} \int\limits_{B(0,\delta)} \left[\frac{1}{\epsilon}\int_0^T f'(s)\cdot dx(s)\right] dQ_\epsilon\right] \\
&\geq e^{-\frac{\ell}{\epsilon}} Q_\epsilon[B(0,\delta)]
\end{aligned}$$

by Jensen's inequality coupled with symmetry. Since for any $\delta > 0$, $Q_\epsilon[B(0,\delta)] \to 1$ as $\epsilon \to 0$, we are done. □

REMARK 3.3. We will need local uniformity in x in the statement of our large-deviation principle for $P_{\epsilon,x}$. This follows easily from the continuity of the maps ϕ_x in x.

REMARK 3.4. This does not quite solve the exit problem. The estimates are good only for a finite T, and all estimates only show that the probabilities involved are quite small. The solution to the exit problem is slightly more subtle. The basic idea is that among a bunch of very unlikely things the least unlikely thing is most likely to occur first!

3.3. The Exit Problem

We start with a lemma that is a variational calculation. Consider any path $h(\cdot)$ that starts from the stable equilibrium x_0 and ends at some $x \in G$.

LEMMA 3.5.

$$\inf_{0<T<\infty} \inf_{\substack{h,\\ h(0)=x,\\ h(T)=x_0}} \int_0^T [h'(t) + \nabla V]^2\,dt = 4[V(x) - V(x_0)]. \tag{3.11}$$

PROOF. We look at the ODE $\dot{x}(t) = -(\nabla V)(x(t))$, $x(0) = x_0$, and reverse it between 0 and T, giving a trajectory $h(t) = x(T-t)$ from $x(T)$ to x satisfying $h'(t) = (\nabla V)(h(t))$:

$$\begin{aligned}
\int_0^T [h'(t) + \nabla(V)(h(t))]^2\,dt &= \int_0^T [h'(t) - \nabla(V)(h(t))]^2\,dt \\
&\quad + 4\int_0^T (\nabla V)(h(t))\cdot h'(t)dt \\
&= 4[V(x) - V(x(T))].
\end{aligned}$$

For T large, $x(T) \simeq x_0$ and therefore

$$\inf_{0<T<\infty} \inf_{\substack{h \\ h(0)=x_0 \\ h(T)=x}} \int_0^T [h'(t) + \nabla V]^2 \, dt \le 4[V(x) - V(x_0)].$$

On the other hand, for any h with $h(T) = x$ and $h(0) = x_0$,

$$\begin{aligned} 4[V(x) - V(x_0)] &= 4 \int_0^T (\nabla V)(h(t)) \cdot h'(t) dt \\ &= \int_0^T [h'(t) + (\nabla V)(h(t)]^2 \, dt - \int_0^T [h'(t) - (\nabla V)(h(t)]^2 \, dt \\ &\le \int_0^T [h'(t) + (\nabla V)(h(t)]^2 \, dt. \end{aligned}$$ □

The next lemma says that it is very unlikely that the path stays away from the equilibrium point for too long.

LEMMA 3.6. *Let U be any neighborhood of the equilibrium x_0 and*

$$\Lambda(U, T) = \inf_{f(\cdot): f(\cdot) \in G \cap U^c} \int_0^T [f'(t) + (\nabla V)(f(t))]^2 \, dt,$$

Then $\liminf_{T\to\infty} \Lambda(U, T) = \infty$.

PROOF. Suppose there are paths in $G \cap U^c$ for long periods with bounded rate

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

Then, there has to be arbitrarily long stretches for which the contribution to $I(f)$ is small. Such trajectories are equicontinuous and produce in the limit solutions of $dx(t) + (\nabla V)(x(t))dt = 0$ that live in $G \cap U^c$ forever, which is a contradiction. □

Now we state and prove the main theorem.

THEOREM 3.7. *Assume that $V(\cdot)$ on the boundary δG achieves its minimum at a unique point y_0. Then, for any $x \in G$,*

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

In other words, irrespective of the starting point, exit will take place near y_0 with probability nearly 1.

PROOF. Let us fix a neighborhood N of y_0 on the boundary. Let $\inf_{y\in\delta G\cap N^c} V(y) = V(y_0) + \theta$ for some $\theta > 0$. Let us take two neighborhoods, U_1, U_2, of x_0 such that $\overline{U}_1 \subset U_2$ and $V(x) - V(x_0) \le \frac{\theta}{10}$ on $\overline{U}_2$. Let τ be the exit time from G. We will show that, for any $x \in G$,

$$\lim_{\epsilon\to 0} P_{x,\epsilon}[x(\tau) \notin N] = 0.$$

Let us define the following stopping times:

$$\begin{aligned} \tau &= \inf\{t : x(t) \notin G\}, \\ \tau_1 &= \inf\{t : x(t) \notin \overline{U}_1^c\} \wedge \tau, \\ \tau_2 &= \inf\{t \ge \tau_1 : x(t) \notin U_2\}, \end{aligned}$$

$$\vdots$$

$$\begin{aligned}\tau_{2k+1} &= \inf\{t \geq \tau_{2k} : x(t) \notin \overline{U}_1^{\mathrm{c}}\} \wedge \tau,\\ \tau_{2k+2} &= \inf\{t \geq \tau_{2k+1} : x(t) \notin U_2\}.\end{aligned}$$

For any $x \in G$, $P_{x,\epsilon}[\tau_1 = \tau] \to 0$ as $\epsilon \to 0$ and the path cannot exit from G between τ_{2k+1} and τ_{2k+2}. As for τ_{2k+1}, one of three things can happen: $\tau > \tau_{2k+1}$ and then $x(\tau_{2k+1}) \in \partial U_1$, or $\tau = \tau_{2k+1}$ in which case either $x(\tau_{2k+1}) = x(\tau) \in N$ or $x(\tau_{2k+1}) = x(\tau) \in \partial G \cap N^{\mathrm{c}}$. The first event has probability nearly 1 and the remaining two have probability nearly 0. But one of them has much smaller probability than the other. So the event that has the larger of the two probabilities will happen first. We need to prove only that

$$\lim_{\epsilon\to 0} \frac{\sup_{x\in\delta U_2} P_{x,\epsilon}[\{\tau_1 = \tau\} \cap \{x(\tau) \notin N\}]}{\inf_{x\in\delta U_2} P_{x,\epsilon}[\{\tau_1 = \tau\} \cap \{x(\tau) \in N\}]} = 0.$$

Let us look at the numerator first.

$$\begin{aligned}a(x,\epsilon) &= P_{x,\epsilon}[\{\tau_1 = \tau\} \cap \{x(\tau) \notin N\}]\\ &\leq P_{x,\epsilon}[\{\tau_1 = \tau\} \cap \{x(\tau) \notin N\} \cap \{\tau_1 \leq T\}] + P_{x,\epsilon}[\tau_1 \geq T].\end{aligned}$$

By Lemma 3.6 the second term on the right can be made superexponentially small; i.e.,

$$\limsup_{T\to\infty} \limsup_{\epsilon\to 0} \epsilon \log P_{x,\epsilon}[\tau_1 \geq T] = -\infty.$$

The first term has an explicit exponential rate and, for any T,

$$\begin{aligned}&\limsup_{\epsilon\to 0} \epsilon \log \sup_{x\in\delta U_2} P_{x,\epsilon}[\{\tau_1 = \tau\} \cap \{x(\tau) \notin N\} \cap \{\tau_1 \leq T\}]\\ &\qquad \leq -2 \inf_{y\in N^{\mathrm{c}}} \inf_{x\in\delta U_2} [V(y) - V(x)]\\ &\qquad \leq -\frac{3\theta}{2} - 2[V(y_0) - V(x_0)]\end{aligned}$$

Therefore,

$$\limsup_{\epsilon\to 0} \epsilon \log \sup_{x\in\delta U_2} a(x,\epsilon) \leq -\frac{3\theta}{2} - 2[V(y_0) - V(x_0)].$$

On the other hand, for estimating the denominator,

$$\begin{aligned}\liminf_{\epsilon\to 0} \epsilon \inf_{x\in\delta U_2} \log P_{x,\epsilon}[\{\tau_1 = \tau\} \cap \{x(\tau) \in N\}] &\geq -\sup_{x\in\delta U_2} 2[V(y_0) - V(x)]\\ &\geq -2[V(y_0) - V(x_0)] - \frac{\theta}{5}.\end{aligned}$$

The numerator goes to 0 a lot faster than the denominator and the ratio therefore goes to 0. $\square$

REMARK 3.8. It is not important that $b(x) = -(\nabla V)(x)$ for some V. Otherwise, if x_0 is the unique stable equilibrium in G, for $x \in G$, one can define the "quasi-potential" $V(x)$ by

$$4V(x) = \inf_{0<T<\infty} \inf_{\substack{h(\cdot),\\ h(0)=x_0,\\ h(T)=x}} \int_0^T [x'(t) - b(x(t))]^2\, dt,$$

and it works just as well.

3.4. Superexponential Estimates

Often we have a sequence $X_{n,k}$ of random variables with values in $\mathcal{X}$ that are defined on some (Ω, Σ, P), and, for each fixed k, we have an LDP for $P_{n,k}$, the distribution of $X_{n,k}$ on $\mathcal{X}$ with rate function $I_k(x)$. As $k \to \infty$, for each n, $X_{n,k} \to X_n$; i.e., for each $\delta > 0$ and for each n,

$$\lim_{k\to\infty} P[d(X_{n,k}, X_n) \geq \delta] = 0. \tag{3.12}$$

We want to prove an LDP for P_n the distribution of X_n on $\mathcal{X}$. This involves interchanging two limits and needs additional estimates. The following "superexponential estimate" is enough. For each fixed $\delta > 0$,

$$\limsup_{k\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P[d(X_{n,k}, X_n) \geq \delta] = -\infty. \tag{3.13}$$

THEOREM 3.9. *If for each k the distributions $\{P_{n,k}\}$ of $X_{n,k}$ satisfy a large-deviation principle with a rate function $I_k(x)$, and if* (3.12) *and* (3.13) *hold, then for each x,*

$$\lim_{\delta\to 0} \liminf_{k\to\infty} \inf_{y\in B(x,\delta)} I_k(y) = \lim_{\delta\to 0} \limsup_{k\to\infty} \inf_{y\in B(x,\delta)} I_k(y).$$

The common limit $I(\cdot)$ is a rate function, and the distribution P_n of X_n satisfies LDP with rate $I(\cdot)$.

PROOF. Let us define $I^+(x) \geq I^-(x)$ as

$$I^+(x) = \lim_{\delta\to 0} \limsup_{k\to\infty} \inf_{y\in B(x,\delta)} I_k(y), \qquad I^-(x) = \lim_{\delta\to 0} \liminf_{k\to\infty} \inf_{y\in B(x,\delta)} I_k(y).$$

We will establish the theorem in three steps. We first show that $I^-(x)$ is lower semicontinuous and its level sets $\{x : I^-(x) \leq \ell\}$ are compact. We then prove local upper bounds with $I^+(\cdot)$ and local lower bounds with $I^-(\cdot)$. Since $I^+(x) \geq I^-(x)$, this will show that $I^+(x) \equiv I^-(x)$.

Step 1. We know that, for each k, the set $\{x : I_k(x) \leq \ell\}$ is compact. We begin by showing that any sequence $\{x_k\}$ such that $I_k(x_k) \leq \ell$, for some $\ell < \infty$, has a convergent subsequence. In other words $\bigcup_k \{x : I_k(x) \leq \ell\}$ is totally bounded for each $\ell < \infty$. Let $\ell < \infty$ and $\delta > 0$ be given. Then from (3.13), there exists k_0 such that for $k \geq k_0$

$$\limsup_{n\to\infty} \frac{1}{n} \log P[d(X_{n,k}, X_n) \geq \delta] \leq -2\ell.$$

Clearly,

$$\begin{aligned} &P[d(X_{n,k_0}, x_k) \leq 3\delta] \\ &\quad \geq P\big[[d(X_{n,k_0}, X_n) \leq \delta] \cap [d(X_n, X_{n,k}) \leq \delta] \cap [d(X_{n,k}, x_k) \leq \delta]\big] \\ &\quad \geq P[d(X_{n,k}, x_k) \leq \delta] - P[d(X_{n,k}, X_n) \geq \delta] - P[d(X_{n,k_0}, X_n) \geq \delta]. \end{aligned}$$

Or,

$$\begin{aligned} P[d(X_{n,k}, x_k) \leq \delta] \leq P[d(X_{n,k_0}, x_k) \leq 3\delta] &+ P[d(X_{n,k}, X_n) \geq \delta] \\ &+ P[d(X_{n,k_0}, X_n) \geq \delta]. \end{aligned}$$

This implies, for any fixed $k \geq k_0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log P[d(X_{n,k}, x_k) \leq \delta] \leq$$
$$\limsup_{n\to\infty} \frac{1}{n} \log \max\{P[d(X_{n,k_0}, x_k) \leq 3\delta], P[d(X_{n,k}, X_n) \geq \delta], P[d(X_{n,k_0}, X_n) \geq \delta]\}.$$

Since $I_k(x_k) \leq \ell$ and $\limsup_{n\to\infty} \frac{1}{n} \log P[d(X_{n,k}, X_n) \geq \delta] \leq -2\ell$ for all $k \geq k_0$,

$$\inf_{y\in B(x_k,3\delta)} I_{k_0}(y) \leq \inf_{y\in B(x_k,\delta)} I_k(y) \leq \ell.$$

This shows that for any arbitrary $\delta > 0$, there is a sequence $y_k \in B(x_k, 3\delta)$ with $I_{k_0}(y_k) \leq \ell$, which therefore has a convergent subsequence. By a variant of the diagonalization process, we can find a subsequence x_{k_r} such that there is y_{r,k_r} with $d(x_{k_r}, y_{r,k_r}) \leq 2^{-r}$ and, for each j, $y_{j,k_r} \to y_j$ as $r \to \infty$. In other words, we can assume without loss of generality that for any $\delta > 0$, there is $y_k \in B(x_k, \delta)$ that converges to a limit. It is easy to check now that $\{x_k\}$ must be a Cauchy sequence. Since the space is complete, it converges.

The next step is to show that $C_\ell = \{x : I^-(x) \leq \ell\}$ is compact, i.e., totally bounded and closed. If we denote by $D_{k,\ell} = \{x : I_k(x) \leq \ell\}$, then

$$C_\ell = \bigcap_{\ell'>\ell} \bigcap_{\delta>0} \bigcap_{k'\geq 1} \overline{\Big[\bigcup_{k\geq k'} D_{k,\ell'}\Big]^\delta}.$$

It is clear that C_ℓ is closed. Since $\bigcup_{k\geq k'} D_{k,\ell'}$ is totally bounded, it follows that so is C_ℓ.

Step 2. Let $C \subset \mathcal{X}$ be closed. Then, either $X_{n,k} \in \overline{C}^\delta$ or $d(X_{n,k}, X_n) \geq \delta$. Therefore for any k,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n[C] \leq \max\big\{-\inf_{x\in\overline{C}^\delta} I_k(x), \theta_k\big\},$$

where $\theta_k \to -\infty$ as $k \to \infty$. Consequently,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n[C] \leq -\limsup_{\delta\to 0} \limsup_{k\to\infty} \inf_{x\in\overline{C}^\delta} I_k(x) \leq -\inf_{x\in C} I^+(x).$$

In particular,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log P_n[B(x,\delta)] \leq -I^+(x).$$

Step 3. Let $I(x) = \ell < \infty$. Then, there are $x_k \in B(x, \delta)$ with $I_k(x_k) \leq \ell + \epsilon$ and

$$P_n[B(x, 2\delta)] \geq P_{n,k}[B(x_k, \delta)] - P[d(X_{n,k}, X_n) \geq \delta].$$

We choose k large enough so that the second term on the right is negligible compared to the first. We then obtain

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log P_n[B(x, 2\delta)] \geq -I^-(x).$$

This proves $I^+(x) = I^-(x)$. $\square$

3.5. General Diffusion Processes

We can have processes $P_{x,\epsilon}$ that correspond to more general operators

$$\mathcal{L}_\epsilon u = \frac{\epsilon}{2}\sum_{i,j=1}^{d} a_{i,j}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{j=1}^{d} b_j(x)\frac{\partial u}{\partial x_j}.$$

The rate function for large deviations of $P_{x,\epsilon}$ will now be

$$I(f) = \frac{1}{2}\int_0^T \sum_{i,j=1}^{d} \big\langle a^{-1}(f(t))\big(f'(t) - b(f(t))\big), (f'(t) - b(f(t))\big\rangle dt.$$

The proof would proceed along the following lines. We will assume that all the coefficients are smooth and, in addition, $\{a_{i,j}(x)\}$ is uniformly elliptic. This provides a choice of the square root σ that is smooth as well. The distribution $P_{x,\epsilon}$ is now the distribution of the solution of the SDE

$$x(t) = x + \sqrt{\epsilon}\int_0^t \sigma(x(s))\cdot d\beta(s) + \int_0^t b(x(s))ds,$$

which has (almost surely) a uniquely defined solution. We have a large deviation for $\sqrt{\epsilon}\beta(t)$ with rate function as before

$$I_0(f) = \frac{1}{2}\int_0^T \|f'(t)\|^2\, dt.$$

The map $\beta(\cdot) \to x(\cdot)$ is, however, not continuous in the usual topology on $C[[0,T];\mathbb{R}^d]$. For any N and $\epsilon > 0$, we can approximate $x_\epsilon(t)$ by $x_{N,\epsilon}(t)$, which solves

$$x_{N,\epsilon}(t) = x + \sqrt{\epsilon}\int_0^t \sigma(x_{N,\epsilon}(\pi_N(s))d\beta(s) + \int_0^t b\big(x_{N,\epsilon}(\pi_N(s))\big)ds$$

where $\pi_N(s) = \frac{[Ns]}{N}$. The coefficients are frozen and updated every $\frac{1}{N}$ unit of time. The map $\beta(\cdot) \to x_N(\cdot)$ is continuous, and therefore the distribution of $x_N(t)$ satisfies a large-deviation principle with rate function

$$\begin{aligned} I_N(f) &= \frac{1}{2}\int_0^T \big\|\sigma^{-1}\big(x(\pi_N(s))\big)\big[f'(s) - b\big(x(\pi_N(s))\big)\big]\big\|^2\, ds \\ &= \frac{1}{2}\int_0^T \big\langle a^{-1}\big(x(\pi_N(s))\big)\big[f'(s) - b\big(x(\pi_N(s))\big)\big], \big[f'(s) - b\big(x(\pi_N(s))\big)\big]\big\rangle ds. \end{aligned}$$

The proof is completed (see Theorem 3.9) by proving that for any $\delta > 0$,

$$\lim_{N\to\infty}\limsup_{\epsilon\to 0} \epsilon \log P\Big[\sup_{0\le t\le T}\|x_{N,\epsilon}(t) - x_\epsilon(t)\| \ge \delta\Big] = -\infty \tag{3.14}$$

and

$$I(f) = \inf_{f_N \to f}\liminf_{N\to\infty} I_N(f_N) \tag{3.15}$$

where the infimum is taken over all sequences $\{f_N\}$ that converge to f.

Denoting by $Z_{N,\epsilon}(t) = x_{N,\epsilon}(t) - x_\epsilon(t)$, we have

$$Z_{N,\epsilon}(t) = \sqrt{\epsilon}\int_0^t e_N(s)d\beta(s) + \int_0^t g_N(s)ds$$

where

$$\|e_N(s)\| = \big\|\sigma\big(x_{N,\epsilon}(\pi_N(s))\big) - \sigma(x(s))\big\| \le A\|Z_{N,\epsilon}(s)\| + A\|x_{N,\epsilon}(\pi_N(s)) - x_{N,\epsilon}(s)\|$$

and

$$\|g_N(s)\| = \big\|b\big(x_{N,\epsilon}(\pi_N(s))\big) - b(x_\epsilon(s))\big\| \le A\|Z_{N,\epsilon}(s)\| + A\|x_{N,\epsilon}(\pi_N(s)) - x_{N,\epsilon}(s)\|.$$

If we define the stopping time τ as

$$\tau = \inf\{s : \|x_{N,\epsilon}(\pi_N(s)) - x_{N,\epsilon}(s)\| \ge \eta\} \wedge T,$$

$$\begin{aligned}
&P\Big[\sup_{0\le t\le T} \|x_{N,\epsilon}(t) - x_\epsilon(t)\| \ge \delta\Big] \\
&\quad\le P\Big[\sup_{0\le t\le \tau} \|x_{N,\epsilon}(t) - x_\epsilon(t)\| \ge \delta\Big] + P[\tau < T] \\
&\quad\le P\Big[\sup_{0\le t\le \tau} \|x_{N,\epsilon}(t) - x_\epsilon(t)\| \ge \delta\Big] + P\Big[\sup_{0\le s\le T} \|x_{N,\epsilon}(\pi_N(s)) - x_{N,\epsilon}(s)\| \ge \eta\Big] \\
&\quad= \Theta_1 + \Theta_2.
\end{aligned}$$

Let us handle each of the two terms separately. First, we need this lemma.

LEMMA 3.10. *Let $z(t)$ be a process satisfying*

$$z(t) = \int_0^s e(s)\cdot d\beta(s) + \int g(s)ds$$

with $\|e(s)\| \le B(\eta^2 + \|z\|^2)^{1/2}$, $\|g(s)\| \le A(\eta^2 + \|z\|^2)^{1/2}$ in some interval $0 \le t \le \tau$ where $\tau \le T$ is a stopping time. Then, for any $\ell \ge 0$,

$$P\Big[\sup_{0\le t\le \tau} \|z(s)\| \ge \delta\Big] \le \left[\frac{\delta^2}{\delta^2+\eta^2}\right]^\ell e^{T(2A\ell + 4B^2\ell^2)}.$$

PROOF. Consider the function

$$f(x) = (\eta^2 + \|x\|^2)^\ell.$$

By Itô's formula,

$$df(z(t)) = (\nabla f)(z(t))\cdot dz(t) + \frac{1}{2}\operatorname{Tr}\big[(\nabla^2 f)(z(t))e(t)e^*(t)\big]dt = a(t)dt + m(t)$$

where $m(t)$ is a martingale and

$$|a(t)| \le (2B\ell + 4A^2\ell^2)(\eta^2 + \|z(t)\|^2)^\ell.$$

Therefore,

$$f(z(t))e^{-t(2A\ell + 4B^2\ell^2)}$$

is a supermartingale and

$$P\Big[\sup_{0\le s\le \tau} \|z(s)\| \ge \delta\Big] \le \left[\frac{\eta^2}{\delta^2+\eta^2}\right]^\ell e^{T(2A\ell + 4B^2\ell^2)}. \qquad \square$$

In $0 \le t \le \tau$, we have

$$\|e_n\| \le 2A\big[\|Z_{N,\epsilon}\|^2 + \eta^2\big]^{\frac{1}{2}};\ \|g_n\| \le 2A\big[\|Z_{N,\epsilon}\|^2 + \eta^2\big]^{\frac{1}{2}}.$$

Applying the lemma with $2A$ and $2\sqrt{\epsilon}A$ replacing A and B, we obtain with $\ell = \frac{1}{\epsilon}$

$$\epsilon \log \Theta_1 \le \log \frac{\eta^2}{\delta^2 + \eta^2} + T[2A + 4A^2]. \tag{3.16}$$

Now, we turn to Θ_2. We will use the following lemma.

LEMMA 3.11. *Let*

$$z(t) = x + \sqrt{\epsilon}\int_0^t e(s)\cdot d\beta(s) + \int_0^t g(s)ds$$

where $\|e(s)\|, \|g(s)\|$ *are bounded by* C. *Then, for any* $\eta > 0$,

$$\limsup_{N\to\infty}\limsup_{\epsilon\to 0}\epsilon\log P\Big[\sup_{0\le s\le T}\|z(\pi_N(s)) - z(s)\| \ge \eta\Big] = -\infty.$$

PROOF. We can choose N large enough so that $\frac{C}{N} \le \frac{\eta}{2}$. Then, we need only show that

$$\limsup_{N\to\infty}\limsup_{\epsilon\to 0}\log P\left[\sup_{0\le t\le\frac{1}{N}}\left\|\int_0^t e(s)\cdot d\beta\right\| \ge \frac{\eta}{2\sqrt{\epsilon}}\right] = -\infty,$$

which is an immediate consequence of the following observation. If $e(s)e^*(s) \le CI$, for any $k > 0$ and unit vector $\theta \in \mathbb{R}^d$,

$$\exp\left[k\left\langle\theta, \int_0^t e(s)\cdot d\beta(s)\right\rangle - \frac{Ck^2t}{2}\right]$$

is a supermartingale and

$$P\left[\sup_{0\le t\le T}\left\langle\theta, \int_0^t e(s)\cdot d\beta(s)\right\rangle \ge \rho\right] \le \exp\left[-k\rho + \frac{Ck^2T}{2}\right].$$

Optimizing over k, we get

$$P\left[\sup_{0\le t\le T}\left\langle\theta, \int_0^t e(s)\cdot d\beta(s)\right\rangle \ge \rho\right] \le \exp\left[-\frac{\rho^2}{2CT}\right].$$ □

This shows that, for any $\eta > 0$,

$$\limsup_{N\to\infty}\limsup_{\epsilon\to 0}\epsilon\log\Theta_2 = -\infty. \tag{3.17}$$

We conclude by letting $\epsilon \to 0$, then $N \to \infty$, and finally $\eta \to 0$. Estimates (3.16) and (3.17) imply (3.14).

Finally, it is not difficult to show that

$$\inf_{f_N(\cdot)\to f(\cdot)}\liminf_{N\to\infty}\int_0^T \langle f_N'(t), a^{-1}(f_N(\pi_N(t)))f_N'(t)\rangle dt = \int_0^T \langle f'(t), a^{-1}(f(t))f'(t)\rangle dt.$$

We have therefore proved the following theorem.

THEOREM 3.12. *Let* $\{a_{i,j}(x)\}$ *be smooth and uniformly elliptic, and let* $b(x)$ *be smooth and bounded. Then the distribution* $P_{\epsilon,x}$ *of diffusion with generator*

$$(\mathcal{L}_\epsilon u)(x) = \frac{\epsilon}{2}\sum_{i,j=1}^d a_{i,j}(x)\frac{\partial^2 u}{\partial x_i\partial x_j}(x) + \sum_{j=1}^d b_j(x)\frac{\partial u}{\partial x_j}(x)$$

satisfies on $C[[0,T],\mathbb{R}^d]$ *as* $\epsilon \to 0$ *a large-deviation principle with rate*

$$I(f) = \frac{1}{2}\int_0^T \sum_{i,j=1}^d \big\langle a^{-1}(f(t))\big(f'(t) - b(f(t))\big), (f'(t) - b(f(t))\big\rangle dt$$

if $f(0) = x$ *and* $f(t)$ *is absolutely continuous with a square integrable derivative. Otherwise* $I(f) = +\infty$.

REMARK 3.13. In our case it is easy to show directly that $\bigcup_N \{f : I_N(f) \leq \ell\}$ is totally bounded. From the bounds on b and a^{-1}, it is easy to conclude that

$$\bigcup_N \{f : I_N(f) \leq \ell\} \subset \left\{ f : \int_0^T \|f'(t)\|^2 \, dt \leq \ell' \right\}$$

for an ℓ' depending on ℓ and the bounds on a^{-1} and b.

3.6. Short-Time Behavior of Diffusions

Brownian motion on $\mathbb{R}^d$ has the transition density

$$p(t,x,y) = \exp\left[-\frac{|x-y|^2}{2t} + o\left(\frac{1}{t}\right) \right] = \exp\left[-\frac{d(x,y)^2}{2t} + o\left(\frac{1}{t}\right) \right],$$

where $d(x,y)$ is the Euclidean distance. If we replace the Brownian motion with independent components by one with positive definite covariance A, then the metric gets replaced by $d(x,y) = \sqrt{\langle A^{-1}(x-y), (x-y)\rangle}$ and a similar formula for $p_A(t,x,y)$ is still valid, as seen by a simple linear change of coordinates. The natural question that arises is whether there is a similar relation between the transition probability density $p_{\mathcal{L}}(t,x,y)$ of the diffusion with generator

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

and the geodesic distance $d_{\mathcal{L}}(x,y)$ between x and y in the Riemannianian metric

$$ds^2 = \sum_{i,j} a_{i,j}^{-1}(x) dx_i \, dx_j .$$

We will show that indeed there is. In the special case when $b \equiv 0$, for the generator

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

the rate function takes the form

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

This is not changed if we add a small first-order term; i.e.,

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

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Denoting the two measures by Q_ϵ and P_ϵ, the Radon-Nikodym derivative is

$$\frac{dQ_\epsilon}{dP_\epsilon} = \exp\left[\frac{\delta(\epsilon)}{\epsilon} \int_0^T \langle a^{-1}(x(s)) b(x(s)), dx(s) \rangle - \frac{\delta(\epsilon)^2}{2\epsilon} \int_0^T \langle a^{-1}(x(s)) b(x(s)), b(x(s)) \rangle ds \right].$$

From the boundedness of a, a^{-1}, and b, it is easy to deduce (see exercise at the end) that for any k,

$$\lim_{\epsilon \to 0} \epsilon \log E^{P_\epsilon}\left[\left[\frac{dQ_\epsilon}{dP_\epsilon} \right]^k \right] = 0$$

and

$$\lim_{\epsilon\to 0} \epsilon \log E^{Q_\epsilon}\left[\left[\frac{dP_\epsilon}{dQ_\epsilon}\right]^k\right] = 0.$$

We can now estimate, by Hölder's inequality,

$$Q_\epsilon(A) = \int_A \frac{dQ_\epsilon}{dP_\epsilon}\, dP_\epsilon \le [P_\epsilon(A)]^{\frac{1}{p}} \left\| \frac{dQ_\epsilon}{dP_\epsilon} \right\|_{q,P_\epsilon},$$

as well as,

$$P_\epsilon(A) = \int_A \frac{dP_\epsilon}{dQ_\epsilon}\, dQ_\epsilon \le [Q_\epsilon(A)]^{\frac{1}{p}} \left\| \frac{dP_\epsilon}{dQ_\epsilon} \right\|_{q,Q_\epsilon}.$$

By choosing $p > 1$ but arbitrarily close to 1, $P_\epsilon(A)$ and $Q_\epsilon(A)$ are seen to have the same exponential decay rate.

The process with generator $\epsilon\mathcal{L}$ is the same as the process for $\mathcal{L}$ slowed down. Therefore, the transition probability $p(\epsilon, x, dy)$ is the same as the transition probability $p_\epsilon(1, x, dy)$ of $\epsilon\mathcal{L}$. By the contraction principle, we can conclude that for G open

$$\liminf_{\epsilon\to 0} 2\epsilon \log p(\epsilon, x, G) \ge -\inf_{\substack{f: f(0)=x,\\ f(1)\in G}} I(f), \tag{3.18}$$

and for C closed

$$\limsup_{\epsilon\to 0} 2\epsilon \log p(\epsilon, x, C) \ge -\inf_{\substack{f: f(0)=x,\\ f(1)\in C}} I(f). \tag{3.19}$$

Moreover, an elementary calculation shows that

$$\inf_{\substack{f(0)=x,\\ f(1)=y}} I(f) = \frac{1}{2} d(x, y)^2$$

where $d(x, y)$ is the geodesic distance in the metric $ds^2 = \sum_{i,j} a_{i,j}^{-1}(x) dx_i\, dx_j$. One can then use the Chapman-Kolmogorov equation

$$p(t, x, y) = \int p(t_1, x, dz) p(t - t_1, z, y)$$

and improve the estimate on $p(t, x, A)$ to an estimate on $p(t, x, y)$ that takes the form

$$p(t, x, y) = \exp\left[-\frac{d(x, y)^2}{2t} + o\left(\frac{1}{t}\right)\right].$$

Another way of looking at this is, if we have a Riemannian metric $ds^2 = \sum g_{i,j}(x) dx_i\, dx_j$ on $\mathbb{R}^d$ where $\{g_{i,j}(x)\}$ are smooth, bounded, and uniformly positive definite, then the diffusion with generator $\frac{1}{2}\Delta_g$ where Δ_g is Laplacian in the metric g has transition probability density that satisfies

$$p(t, x, y) = \exp\left[-\frac{d_g(x, y)^2}{2t} + o\left(\frac{1}{t}\right)\right] \tag{3.20}$$

where $d_g(x, y)$ is the geodesic distance between x and y in the metric $\{g_{i,j}(x)\}$.

3.7. Supplementary Material

The work on small-time behavior of diffusions was suggested by a result of Cieselski [**1**] that if $p_G(t,x,y)$ is the fundamental solution of the heat equation $u_t = \frac{1}{2}\Delta$ with Dirichlet boundary condition on the boundary ∂G of an open set G and $p(t,x,y)$ the whole space solution, then

$$\lim_{t\to 0} \frac{p_G(t,x,y)}{p(t,x,y)} = 1$$

for all $x, y \in G$ if and only if G is essentially convex. Intuitively, this says that if a Brownian path goes from $x \to y$ in a short time, then it did so in a straight line. The analogue for diffusions would be that the geodesic replaces the straight line. This is a consequence of the large-deviation result as is shown in the following exercises.

Exercise. Assuming a PDE estimate of the form

$$\lim_{\delta\to 0} \limsup_{t\to 0} t \sup_{|x-y|\le\delta} \log p(t,x,y) = \lim_{\delta\to 0} \liminf_{t\to 0} t \inf_{|x-y|\le\delta} \log p(t,x,y) = 0$$

for the fundamental solution of $p(t,x,y)$ of $u_t = \mathcal{L}u$, use (3.18) and (3.19) to prove (3.20).

Exercise. Deduce that the measure $Q_{\epsilon,x,y}$ on path space $C[[0,1],\mathbb{R}^d]$ starting from x with transition probability

$$q_{\epsilon,x,y}(s,x',t,y') = \frac{p(\epsilon(t-s),x',y')p(\epsilon(1-t),y',y)}{p(\epsilon(1-s)x',y)}$$

concentrates as $\epsilon \to 0$ on the set of geodesics connecting x and y.

Strassen in [**27**] used the large-deviation estimate to prove a functional form of the law of the iterated logarithm. Let $\beta(t)$ be the one-dimensional Brownian motion. Let

$$\beta_\lambda(t) = \frac{\beta(\lambda t)}{\sqrt{\lambda \log\log\lambda}}.$$

Then, on the space $C[0,1]$ with probability 1, the set $\{\beta_\lambda(\cdot) : \lambda \ge 10\}$ is conditionally compact and the set of limit points as $\lambda \to \infty$ is precisely the set of f such that $f(0) = 0$ and $I(f) = \frac{1}{2}\int_0^1 [f'(t)]^2\,dt \le 1$. The proof is very similar to the proof of the usual law of the iterated logarithm. Due to the slow change in λ, it is enough to look at $\lambda_n = \rho^n$ for $\rho > 1$. Then,

$$P[\beta_{\rho^n}(\cdot) \in B(f,\delta)] \simeq n^{-I(f)},$$

and we now apply the Borel-Cantelli lemma. One half requires $\rho \to 1$, and the other half requires $\rho \to \infty$ to generate near independence. Now using Skorohod imbedding, one can deduce a similar result for sums of i.i.d. random variables with any arbitrary common distribution with mean 0 and variance 1.

https://doi.org/10.1090//cln/027/04

CHAPTER 4

Large Time

4.1. Introduction

The goal of this chapter is to prove the following theorem:

THEOREM 4.1. *Let $S_n = X_1 + \cdots + X_n$ be the nearest-neighbor random walk on $\mathbb{Z}^d$ with each $X_i = \pm e_r$ with probability $\frac{1}{2d}$ where $\{e_r\}$ are the unit vectors in the d positive coordinate directions. Let D_n be the range of $S_1, \ldots, S_n$ on $\mathbb{Z}^d$. $|D_n|$ is the cardinality of the range D_n. Then*

$$\lim_{n\to\infty} \frac{1}{n^{\frac{d}{d+2}}} \log E\big[e^{-\nu|D_n|}\big] = -k(\nu, d)$$

where

$$k(\nu, d) = \inf_{\ell} \big[v(d)\ell^d + \lambda(d)\ell^{-2}\big]$$

with $v(d)$ equal to the volume of the unit ball, and $\lambda(d)$ is the smaller eigenvalue of $-\frac{1}{2d}\Delta$ in the unit ball with Dirichlet boundary conditions.

The starting point of the investigation is a result on large deviations from the ergodic theorem for Markov chains. Let $\Pi = \{p(x, y)\}$ be the matrix of transition probability of a Markov chain on a finite state space $\mathcal{X}$. We will assume that for any pair x, y, there is some power n with $\Pi^n(x, y) > 0$. Then, there is a unique invariant probability distribution $\pi = \{\pi(x)\}$ such that $\sum_x \pi(x)p(x, y) = \pi(y)$ and, according to the ergodic theorem, for any $f : \mathcal{X} \to \mathbb{R}$ almost surely with respect to the Markov chain P_x starting at time 0 from any $x \in \mathcal{X}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(X_j) = \sum_x f(x)\pi(x).$$

The natural question on large deviations here is to determine the rate of convergence to 0 of

$$P\left[\left|\frac{1}{n}\sum_{j=1}^{n} f(X_j) - \sum_x f(x)\pi(x)\right| \geq \delta\right]. \tag{4.1}$$

More generally, on the space $\mathcal{M}$ of probability distributions on $\mathcal{X}$ we can define a probability measure $Q_{n,x}$ defined as the distribution of the empirical distribution $\frac{1}{n}\sum_{j=1}^{n} \delta_{X_j}$ which is a random point in $\mathcal{M}$. The ergodic theorem can be reinterpreted as

$$\lim_{n\to\infty} Q_{n,x} = \delta_\pi;$$

i.e., the empirical distribution is close to the invariant distribution with probability nearly 1 as $n \to \infty$. We want to establish a large deviation principle and determine the corresponding rate function $I(\mu)$. One can then determine the behavior of (4.1) as

$$\lim_{n\to\infty} \frac{1}{n} \log P\left[\left|\frac{1}{n}\sum_{j=1}^{n} f(X_j) - \sum_x f(x)\pi(x)\right| \geq \delta\right] = \inf\left[I(\mu) : \left|\sum f(x)[\mu(x) - \pi(x]\right| \geq \delta\right].$$

With a little extra work, under suitable transitivity conditions, one can show that for $x \in A$,

$$\lim_{n\to\infty} \frac{1}{n} \log P_{n,x}\left[X_j \in A \text{ for } 1 \leq j \leq n\right] = -\inf_{\mu:\mu(A)=1} I(\mu).$$

There are two ways of looking at $I(\mu)$. For the upper bound, if we can estimate for $V \in \mathcal{V}$

$$\limsup_{n\to\infty} \frac{1}{n} \log E^P[V(X_1) + \cdots + V(X_n)] \leq \lambda(V);$$

then, by standard Chebyshev-type estimate,

$$-I(\mu) = \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log P\left[\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \in N(\mu,\delta)\right] \leq -\sup_{V\in\mathcal{V}}\left[\int V(x)d\mu(x) - \lambda(V)\right].$$

Of course, when the state space is finite, $\int V(x)d\mu(x) = \sum_x V(x)\mu(x)$. Notice that $\lambda(V + c) = \lambda(V) + c$ for any constant c. Therefore,

$$I(\mu) \geq \sup_{\substack{V\in\mathcal{V},\\ \lambda(V)=0}} \int V(x)d\mu(x). \tag{4.2}$$

It is not hard to construct V such that $\lambda(V) = 0$.

LEMMA 4.2. *Suppose $u : \mathcal{X} \to \mathbb{R}$ satisfies $C \geq u(x) \geq c > 0$ for all x. Then, uniformly in x and n,*

$$\frac{c}{C} \leq E_x\left[\exp\left[\sum_{i=1}^{n} V(X_i)\right]\right] \leq \frac{C}{c} \tag{4.3}$$

where $V(x) = \log \frac{u(x)}{(\pi u)(x)}$ with $(\pi u)(x) = \sum_y \pi(x,y)u(y)$. In particular, $\lambda(V) = 0$ and (4.2) *holds.*

PROOF. An elementary calculation shows that

$$E_x\left[\prod_{i=1}^{n} \frac{u(X_i)}{(\pi u)(X_i)}\right] \leq \frac{1}{c} E_x\left[\prod_{i=1}^{n-1} \frac{u(X_i)}{(\pi u)(X_i)} u(X_n)\right] = \frac{u(x)}{c} \leq \frac{C}{c}$$

and

$$E_x\left[\prod_{i=1}^{n} \frac{u(X_i)}{(\pi u)(X_i)}\right] \geq \frac{1}{C} E_x\left[\prod_{i=1}^{n-1} \frac{u(X_i)}{(\pi u)(X_i)} u(X_n)\right] = \frac{u(x)}{C} \leq \frac{c}{C}. \qquad \square$$

To prove the converse, one "tilts" the measure P_x with transition probability $\pi(x,y)$ to a measure Q_x with transition probability $q(x,y) > 0$ that has μ as the invariant probability. Then, by the law of large numbers, if

$$A_{n,\mu,\delta} = \{(x_1,\ldots,x_n\} : \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \in N(\mu,\delta)\},$$

then

$$Q_x[A_{n,\mu,\delta}] \to 1$$

as $n \to \infty$. On the other hand, with $x_0 = x$, using Jensen's inequality,

$$\begin{aligned}
P_x[A_{n,\mu,\delta}] &= \int_{A_{n,\mu,\delta}} \left[\prod_{i=0}^{n-1} \frac{\pi(x_i, x_{i+1})}{q(x_i, x_{i+1})}\right] dQ_x \\
&= Q_x[A_{n,\mu,\delta}] \frac{1}{Q_x[A_{n,\mu,\delta}]} \int_{A_{n,\mu,\delta}} \left[\prod_{i=0}^{n-1} \frac{\pi(x_i, x_{i+1})}{q(x_i, x_{i+1})}\right] dQ_x \\
&\geq Q_x[A_{n,\mu,\delta}] \exp\left[-\frac{1}{Q_x[A_{n,\mu,\delta}]} \int_{A_{n,\mu,\delta}} \left[\sum_{i=0}^{n-1} \log \frac{q(x_i, x_{i+1})}{\pi(x_i, x_{i+1})}\right] dQ_x\right].
\end{aligned}$$

A simple application of the ergodic theorem yields

$$\liminf_{n\to\infty} \frac{1}{n} \log P_x[A_{n,\mu,\delta}] \geq -\sum_{x,y} \mu(x) q(x,y) \log \frac{q(x,y)}{\pi(x,y)}.$$

Since $q(\cdot\,,\cdot)$ can be arbitrary, provided $\mu q = q$; i.e., $\sum_x \mu(x) q(x,y) = \mu(y)$, we have

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log P_x\left[\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \in N(\mu,\delta)\right] \geq -\inf_{q:\mu q=\mu} \sum_{x,y} \mu(x) q(x,y) \log \frac{q(x,y)}{\pi(x,y)}.$$

In the next lemma, we will prove that, for any μ,

$$\sup_{u>0} \sum_x \mu(x) \log \frac{u(x)}{(\pi u)(x)} = \inf_{\mu:\mu q=\mu} \sum_{x,y} \mu(x) q(x,y) \log \frac{q(x,y)}{\pi(x,y)}.$$

With that, we will have the following theorem:

THEOREM 4.3. *Let $\pi(x,y) > 0$ be the transition probability of a Markov chain $\{X_i\}$ on a finite state space $\mathcal{X}$. Let $Q_{n,x}$ be the distribution of the empirical distribution $\gamma_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ of the Markov chain started from x on the space $\mathcal{M}(\mathcal{X})$ of probability measures on $\mathcal{X}$. Then, it satisfies a large deviation principle with rate function*

$$I(\mu) = \sup_{u>0} \sum_x \mu(x) \log \frac{u(x)}{(\pi u)(x)} = \inf_{\mu:\mu q=\mu} \sum_{x,y} \mu(x) q(x,y) \log \frac{q(x,y)}{\pi(x,y)}.$$

Since we have already proved the upper and lower bounds, we only need the following lemma:

LEMMA 4.4.

$$\sup_{u>0} \sum_x \mu(x) \log \frac{u(x)}{(\pi u)(x)} = \inf_{q:\mu q=\mu} \sum_{x,y} \mu(x) q(x,y) \log \frac{q(x,y)}{\pi(x,y)}.$$

PROOF. The proof depends on the following minimax theorem: Let $F(x,y)$ be a function defined on $C \times D$, which are convex sets in some nice topological vector space. For each fixed y, let F be lower semicontinuous and convex in x and, for each fixed x, upper semicontinuous and concave in y. Let either C or D be compact. Then,

$$\inf_{x\in C} \sup_{y\in D} F(x,y) = \sup_{y\in D} \inf_{x\in C} F(x,y).$$

We take $C = \{v : \mathcal{X} \to \mathbb{R}\}$, $D = \mathcal{M}(\mathcal{X} \times \mathcal{X})$, and for $v \in C, m \in D$,

$$\begin{aligned}
&\inf_{q:\mu q=\mu} \sum_{x,y} \mu(x)q(x,y)\log\frac{q(x,y)}{\pi(x,y)} \\
&= \inf_q \sup_v \Big[\sum_{x,y}[v(x)-v(y)]q(x,y)\mu(x) + \sum_{x,y}\mu(x)q(x,y)\log\frac{q(x,y)}{\pi(x,y)}\Big] \\
&= \sup_v \inf_q \Big[\sum_{x,y}[v(x)-v(y)]q(x,y)\mu(x) + \sum_{x,y}\mu(x)q(x,y)\log\frac{q(x,y)}{\pi(x,y)}\Big].
\end{aligned}$$

The function F is clearly linear and hence concave in v while being convex in q. Here the supremum over v of the first term is either 0 or infinite. It is 0 when $\mu q = \mu$ and infinite otherwise. The infimum over q is over all transition matrices $q(x, y)$. The infimum over q can be explicitly carried out and yields, for some u and v,

$$\log\frac{q(x,y)}{\pi(x,y)} = u(y) - v(x).$$

The normalization $\sum_y q(x, y) \equiv 1$ implies $e^{v(x)} = (\pi e^u)(x)$. The supremum over v turns into

$$\sup_{u>0}\sum_x \mu(x)\log\frac{u(x)}{(\pi u)(x)}. \qquad \square$$

REMARK 4.5. It is useful to note that the function $f \log f$ is bounded below by its value at $f = e^{-1}$ which is $-e^{-1}$. For any set A, any function f, and any probability measure μ,

$$\int_A f\log f d\mu \le \int f \log f d\mu + e^{-1}.$$

4.2. Large Deviations and the Principal Eigenvalues

Let $\{p(x, y)\}$, for $x, y \in \mathcal{X}$, be a matrix with strictly positive entries. Then there is a positive eigenvalue ρ such that it is simple, has a corresponding eigenvector with positive entries, and the remaining eigenvalues are of modulus strictly smaller than ρ. If $p(\cdot, \cdot)$ is a stochastic matrix, then $\sum_y p(x, y) = 1$; i.e., $\rho = 1$ and the corresponding eigenvector $u(x) \equiv 1$. In general, if $\sum p(x, y)u(y) = \rho u(x)$, then $\pi(x, y) = \frac{p(x,y)u(y)}{\rho u(x)}$ is a stochastic matrix. An elementary calculation yields

$$\sum_y p^{(n)}(x,y)u(y) = \rho^n u(x),$$

and consequently,

$$\frac{\inf_x u(x)}{\sup_x u(x)}\rho^n \le \inf_x \sum_y p^{(n)}(x,y) \le \sup_x \sum_y p^{(n)}(x,y) \le \frac{\sup_x u(x)}{\inf_x u(x)}\rho^n.$$

Combined with the recurrence relation

$$p^{(n+1)}(x,y) = \sum_z p^{(n)}(x,z)p(z,y),$$

it is easy to obtain a lower bound

$$p^{(n+1)}(x,y) \ge \inf_{z,y} p(z,y) \inf_x \sum_z p^{(n)}(x,z) \ge \inf_{z,y} p(z,y)\frac{\sup_x u(x)}{\inf_x u(x)}\rho^n.$$

In any case, there are constants C, c such that

$$c\rho^n \le p^{(n)}(x, y) \le C\rho^n;$$

$\rho = \rho(p(\cdot))$ is the spectral radius of $p(\cdot, \cdot)$. Of special interest will be the case when $p(x, y) = p_V(x, y) = \pi(x, y)e^{V(y)}$; i.e., p multiplied on the right by the diagonal matrix with entries $\{e^{V(x)}\}$. The following lemma is a simple computation, easily proved by induction on n.

LEMMA 4.6. *Let P_x be the Markov process with transition probability $\pi(x, y)$ starting from x. Then,*

$$E^{P_x}\left[\exp\left[\sum_{i=1}^{n} V(X_i)\right]\right] = \sum_y p_V^{(n)}(x, y)$$

where $p_V(x, y) = \pi(x, y)e^{V(y)}$.

It is now easy to connect large deviations and principal eigenvalues.

THEOREM 4.7. *The principal eigenvalue of a matrix $p(\cdot, \cdot)$ with positive entries is its spectral radius $\rho(p(\cdot, \cdot))$ and the large deviation rate function $I(\mu)$ for the distribution of the empirical distribution $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ on the space $\mathcal{M}(\mathcal{X})$ is the convex dual of*

$$\lambda(V) = \log \rho(p_V(\cdot, \cdot)).$$

REMARK 4.8. It is not necessary to demand that $\pi(x, y) > 0$ for all x, y. It is enough to demand only that for some $k \ge 1$, $\pi^{(k)}(x, y) > 0$ for all x, y. One can allow periodicity by allowing k to depend on x, y. These are straightforward modifications carried out in the study of Markov chains.

4.3. More General State Spaces

It is natural to assume that the state space $\mathcal{X}$ is more general than a finite set. We could assume that it is a compact metric space and that the transition probability is given by $\pi(x, dy)$. One needs to assume that the map $x \to \pi(x, \cdot)$ is weakly continuous as a map from $\mathcal{X} \to \mathcal{M}(\mathcal{X})$, the space of probability measures on $\mathcal{X}$ with the topology of weak convergence. Then, the probability distribution $Q_{n,x}$ of $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ on $\mathcal{M}(\mathcal{X})$ starting from a point $x \in X$ is expected to have a large deviation behavior in the weak topology with the rate function

$$I(\mu) = \sup_{\substack{u \in C(\mathcal{X}), \\ \inf_x u(x) > 0}} \int_{\mathcal{X}} \log \frac{u(x)}{(\pi u)(x)}\, d\mu(x).$$

The upper bound proceeds the same way and the lower bound and can be easily proved if one assumes that $\pi(x, dy) = \pi(x, y)d\lambda(y)$ has a density $\pi(x, y)$ with respect to a reference measure λ and $\pi(x, y) \ge c > 0$. Actually, weaker conditions will suffice but will involve a little bit more work. The proof is not all that different from the finite state space case.

One can deal with Markov processes in continuous time with transition probabilities $\pi(t, x, dy)$, and we are looking at large deviations of $Q_{t,x}$, the distribution of $\frac{1}{t}\int_0^t \delta_{X(s)}\,ds$ on $\mathcal{M}(\mathcal{X})$ under a Markov process with transition probability π starting from the point x. The rate function now takes the form

$$I(\mu) = -\inf_{\substack{u \in C(\mathcal{X}) \cap \mathcal{D}, \\ \inf_x u(x) > 0}} \int_{\mathcal{X}} \frac{(\mathcal{L}u)(x)}{u(x)} d\mu(x)$$

where $\mathcal{L}$ is the infinitesimal generator for the semigroup $(T_t f)(x) = \int f(y)\pi(t, x, dy)$. One way of proving it is to approximate $\frac{1}{t}\int_0^t \delta_{X(s)} ds$ by $\frac{h}{t}\sum_{1\le j\le \frac{t}{h}} \delta_{X(jh)}$. Its distribution will have a rate function

$$I_h(\mu) = \sup_{\substack{u\in C(\mathcal{X}),\\ \inf_x u(x)>0}} \int_{\mathcal{X}} \log \frac{u(x)}{(\pi_h u)(x)}\, d\mu(x).$$

Since time has step size h, the proof reduces to showing

$$\lim_{h\to 0} \frac{I_h(\mu)}{h} = I(\mu).$$

Roughly speaking, $\pi_h = I + h\mathcal{L} + o(h)$ and

$$\log \frac{u}{\pi_h u} = -\log \frac{u + h\mathcal{L}u + o(h)}{u} = -h\frac{\mathcal{L}u}{u} + o(h).$$

One often has to deal with noncompact spaces. Since the lower bound is local, there is generally no problem with it. The upper bound requires some compactness or, if we do not have it, then the process should have strong positive recurrence that allows us to control time spent by the process outside a compact set. These types of results have been worked out in detail in many places. See the end of the chapter for references. We will not use them directly, but establish what we need in our special case.

4.4. Dirichlet Eigenvalues

Let $F \subset \mathcal{X}$. Our aim is to estimate, for a Markov chain P_x with transition probability $\pi(x, y)$ and starting from $x \in F$,

$$\begin{aligned} P_x\big[X_i \in F,\ i = 1, \dots, n\big] &= \sum_{x_1,\dots,x_n\in F} \pi(x, x_1) p(x_1, x_2) \cdots \pi(x_{n-1}, x_n) \\ &= \sum_y p_F^{(n)}(x, y) \end{aligned}$$

where $p_F(x, y) = \pi(x, y)$ if $x, y \in F$ and 0 otherwise. In other words, p_F is a substochastic matrix on F. In some sense, this corresponds to p_V where $V = 0$ on F and $-\infty$ on F^c. The spectral radius $\rho(F)$ of p_F has the property that for $x \in F$,

$$\lim_{n\to\infty} \frac{1}{n} \log P_x[X_i \in F,\ i = 1, \dots, n] = \log \rho(p_F).$$

In our case, it is a little more complicated because we have a ball of radius cn^α and we want our random walk in n steps to be confined to this ball. The set F of the previous discussion depends on n. The idea is if we scale space and time and use the invariance principle as our guide, this should be roughy the same as the probability that a Brownian motion with covariance $\frac{1}{d}I$ remains inside a ball of radius c during the time interval $0 \le t \le n^{1-2\alpha}$. We have done the Brownian rescaling by factors $n^{2\alpha}$ for time and n^α for space. This will have probability decaying like $\lambda_d(c)n^{1-2\alpha}$ where $\lambda_d(c) = \frac{\lambda_d}{c^2}$ is the eigenvalue of $\frac{\Delta}{2d}$ for the unit ball in $\mathbb{R}^d$ with Dirichlet boundary conditions. The volume of the ball of radius cn^α is $v_d c^d n^{d\alpha}$ and that is roughly the maximum number of lattice points that can be visited by a random walk confined to the ball of radius cn^α. The contribution from such paths is $\exp\big[-\nu v_d c^d n^{\alpha d} - \frac{\lambda_d}{c^2} n^{1-2\alpha}\big]$. It is clearly best to choose $\alpha = \frac{1}{d+2}$ so that we have

$$\liminf_{n\to\infty} \frac{1}{n^{\frac{d}{d+2}}} \log E\big[\exp[-\nu|D_n|]\big] \ge -\left[\nu v_d c^d + \frac{\lambda_d}{c^2}\right].$$

If we compute

$$\inf_{c>0}\left[\nu v_d c^d + \frac{\lambda_d}{c^2}\right] = k(d)\nu^{\frac{2}{d+2}},$$

then

$$\liminf_{n\to\infty} \frac{1}{n^{\frac{d}{d+2}}} \log E\big[\exp[-\nu|D_n|]\big] \ge -k(d)\nu^{\frac{2}{d+2}}.$$

We will first establish this lower bound rigorously and then prove the upper bound.

4.5. Lower Bound

We begin with a general inequality that we will use repeatedly. Let P, Q be two probability measures on some probability space. Assume that $Q \ll P$ and, with $\psi = \frac{dQ}{dP}$, we have $\int \psi \log \psi \, dP = H(Q,P) = H < \infty$. Then:

LEMMA 4.9. *For any function f,*

$$E^Q[f] \le \log E^P[e^f] + H.$$

Moreover, for any measurable set A,

$$Q(A) \le \frac{H+2}{\log \frac{1}{P(A)}}$$

and

$$P(A) \ge Q(A) \exp\left[-H - \frac{1}{Q(A)} \int |H - \log\psi| dQ\right]. \tag{4.4}$$

PROOF. It is a simple inequality to check that for any x and $y > 0$, $xy \le e^x + y\log y - y$. Therefore,

$$E^Q[f] = E^P[f\psi] \le E^P[e^f + \psi\log\psi - \psi] = E^P[e^f] + H - 1.$$

Replacing f by $f + c$,

$$E^Q[f] \le e^c E^P[e^f] + H - 1 - c.$$

With the choice of $c = -\log E^P[e^f]$, we obtain

$$E^Q[f] \le \log E^P[e^f] + H.$$

If we take $f = c\chi_A$,

$$Q(A) \le \frac{1}{c}\big[\log[e^c P(A) + 1 - P(A)] + H\big]$$

with $c = -\log P(A)$,

$$Q(A) \le \frac{H+2}{\log\frac{1}{P(A)}}.$$

Finally,

$$P(A) \ge \int_A e^{-\log\psi} dQ \ge Q(A)\frac{1}{Q(A)}\int_A e^{-\log\psi} dQ \ge Q(A)\exp\left[-\frac{1}{Q(A)}\int_A \log\psi \, dQ\right]$$

and

$$\frac{1}{Q(A)}\int_A \log\psi dQ = H - \frac{1}{Q(A)}\int_A [H - \log\psi] dQ \le H + \frac{1}{Q(A)}\int |H - \log\psi| dQ. \quad \square$$

LEMMA 4.10. *Let c, c_1, c_2 be constants, with $c_2 < c$. Then, for the random walk $\{P_x\}$,*

$$\begin{aligned}\lim_{\substack{n\to\infty,\\ n^{-\alpha}x_n\to x}} & P_{x_n}[X_i \in B(0, c\, n^\alpha)\ \forall\ 1 \le i \le c_1\, n^{2\alpha}, X_{c_1 n^{2\alpha}} \in B(0, c_2\, n^\alpha)] \\ &= Q_x[x(t) \in B(0,c)\ \forall\ t \in [0, c_1], x(c_1) \in B(0, c_2)] \\ &= f(x, c, c_1, c_2)\end{aligned}$$

where Q_x is Brownian motion with covariance $\frac{1}{d} I$.

This is just the invariance principle asserting the convergence of random walk to Brownian motion under suitable rescaling. The set of trajectories confined to a ball of radius c for time c_1 and that end up, at time c_1, inside a ball of radius c_2 is easily seen to be a continuity set for Brownian motion. The convergence is locally uniform and consequently

$$\lim_{n\to\infty} \inf_{x\in B(0,c_2 n^\alpha)} P_x[X_i \in B(0, c\, n^\alpha)\ \forall\ 1 \le i \le c_1\, n^{2\alpha}, X_{c_1 n^{2\alpha}} \in B(0, c_2\, n^\alpha)] = \inf_{x\in B(0,c_2)} f(x, c, c_1, c_2).$$

In particular, from the Markov property

$$P_0[X_i \in B(0, c\, n^\alpha)\ \forall 1 \le i \le n)] \ge \inf_{x\in B(0,c_2 n^\alpha)} P_x[X_i \in B(0, c\, n^\alpha)\ \forall\ 1 \le i \le c_1 n^{2\alpha}, X_{c_1 n^{2\alpha}} \in B(0, c_2\, n^\alpha)]^{\frac{n^{1-2\alpha}}{c_1}},$$

showing

$$\lim_{n\to\infty} \frac{1}{n^{1-2\alpha}} \log P_0[X_i \in B(0, c\, n^\alpha)\ \forall\ 1 \le i \le n)] \ge \inf_{x\in B(0,c_2)} \frac{1}{c_1} \log f(x, c, c_1, c_2).$$

Since the left-hand side is independent of c_1, we can let $c_1 \to \infty$.

LEMMA 4.11. *For any $c_2 < c$,*

$$\lim_{c_1\to\infty} \inf_{x\in B(0,c_2)} \frac{1}{c_1} \log f(x, c, c_1, c_2) \ge -\frac{\lambda_d}{c^2}.$$

PROOF. Because of the scaling properties of the Brownian motion, we can assume without loss of generality that $c = 1$. The idea of the proof is to construct a process Q_x with the following properties:

- The process Q_x is a diffusion with generator $\frac{1}{2d}\Delta + b(x)\cdot\nabla$, which is absolutely continuous with respect to P_x, the Brownian motion with generator $\frac{1}{2d}\Delta$.
- The first-order term $b(\cdot)$ is strong enough near the boundary that the paths under Q_x almost surely do not reach the boundary of the unit ball in finite time.
- There is an invariant density $g(x)$ for the process with generator $\frac{1}{2d}\Delta + b(x)\cdot\nabla$ that is the unique normalized positive solution of $\frac{1}{2d}\Delta g - \nabla\cdot b(x)g(x) = 0$.
- The Radon-Nikodym derivative $\frac{dQ_x}{dP_x}$ on the σ-field of events up to time t is given by

$$\log \frac{dQ_x}{dP_x} = \psi(t) = d\int_0^t b(x(s))\cdot dx(s) - \frac{d}{2}\int_0^t |b(x(s))|^2\, ds;$$

the ergodic theorem implies

$$\lim_{t\to\infty} E^{Q_x}\left[\left|\frac{\psi(t)}{t} - \frac{d}{2}\int\limits_{B(0,1)} |b(x)|^2 g(x)dx\right|\right] = 0,$$

and the convergence is uniform over $x \in B(0,r)$ for $r < 1$.

Let A_t be a set measurable with respect to the σ-field unto time t. We can then use (4.4) to conclude that if, for some $r < 1$, $\inf_{x\in B(0,r)} Q_x(A_t) \ge q > 0$, then

$$\inf_{x\in B(0,r)} \frac{1}{t}\log P_x[A_t] \ge -\frac{d}{2}\int\limits_{B(0,1)} |b(x)|^2 g(x)dx.$$

We now produce such a $b(x)$. Let $-\lambda_d$ be the ground state eigenvalue of $\frac{1}{2d}\Delta$ on $B(0,1)$ with Dirichlet boundary conditions and with a corresponding eigenfunction $\phi(x) > 0$ inside $B(0,1)$. We take $b(x) = \frac{1}{d}\frac{\nabla\phi}{\phi}$. The function $u = \log\phi$ satisfies

$$\frac{1}{2d}\Delta u + \frac{1}{d}\frac{\nabla\phi}{\phi}\nabla u = \frac{1}{2d}\frac{|\nabla\phi|^2}{\phi^2} - \lambda_d.$$

Therefore $-u(x(t)) - \lambda_d t$ is a supermartingale and u is bounded below. It cannot blow up in a finite time. Hence, the zero set of ϕ is never hit in finite time. This shows the absolute continuity of Q_x with respect to P_x. One can verify that with $g = \phi^2$

$$\frac{1}{2d}\Delta g - \frac{1}{2d}\nabla\cdot\frac{\nabla g}{g}\cdot g = 0.$$

Finally,

$$\frac{d}{2}\int |b(x)|^2 g(x)dx = \frac{1}{2d}\int |\nabla\phi|^2 dx = -\frac{1}{2d}\int \phi\Delta\phi dx = -\lambda_d\int \phi^2\, dx = -\lambda_d.$$

□

4.6. Upper Bounds

Let us map our random walk on $\mathbb{Z}^d$ into a walk on the unit torus by rescaling $z \to \frac{z}{N} \in \mathbb{R}^d$ and then on to the torus $\mathbb{T}^d$ by sending each coordinate x_i to x_i (mod) 1. The transition probabilities $\Pi_N(x,dy)$ are $x \to x \pm \frac{e_i}{N}$ with probability $\frac{1}{2d}$. Let $u > 0$ be a smooth function on the torus. Then,

$$\begin{aligned}\log\frac{u}{\Pi u}(x) &= -\log\frac{\frac{1}{2d}\sum_{i,\pm} u(x\pm\frac{e_i}{N})}{u(x)}\\ &= -\log\left[1 + \frac{\frac{1}{2d}\sum_{i,\pm}[u(x\pm\frac{e_i}{N}) - u(x)]}{u(x)}\right]\\ &\simeq -\frac{1}{2dN^2}\frac{\Delta u}{u}(x) + o(N^{-2}).\end{aligned}$$

Denoting the distribution of the scaled random walk on the torus starting from x by $P_{N,x}$, we first derive a large deviation principle for the empirical distribution

$$\alpha(n,\omega) = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$$

where $X_i \in \mathbb{T}^d$ are already rescaled. We denote by $Q_{n,N,x}$ the distribution of α_n on $\mathcal{M}(\mathbb{T}^d)$. If $n \to \infty$, $N \to \infty$, and $k = \frac{n}{N^2} \to \infty$, then we have a large deviation principle for $Q_{n,N,x}$ on $\mathcal{M}(\mathbb{T}^d)$.

THEOREM 4.12. *For any closed set $C \in \mathcal{M}(\mathbb{T}^d)$,*

$$\limsup_{\substack{N\to\infty,\\ k=\frac{n}{N^2}\to\infty}} \frac{1}{k} \log Q_{n,N,x}(C) \le -\inf_{\mu\in C} I(\mu)$$

and, for any open set $G \in \mathcal{M}(\mathbb{T}^d)$,

$$\liminf_{\substack{N\to\infty,\\ k=\frac{n}{N^2}\to\infty}} \frac{1}{k} \log Q_{n,N,x}(G) \ge -\inf_{\mu\in G} I(\mu)$$

where, if $d\mu = f dx$ and $\nabla\sqrt{f} \in L_2(\mathbb{T}^d)$,

$$I(\mu) = \frac{1}{8d}\int \frac{\|\nabla f\|^2}{f} dx = \frac{1}{2d}\int \|\nabla\sqrt{f}\|^2\, dx.$$

Otherwise, $I(\mu) = +\infty$.

PROOF. *Lower Bound.* We need to add a bias so that the invariant probability for the perturbed chain on the imbedded lattice $\frac{1}{N}\mathbf{Z}_N^d$ is close to a distribution with density f on the torus. We take $v(x) = \sum_y \pi(x,y) f(y)$ and the transition probability to be

$$\hat{\pi}(x,y) = \pi(x,y)\frac{f(y)}{v(x)}.$$

Then, $\sum_x v(x)\hat{\pi}(x,y) = \frac{1}{2d}\sum_{i,\pm} f(x \pm e_i) = v(x)$; so, the invariant probability is $\frac{v(x)}{\sum_x v(x)}$. It is not hard to prove (see exercise) that if $N \to \infty$ and $\frac{n}{N^2} \to \infty$, then

$$\frac{1}{n}\sum_{i=1} V(X_i) \to \int V(x) f(x) dx$$

in probability under $\widehat{Q}_{n,N,x}$, provided V is a bounded continuous function and $\int f(x)dx = 1$. So, the probability of large deviation will have a lower bound with the rate function computed from the entropy

$$\begin{aligned} n\sum_{x,y} v(x)\pi(x,y)\frac{f(y)}{v(x)}\log\frac{f(y)}{v(x)} &\simeq \frac{n}{N^2}\sum_{x,y}\pi(x,y)f(y)\log\frac{f(y)}{\sum_y \pi(x,y)f(y)} \\ &\simeq \frac{n}{N^2}\frac{1}{8d}\int \frac{\|\nabla f\|^2}{f}\, dx. \end{aligned}$$

Upper Bound. We start with the identity

$$\begin{aligned} &E^{P_{n,x}}\left[\exp\left[-\sum_{j=1}^{n}\log\frac{\frac{1}{2d}\sum_{i,\pm} u\big(X_j \pm \frac{1}{N}e_i\big)}{u(X_j)}\right]\right] \\ &= E^{Q_{n,N,x}}\left[\exp\left[-n\int \log\frac{\frac{1}{2d}\sum_{i,\pm} u\big(x \pm \frac{1}{N}e_i\big)}{u(x)} d\alpha\right]\right] \\ &= 1. \end{aligned}$$

Since

$$N^2 \log \frac{\frac{1}{2d}\sum_{i,\pm} u\big(x \pm \frac{1}{N}e_i\big)}{u(x)} \to \frac{1}{2d}\frac{\Delta u}{u}(x)$$

uniformly over $x \in \mathbb{T}^d$, it follows that

$$\lim_{\delta\to 0} \limsup_{\substack{N\to\infty,\\ k=\frac{n}{N^2}\to\infty}} \frac{1}{k} Q_{n,N,x}[B(\alpha,\delta)] \le \int \Big[\frac{\Delta u}{u}(x)\Big] d\alpha.$$

Optimizing over u, we obtain

$$\lim_{\delta\to 0} \limsup_{\substack{N\to\infty,\\ k=\frac{n}{N^2}\to\infty}} \frac{1}{k} Q_{n,N,x}[B(\alpha,\delta)] \le -I(\alpha)$$

where

$$I(\alpha) = \frac{1}{2d}\sup_{u>0}\int\Big[-\frac{\Delta u}{u}(x)\Big]d\alpha. \tag{4.5}$$ □

A routine covering argument, of closed sets that are really compact in the weak topology, by small balls completes the proof of the upper bound. It is easy to see that $I(\alpha)$ is convex, lower semicontinuous, and translation invariant. By replacing α by $\alpha_\delta = (1-\delta)\alpha * \phi_\delta + \delta$, we see that $I(\alpha_\delta) \le I(\alpha)$, $\alpha_\delta \to \alpha$ as $\delta \to 0$, and α_δ has a nice density f_δ. It is therefore sufficient to prove that for smooth, strictly positive f,

$$\frac{1}{2d}\sup_{u>0}\int\Big[-\frac{\Delta u}{u}(x)\Big]f(x)dx = \frac{1}{8d}\int\frac{\|\nabla f\|^2}{f}\,dx.$$

Writing $u = e^h$, the calculation reduces to

$$\begin{aligned}\frac{1}{2d}\sup_h\Big[\int[-\Delta h - |\nabla h|^2]f(x)dx\Big] &= \frac{1}{2d}\sup_h\Big[\int[\langle\nabla h,\nabla f\rangle dx - \int|\nabla h|^2]f(x)dx\Big]\\ &= \frac{1}{8d}\int\frac{\|\nabla f\|^2}{f}\,dx.\end{aligned}$$

One inequality is just obtained by Schwarz and the other by the choice of $h = \sqrt{f}$.

EXERCISE 4.13. Let Π_h be transition probabilities of a Markov chain $P_{h,x}$ on a compact space $\mathcal{X}$ such that $\frac{1}{h}[\Pi_h - I] \to \mathcal{L}$ where $\mathcal{L}$ is a nice diffusion generator with a unique invariant distribution μ. Then, for any continuous function $f : \mathcal{X} \to \mathbb{R}$, for any $\epsilon > 0$

$$\limsup_{\substack{h\to 0,\\ nh\to\infty}} \sup_x P_{h,x}\bigg[\bigg|\frac{1}{n}\sum_{j=1}^n f(X_j) - \int f(x)d\mu(x)\bigg| \ge \epsilon\bigg] = 0.$$

Hint. If we denote by $\mu_{n,h,x}$ the distribution $\frac{1}{n}\sum_{j=1}^n \Pi^j(x,\cdot)$, then verify that any limit point of $\mu_{n,h,x'}$ as $h \to 0$, $nh \to \infty$, and $x' \to x$ is an invariant distribution of $\mathcal{L}$ and, therefore, is equal to μ. This implies

$$\lim_{\substack{h\to 0,\\ nh\to\infty}} \mu_{n,h,x} = \mu$$

uniformly over $x \in \mathcal{X}$. The ergodic theorem is a consequence of this. If $\int V(x)d\mu(x) = 0$, then ignoring the n diagonal terms,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n^2} E_x[(V(X_1) + \cdots V(X_n))^2] \\
&\leq 2 \lim_{n\to\infty} \frac{1}{n} \sup_{x,i} \big| V(x) E[V(X_{i+1}) + \cdots V(X_n) | X_i = x] \big| \\
&= 0.
\end{aligned}$$

4.7. The Role of Topology

We are really interested in the number of sites visited. If α_n is the empirical distribution, then we can take the convolution $g_{n,N,\omega}(x) = \alpha_n(dx) * N^d \mathbb{1}_{C_N}(x)$ where C_N is the cube of size $\frac{1}{N}$ centered at the origin. Then,

$$|\{x : g_{n,N,\omega}(x) > 0\}| = \frac{1}{N^d} |D_n(\omega)|$$

where $|D_n(\omega)|$ is the cardinality of the set $D_n(\omega)$ of the sites visited. We are looking for a result of the form

THEOREM 4.14.

$$\limsup_{\substack{k\to\infty,\\ N\to\infty}} \frac{1}{k} \log E^{Q_{kN^2,N}}\big[\exp[-\nu|\{x : g(x) > 0\}|]\big] \leq$$
$$- \inf_{\substack{g\geq 0,\\ \int g=1}} \left[\nu|\{x : g(x) > 0\}| + \frac{1}{8d} \int \frac{|\nabla g|^2}{g}\, dx \right]$$

where $Q_{kN^2,N}$ is the distribution of $g_{n,N,\omega}(x) = \alpha_n(dx) * N^d \mathbb{1}_{C_N}(x)$ on $L_1(\mathbb{T}^d)$ induced by the random walk with $n = kN^2$ starting from the origin.

We do have a large deviation result for $Q_{kN^2,N}$ with rate function $I(g) = \frac{1}{8d} \int \frac{|\nabla g|^2}{g} dx$. We proved it for the distribution of $\alpha_{n,\omega}$ on $\mathcal{M}(\mathbb{T}^d)$ in the weak topology. In the weak topology, the map $\alpha \to \alpha * N^d \mathbb{1}_{C_N}(x)$ of $\mathcal{M}(\mathbb{T}^d) \to \mathcal{M}(\mathbb{T}^d)$ is uniformly close to identity that the large deviation principle holds for $Q_{kN^2,N}$ that are supported on $L_1(\mathbb{T}^d) \subset \mathcal{M}(\mathbb{T}^d)$ in the weak topology.

If we had the large deviation result for $Q_{kN^2,N}$ in the L_1 topology, we will be in good shape. The function $F(g) = |\{x : g(x) > 0\}|$ is lower semicontinuous in L_1. We can use Theorem 2.6. It is not hard to prove the following general fact: Let $\delta > 0$ be arbitrary. Let $\phi_\delta(x)$ be an approximation of the identity. Then $g_\delta = g * \phi_\delta$ is a map of $L_1 \to L_1$. This is a continuous map from $L_1 \subset \mathcal{M}(\mathbb{T}^d)$ with the weak topology to L_1 with the strong topology. If we denote the image of $Q_{kN^2,N}$ by $Q^\delta_{kN^2,N}$, it is easy to deduce the following:

THEOREM 4.15. *For any $\delta > 0$, the distributions $Q^\delta_{kN^2,N}$ satisfy a large deviation principle as $k \to \infty$ and $N \to \infty$, so that for $C \in L_1(\mathbb{T}^d)$ that are closed we have*

$$\limsup_{\substack{k\to\infty,\\ N\to\infty}} \frac{1}{k} \log Q^\delta_{kN^2,N}[C] \leq \inf_{g : g_\delta \in C} I(g)$$

and for G that are open

$$\liminf_{\substack{k\to\infty,\\ N\to\infty}} \frac{1}{k} \log Q^\delta_{kN^2,N}[G] \geq \inf_{g : g_\delta \in G} I(g).$$

But, we need the results for $\delta = 0$, and this involves interchanging the two limits. This can be done through the superexponential estimate.

THEOREM 4.16.
$$\limsup_{\delta\to 0}\limsup_{\substack{k\to\infty,\\ N\to\infty}} \frac{1}{k}\log Q^{\delta}_{kN^2,N}[g : \|g_\delta - g\|_1 \ge \epsilon] \le -\infty.$$

Once we have the above result, it is not difficult to verify that the rate function for $Q_{kN^2,N}$ in L_1 is also $I(g)$, and we would complete our proof. We will outline first the idea of the proof and reduce it to some lemmas. Denoting $N^d \mathbb{1}_{C_N}$ by χ_N, the quantity

$$\begin{aligned}\left\|\alpha * N^d \mathbb{1}_{C_N} * \phi_\delta - \alpha * N^d \mathbb{1}_{C_N}\right\|_1 &= \sup_{V:|V(x)|\le 1}\left|\int V * \chi_N * \phi_\delta\, d\alpha - \int V * \chi_N\, d\alpha\right| \\ &= \sup_{V\in K_N}\left|\int V * \phi_\delta\, d\alpha - \int V\, d\alpha\right|\end{aligned}$$

where K_N, the image of $V : |V(x)| \le 1$ under convolution with χ_N, is a compact set in $C(\mathbb{T}^d)$. Given $\epsilon > 0$, it can be covered by a finite number $\tau(N, \epsilon)$ of balls of radius $\frac{\epsilon}{2}$. Let us denote the set of centers by $D_{N,\epsilon}$, whose cardinality is $\tau(N, \epsilon)$. Then we can estimate

$$Q_{kN^2,N}[g : \|g_\delta - g\|_1 \ge \epsilon] \le \tau(N,\epsilon) \sup_{V\in D_{N,\epsilon}} Q_{kN^2,N}\left[\left|\int (V * \phi_\delta - V)d\alpha\right| \ge \frac{\epsilon}{2}\right].$$

We begin by estimating the size of $\tau(N, \epsilon)$. The modulus continuity of any $W \in D_{N,\epsilon}$ satisfies
$$|W(x) - W(y)| \le \int |\chi(x - z) - \chi(y - z)|dz \le \frac{\epsilon}{4}$$
provided $|x - y| \le \eta/N$ for some $\eta = \eta(\epsilon)$. We can chop the torus into $[N/\eta]^d$ subcubes and divide each interval $[-1, 1]$ into $4/\epsilon$ subintervals. Then, balls around $[4/\epsilon]^{[N/\eta]^d}$ simple functions will cover $D_{N,\epsilon}$. So, we have proved

LEMMA 4.17.
$$\log \tau(N, \epsilon) \le C(\epsilon)N^d.$$

Let $J_\delta = \{W : W = V * \phi_\delta - V\}$ and $\|V\|_\infty \le 1$. We now try to get a uniform estimate on

$$Q_{kN^2,N}\left[\left|\int W d\alpha\right| \ge \frac{\epsilon}{2}\right] = P_{N,x}\left[\frac{1}{kN^2}\sum_{i=1}^{kN^2} W(X_i) \ge \frac{\epsilon}{2}\right]$$

where $P_{N,x}$ is the probability measure that corresponds to the random walk on $\mathbf{Z}^d_N$ starting from x at time 0. We denote by

$$\Theta(k, N, \lambda, \delta) = \sup_{x\in\mathbf{Z}^d_N}\sup_{W\in J_\delta} E^{P_{N,x}}\left[\exp\left[\frac{\lambda}{N^2}\sum_{i=1}^{kN^2} W(X_i)\right]\right]$$

If we can show that
$$\lim_{\delta\to 0}\lim_{\substack{k\to\infty,\\ N\to\infty}} \frac{1}{k}\log\Theta(k, N, \lambda, \delta) = 0$$
for every λ, then
$$\frac{1}{k}\log Q_{kN^2,N}\left[\left|\int W d\alpha\right| \ge \frac{\epsilon}{2}\right] \le -\left[\lambda\frac{\epsilon}{2} - \frac{1}{k}\log\Theta(k, N, \lambda, \delta)\right]$$

and

$$\limsup_{\delta\to 0}\limsup_{\substack{k\to\infty,\\ N\to\infty}}\sup_{W\in J_\delta}\frac{1}{k}\log Q_{kN^2,N}\left[\left|\int W d\alpha\right|\geq\frac{\epsilon}{2}\right]\leq -\lambda\frac{\epsilon}{2}.$$

Since $\lambda>0$ is arbitrary, it would follow that

$$\limsup_{\delta\to 0}\limsup_{\substack{k\to\infty,\\ N\to\infty}}\sup_{W\in J_\delta}\frac{1}{k}\log Q_{kN^2,N}\left[\left|\int W d\alpha\right|\geq\frac{\epsilon}{2}\right]=-\infty.$$

Finally,

LEMMA 4.18. *For any $\lambda>0$,*

$$\lim_{\delta\to 0}\lim_{\substack{k\to\infty,\\ N\to\infty}}\frac{1}{k}\log\Theta(k,N,\lambda,\delta)=0.$$

PROOF. We note that for any Markov chain, for any V

$$\log\sup_x E^{P_x}\left[\sum_{i=1}^{n}V(X_i)]\right]$$

is subadditive and so it is enough to prove

$$\lim_{\delta\to 0}\lim_{N\to\infty}\inf_k\frac{1}{k}\log\Theta(k,N,\lambda,\delta)=0.$$ □

We can let $k\to t$ (keeping $n=N^2k$ an integer) and consider the limit

$$\widehat{\Theta}(t,\lambda,\delta)=\sup_{x\in\mathbb{T}^d}\sup_{W\in J_\delta}E^{P_x}\left[\exp\left[\lambda\int_0^t W(\beta(s))ds\right]\right]$$

where P_x is the distribution of Brownian motion with covariance $\frac{1}{d}I$ on the torus $\mathbb{T}^d$. Since the space is compact and the Brownian motion is elliptic, the transition probability density has a uniform upper and lower bound for $t>0$, and this enables us to conclude that

$$\limsup_{\delta\to 0}\limsup_{t\to\infty}\log\frac{1}{t}\widehat{\Theta}(t,\lambda,\delta)=0,$$

provided we show that for any $\lambda>0$,

$$\limsup_{\delta\to 0}\sup_{W\in J_\delta}\sup_{\substack{\|f\|_1=1,\\ f\geq 0}}\left[\lambda\int Wf\,dx-\frac{1}{8d}\int\frac{|\nabla f|^2}{f}dx\right]=0.$$

But,

$$\left|\int Wf\,dx\right|=\left|\int(V*\phi_\delta-V)f\,dx\right|\leq\int V|f_\delta-f|dx\leq\|f_\delta-f\|_1.$$

On the other hand, in the variational formula we can limit ourselves to f with $\int\frac{\|\nabla f\|^2}{f}dx\leq 8\lambda g$. But that set is compact in L_1, and therefore for any $C<\infty$,

$$\lim_{\delta\to 0}\sup_{f:\int\frac{\|\nabla f\|^2}{f}dx\leq C}\|f_\delta-f\|_1=0.$$

4.8. Finishing Up

We have now shown that

$$\frac{1}{n^{\frac{d}{d+2}}}\log E\big[\exp[-\nu|D_n|]\big] \le -\inf_{\substack{f\ge 0,\\ \|f\|_1=1}}\left[\nu|\text{supp}\, f| + \frac{1}{8d}\int_{\mathbb{T}_\ell^d}\frac{\|\nabla f\|^2}{f}\,dx\right].$$

The torus $\mathbb{T}_\ell^d$ can be of any size ℓ. We will next show that we can let $\ell\to\infty$ and obtain

$$\lim_{\ell\to\infty}\inf_{\substack{f\ge 0,\\ \|f\|_1=1}}\left[\nu|\text{supp}\, f| + \frac{1}{8d}\int_{\mathbb{T}_\ell^d}\frac{\|\nabla f\|^2}{f}\,dx\right] = \inf_r\left[\nu v_d r^d + \frac{\lambda_d}{r^2}\right].$$

Here v_d is the volume of the unit ball in $\mathbb{R}^d$ and λ_d is the first eigenvalue of $-\frac{1}{2d}\Delta$ in the unit ball of $\mathbb{R}^d$ with Dirichlet boundary condition. One side of this, namely

$$\limsup_{\ell\to\infty}\inf_{\substack{f\ge 0,\\ \|f\|_1=1}}\left[\nu|\text{supp}\, f| + \frac{1}{8d}\int_{\mathbb{T}_\ell^d}\frac{\|\nabla f\|^2}{f}\,dx\right] \le \inf_r\left[\nu v_d r^d + \frac{\lambda_d}{r^2}\right]$$

is obvious, because, if $\ell > 2r$, the ball can be placed inside the torus without distortion. For the other side, given a periodic f on $\mathbb{T}_\ell^d$ supported on a set of certain volume, it has to be transplanted as a function with compact support on $\mathbb{R}^d$ without increasing either the value of $\int_{\mathbb{T}_\ell^d}\frac{\|\nabla\|^2}{f}dx$ or the volume of the support of f by more than a negligible amount, more precisely, by an amount that can be made arbitrarily small, if ℓ is large enough. We do a bit of surgery by opening up the torus. Cut out the set $\bigcup_{i=1}^d |x_i|\le 1$. This is done by multiplying $f = g^2$ by $\prod(1-\phi(x_i))$ where $\phi(\cdot)$ is a smooth function with $\phi(x)=1$ on $[-1,1]$ and 0 outside $[-2,2]$. It is not hard to verify that if $\int_{\bigcup_i\{x:|x_i|\le 2\}}[g^2+\|\nabla g\|^2]\,dx$ is small, then $[g\prod_{i=1}^d(1-\phi(x_i))]^2$ normalized to have integral 1 works. While $A=\bigcup_i\{x: |x_i|\le 2\}$ may not work, we can always find some translate of it that will work because, for any f,

$$\ell^{-d}\int_{\mathbb{T}_\ell^d}\left[\int_{A+x} f(y)dy\right]dx = \ell^{-d}|A|\int f\,dx.$$

Now, we end up with

$$\inf_{\substack{f\ge 0,\\ \|f\|_1=1}}\left[\nu|\text{supp}\, f| + \frac{1}{8d}\int_{\mathbb{R}^d}\frac{\|\nabla f\|^2}{f}dx\right] = \inf_{G\subset\mathbb{R}^d}[\nu|G|+\lambda_d(G)] = \inf_r\left[\nu v_d r^d + \frac{\lambda_d}{r^2}\right],$$

the last step following from an isoperimetric inequality that allows us to limit G to spheres.

4.9. Remarks

The theory of large deviations from the ergodic averages of Markov processes started with the work of Donsker and Varadhan for Brownian motion [**5**], and they generalized it over the next several years to more general situations [**3, 4, 7**]. While the upper bound is easily proved for Feller processes on a compact state space, one needs in addition a strong positive recurrence condition to guarantee an upper bound. The lower bound being local needs some sort of Doeblin or Harris condition. The theory was developed with different applications in mind. The polaron problem needed a slightly different formulation [**11**]. The problem with the number of distinct sites visited required strengthening the topology and was carried out in [**9**]. There was related work carried out independently by Jurgen Gertner [**14**]. In one dimension where local time exists for Brownian motion, one can prove a large-deviation result for the local time in the uniform topology [**8**].

https://doi.org/10.1090//cln/027/05

CHAPTER 5

Hydrodynamic Scaling

5.1. From Classical Mechanics to Euler Equations

The basic example of hydrodynamical scaling is naturally hydrodynamics itself. Let us start with a collection of $N \simeq \bar{\rho}\ell^3$ classical particles in a large periodic cube Λ_ℓ of side ℓ in $\mathbb{R}^3$. The motion of the particles is governed by the equations of motion of a classical Hamiltonian dynamical system with energy given by

$$H(p\,,q) = \frac{1}{2}\sum_{i=1}^{N} \|p_i\|^2 + \frac{1}{2}\sum_{i\neq j} V(q_i - q_j). \tag{5.1}$$

Here, $q_i \in \Lambda_\ell$ is the position of the i^{th} particle, $p_i \in \mathbb{R}^3$ is its velocity, and $k = 1, 2, 3$ refers to the three components of position or velocity. The repulsive potential $V \geq 0$ is even and has compact support in $\mathbb{R}^3$. The interaction, in particular, is assumed to be short range. The equations of motion are

$$\begin{cases} \dfrac{dq_i^k}{dt} = \dfrac{\partial H(p\,,q)}{\partial p_i^k} = p_i^k, \\ \dfrac{dp_i^k}{dt} = -\dfrac{\partial H(p\,,q)}{\partial q_i^k} = -\sum_{i=1}^{N} V_k(q_i - q_j), \end{cases} \tag{5.2}$$

where

$$V_k(q) = \frac{\partial V(q)}{\partial q^k}$$

is the gradient of V. The dynamical system has five conserved quantities: the total number N of particles, the total momenta $\sum_{i=1}^{N} p_i^k$ for $k = 1, 2, 3$, and the total energy $H(p, q)$ given by (5.1). The hydrodynamic scaling in this context consists of rescaling space and time by a factor of ℓ. The rescaled space is the unit torus $\mathbb{T}^3$ in three dimensions. The macroscopic quantities to be studied correspond to conserved quantities. The first one of these is the density, and is measured by a function $\rho(t, x)$ of t and x. For each $\ell < \infty$, it is approximated by $\rho_\ell(t, x)$, defined by

$$\int_{\mathbb{T}^d} J(x)\rho_\ell(t, x)dx = \frac{1}{\ell^3}\sum_{i=1}^{N} J\left(\frac{q_i(\ell t)}{\ell}\right).$$

A straightforward differentiation using (5.2) yields

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{T}^d} J(x)\rho_\ell(t,x)dx &= \frac{d}{dt}\frac{1}{\ell^3}\sum_{i=1}^{N} J\left(\frac{q_i(\ell t)}{\ell}\right)\\ &= \frac{1}{\ell^3}\sum_{i=1}^{N}(\nabla J)\left(\frac{q_i(\ell t)}{\ell}\right)\cdot p_i(\ell t)\\ &\simeq \int_{\mathbb{T}^d}(\nabla J)(x)\cdot\rho_\ell(t,x)u(t,x)dx\end{aligned}$$

where $u = u^k$, $k = 1,2,3$, is the average velocity of the fluid. This introduces three other macroscopic variables, which represent three coordinates of the momenta that are conserved.

We can now write down the first of our five equations,

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho u) = 0. \tag{5.3}$$

To derive the next three equations, we start with test functions J and differentiate, again using (5.2) for $k = 1,2,3$,

$$\begin{aligned}\frac{d}{dt}\frac{1}{\ell^3}\sum_{i=1}^{N} J\left(\frac{q_i(\ell t)}{\ell}\right)p_i^k(\ell t) &= \frac{1}{\ell^3}\sum_{i=1}^{N} p_i^k(\ell t)(\nabla J)\left(\frac{q_i(\ell t)}{\ell}\right)\cdot p_i(\ell t)\\ &\quad - \frac{1}{\ell^2}\sum_{i=1}^{N}\sum_{j=1}^{N} J\left(\frac{q_i(\ell t)}{\ell}\right)V_k(q_i(\ell t) - q_j(\ell t)).\end{aligned} \tag{5.4}$$

The next step is rather mysterious and requires considerable explanation. Quantities

$$\sum_{i,j=1}^{N}\psi(q_i(t) - q_j(t)) \quad\text{and}\quad \sum_{i=1}^{N} p_i^k\, p_i^r$$

are not conserved. They depend on spacings between particles or velocities of individual particles that change in the microscopic time scale and hence do so rapidly in the macroscopic time scale. They can therefore be replaced by their space-time averages. By appealing to an "ergodic theorem," they can be replaced by ensemble averages with respect to their equilibrium distributions. The "ensemble" consists of an infinite collection of points $\{p_\alpha, q_\alpha\}$ in the phase space $\mathbb{R}^3\times\mathbb{R}^3$. There is a natural five-parameter family of measures $\mu_{\rho,u,T}$ that are invariant under spatial translation, as well as Hamiltonian dynamics. The points $\{p_\alpha\}$ are distributed according to a Gibbs distribution with density ρ and formal interaction

$$\frac{1}{2T}\sum_{\alpha,\beta} V(q_\alpha - q_\beta).$$

In other words, $\{q_\alpha\}$ is a point process obtained by taking an infinite volume limit of $N = \ell^3\rho$ particles distributed in a cube of side ℓ according to the joint density

$$\frac{1}{Z}\exp\left[-\frac{1}{2T}\sum_{i,j=1}^{N} V(q_i - q_j)\right]$$

where Z is the normalization constant. The velocities $\{p_\alpha\}$ are distributed independently of each other, as well as of $\{q_\alpha\}$, having a common three-dimensional Gaussian distribution with mean u and covariance TI. Assuming that the infinite volume limit exists in a reasonable sense, it will be a point process defined as an infinite volume Gibbs measure $\mu_{\rho,T}$. The velocities $\{p_\alpha\}$ will be an independent Gaussian ensemble $\nu_{u,T}$.

The first term in (5.4) involves sums of $p_i^k p_i^r$ that are replaced by their expectations in the Gaussian ensemble $u^k u^r + T\delta_{k,r}$.

If we now use the skew-symmetry of $V_k = \partial V/\partial q_k$, we can rewrite the second term of (5.4) as

$$\begin{aligned}
&-\frac{1}{2\ell^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(J\left(\frac{q_i(\ell t)}{\ell}\right) - J\left(\frac{q_j(\ell t)}{\ell}\right)\right)V_k(q_i(\ell t) - q_j(\ell t))\\
&\simeq -\frac{1}{2\ell^3}\sum_{i=1}^{N}\sum_{j=1}^{N} J_r\left(\frac{q_i(\ell t)}{\ell}\right)\big(q_i^r(\ell t) - q_j^r(\ell t)\big)V_k(q_i(\ell t) - q_j(\ell t))\\
&= \frac{1}{\ell^3}\sum_{i=1}^{N}\sum_{j=1}^{N} J_r\left(\frac{q_i(\ell t)}{\ell}\right)\psi_k^r(q_i(\ell t) - q_j(\ell t))\\
&\simeq \int_{\mathbb{T}^d} \frac{\partial J}{\partial x_r}(x)\mathbf{P}_k^r(\rho(t,x), T(t,x))dx
\end{aligned}$$

where $\mathbf{P}_k^r(\rho, T)$ is the "pressure" per unit volume in the Gibbs ensemble

$$\mathbf{P}_k^r(\rho, T) = \lim_{\ell\to\infty} E^{\mu_{\rho,T}}\left\{\frac{1}{\ell^3}\sum_{\substack{|q_\alpha|\le\ell,\\ |q_\beta|\le\ell}} \psi_k^r(q_\alpha - q_\beta)\right\}.$$

We now integrate by parts, remove the test function J, and obtain

$$\frac{d}{dt}(\rho u) + \nabla\cdot(\rho\, u\otimes u + \rho\, TI + \mathbf{P}(\rho, T)) = 0. \tag{5.5}$$

There is an equation of state that expresses the total energy per unit volume, e, as

$$e(\rho, u, T) = \frac{1}{2}\rho(|u|^2 + 3T) + f(\rho, T) \tag{5.6}$$

where $f(\rho, T)$, the potential energy per unit volume, is given by

$$f(\rho, T) = \lim_{\ell\to\infty} E^{\mu_{\rho,T}}\left\{\frac{1}{2\ell^3}\sum_{\substack{|q_\alpha|\le\ell,\\ |q_\beta|\le\ell}} V(q_\alpha - q_\beta)\right\}.$$

Although we will not derive it, there is a similar equation for $e(t,x)$ that is obtained by differentiating

$$\frac{d}{dt}\frac{1}{2\ell^3}\sum_{i=1}^{N} J\left(\frac{q_i(\ell t)}{\ell}\right)\left[|p_i(\ell t)|^2 + \sum_{j=1}^{N} V(q_i(\ell t) - q_j(\ell t))\right]$$

and proceeding in a similar fashion. It looks like

$$\frac{de}{dt} + \nabla \cdot [(e+T)u + \mathbf{P}(\rho\,, T)u] = 0. \tag{5.7}$$

The five equations (5.3), (5.5), and (5.7) in five variables (actually in six variables with one relation (5.6)) is a symmetrizable first-order system of nonlinear hyperbolic conservation laws. Given smooth initial data they have local solutions.

Rigorous derivation of these equations does not exist. The ergodic theory is definitely plausible, but hard to establish. The reason is that we have essentially an infinite system of ordinary differential equations representing a classical deterministic dynamical system, and the ergodicity properties are nearly impossible to establish in any general context. However, if we had some noise in the system, i.e., stochastic dynamics instead of deterministic dynamics, then we would be concerned with the ergodic theory of Markov processes of some kind, which is far more accessible. This will be the focus of our future lectures.

5.2. Simple Exclusion Processes

The simplest example is a system of noninteracting particles undergoing independent motions. For instance, we could have on $\mathbb{T}^3$, $L \simeq \bar{\rho}N^3$ particles, all behaving like independent Brownian particles. If the initial configuration of the L particles is such that the empirical distribution

$$\nu_0(dx) = \frac{1}{N^3}\sum_i \delta_{x_i}$$

has a deterministic limit $\rho_0(x)dx$, then the empirical distribution

$$\nu_t(dx) = \frac{1}{N^3}\sum_i \delta_{x_i(t)}$$

of the configuration at time t has a deterministic limit $\rho(t\,,x)dx$ as $N \to \infty$, and $\rho(t\,,x)$ can be obtained from $\rho_0(x)$ by solving the heat equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\Delta\rho$$

with the initial condition $\rho(0\,,x) = \rho_0(x)$. The proof is an elementary law-of-large-numbers argument involving a calculation of two moments. Let $f(x)$ be a continuous function on $\mathbb{T}$ and let us calculate for

$$U = \frac{1}{N^3}\sum_i f(x_i(t))$$

the expectation and variance of U, given the initial configuration $(x_1, \dots, x_L)$.

$$E(U) = \frac{1}{N^3}\sum \int_{\mathbb{T}^3} f(y)p(t\,,x_i\,,y)dy,$$

and an elementary calculation reveals that the expectation converges to the following limit:

$$\int_{\mathbb{T}^3}\int_{\mathbb{T}^3} f(y)p(t\,,x\,,y)\rho_0(x)dy\,dx = \int_{\mathbb{T}^3} f(y)\rho(t\,,y)dy.$$

The independence clearly provides a uniform upper bound of order N^{-3} for the variance and goes to 0. Of course, on $\mathbb{T}^3$, we could have had a process obtained by rescaling a random walk on a large torus of size N. Then, the hydrodynamic scaling limit would be a consequence of the central limit theorem for the scaling limit of a single particle and the law of large numbers resulting from the averaging over a large number of independently moving particles. The situation could be different if the particles interacted with each other.

The next class of examples are called simple exclusion processes. They make sense on any finite or countable set X, and for us X will be either the integer lattice $\mathbb{Z}^d$ in d dimensions or $\mathbf{Z}_N^d$ obtained from it as a quotient by considering each coordinate modulo N. At any given time, a subset of these lattice sites will be occupied by particles with at most one particle at each site. In other words, some sites are empty while others are occupied with one particle. The particles move randomly. Each particle waits for an exponential random time and then tries to jump from the current site x to a new site y. The new site y is picked randomly according to a finitely supported probability distribution $\pi(z)$ on $\mathbb{Z}^d - \{0\}$. In particular, $\sum_y \pi(y-x) = 1$ for every x. Of course, a jump to y is not always possible. If the site is empty the jump is possible and is carried out. If the site already has a particle, the jump cannot be carried out and the particle forgets about it and waits for another chance; i.e., waits for a new exponential waiting time.

If we normalize so that all waiting times have mean 1, the generator of the process can be written down as

$$(\mathcal{A} f)(\eta) = \sum_{x,y} \eta(x)(1-\eta(y))\pi(y-x)[f(\eta^{x,y}) - f(\eta)]$$

where η represents the configuration with $\eta(x) = 1$ if there is a particle at x, and $\eta(x) = 0$ otherwise. For each configuration η and a pair of sites x, y, the new configuration $\eta^{x,y}$ is defined by

$$\eta^{x,y}(z) = \begin{cases} \eta(y) & \text{if } z = x, \\ \eta(x) & \text{if } z = y, \\ \eta(z) & \text{if } z \neq x, y. \end{cases}$$

We will assume that the random walk with increments having distribution $\pi(\cdot)$ is irreducible, or equivalently, the sub-semigroup generated by $\{z : \pi(z) > 0\}$ is all of $\mathbb{Z}^d$. There are various possibilities:

$$\pi \text{ is symmetric, i.e., } \pi(z) = \pi(-z),$$

or, more generally,

$$\pi \text{ has mean zero, i.e., } \sum_z z\pi(z) = 0,$$

and finally

$$\sum_z z\,\pi(z) = m \neq 0.$$

We shall first concentrate on the symmetric case. Let us look at the function

$$V_J(\eta) = \sum J(x)\eta(x)$$

and compute

$$\begin{aligned}(\mathcal{A}V_J)(\eta) &= \sum_{x,y} \eta(x)(1-\eta(y))\pi(y-x)(J(y)-J(x))\\ &= \frac{1}{2}\sum_{x,y}[\eta(x)(1-\eta(y))-\eta(y)(1-\eta(x))]\pi(y-x)(J(y)-J(x))\\ &= \frac{1}{2}\sum_{x,y}(\eta(x)-\eta(y))\pi(y-x)(J(y)-J(x))\\ &= \sum_{x,y}\eta(x)\pi(y-x)(J(y)-J(x))\\ &= \sum_{x,y}\eta(x)[(\mathbf{P}-I)J](x)\\ &= V_{(\mathbf{P}-I)J}(\eta)\end{aligned}$$

where P is the matrix $\{\pi(y-x)\}$. The space of linear functionals is left invariant by the generator. It is not difficult to see that

$$E_\eta[V_J(\eta(t))] = V_{J(t)}(\eta)$$

where

$$J(t) = \exp[t(\mathbf{P}-I)]J$$

is the solution of

$$\frac{d}{dt}J(t\,,x) = (\mathbf{P}-I)J(t\,,x).$$

It is almost as if the interaction has no effect and, in fact, in the calculation of expectations of "one particle" functions, it clearly does not. Let us start with a configuration on $\mathbf{Z}_N^d$ and scale space by N and time by N^2. The generator becomes $N^2\mathcal{A}$, and the particles can be visualized as moving in a lattice imbedded in the unit torus $\mathbb{T}^d$, with spacing $\frac{1}{N}$, and becoming dense as $N \to \infty$.

Let J be a smooth function on $\mathbb{T}^d$. We consider the functional

$$\xi(t) = \frac{1}{N^d}\sum_x J\Big(\frac{x}{N}\Big)\eta_t(x),$$

and we can write

$$\xi(t) - \xi(0) = \int_0^t V_N(\eta(s))ds + M_N(t)$$

where

$$V_N(\eta) = (N^2\mathcal{A}V_J)(\eta) = V_{J_N}(\eta)$$

with

$$\begin{aligned}(J_N)(\theta) &= N^2\sum\Big[J\Big(\theta+\frac{z}{N}\Big)-J(\theta)\Big]\pi(z)\\ &\simeq \frac{1}{2}(\Delta_C J)(\theta)\end{aligned}$$

for $\theta \in \mathbb{T}^d$. Here, Δ_C refers to the Laplacian

$$\sum_{i,j} C_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}$$

with the covariance matrix C given by

$$C_{i,j} = \sum_z z_i z_j \pi(z).$$

$M_N(t)$ is a martingale and a very elementary calculation yields

$$E\{[M_N(t)]^2\} \le C\, t\, N^{-d},$$

essentially completing the proof in this case. Technically, the empirical distribution $\nu_N(t)$ is viewed as a measure on $\mathbb{T}^d$, and $\nu_N(\cdot)$ is viewed as a stochastic process with values in the space $\mathcal{M}(\mathbb{T}^d)$ of nonnegative measures on $\mathbb{T}^d$. In the limit, it lives on the set of weak solutions of the heat equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\Delta_C\, \rho,$$

and the uniqueness of such weak solutions for given initial density establishes the validity of the hydrodynamic limit.

Let us make the problem slightly more complicated by adding a small bias. Let $q(z)$ be an odd function with $q(-z) = -q(z)$, and we will modify the problem by making π depend on N in the form

$$\pi_N(z) = \pi(z) + \frac{1}{N} q(z).$$

Assuming that q is nonzero only when π is so, π_N will be an admissible transition probability for large enough N. A calculation yields that in the slightly modified model, referred to as the weakly asymmetric simple exclusion model, V_N is given by

$$V_N(\eta) \simeq V_{J_N}(\eta) + \frac{1}{N^d} \sum_x \eta(x)(1-\eta(y))q(y-x)\langle \nabla J(x), y-x\rangle.$$

If one thinks of $\rho(t\,,\theta)$ as the density of particles at the (macroscopic) time t and space θ, the first term clearly wants to have the limit

$$\int_{\mathbb{T}^d} \frac{1}{2}(\Delta_C J)(\theta)\rho(t\,,\theta)d\theta.$$

It is not so clear what to do with the second term. The “invariant” measures in this model are the Bernoulli measures with various densities ρ, and the “averaged” version of the second term should be

$$\int_{\mathbb{T}^d} \langle m, (\nabla J)(\theta)\rangle \rho(t\,,\theta)(1-\rho(t\,,\theta))d\theta$$

with

$$m = \sum_z z q(z).$$

The linear heat equation gets replaced by the nonlinear equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\Delta_C \rho - \nabla \cdot m\rho(1-\rho).$$

This requires justification that will be the content of our next section.

Let us now turn to the case where π has mean zero but is not symmetric. In this case,

$$V_N(\eta) = N^{2-d} \sum_{x,y} \eta(x)(1-\eta(y))\pi(y-x)\Big[J\Big(\frac{y}{N}\Big) - J\Big(\frac{x}{N}\Big)\Big],$$

and we get stuck at this point. If π is symmetric, as we saw, we gain a factor of N^{-2}. Otherwise, the gain is only a factor of N^{-1}, which is not enough.

We seem to end up with

$$N^{-d}\sum_x\sum_y \eta(x)\Big\langle\frac{1}{2}\Big[(\nabla J)\Big(\frac{x}{N}\Big)+(\nabla J)\Big(\frac{y}{N}\Big)\Big], N(1-\eta(y))(y-x)\pi(y-x)\Big\rangle =$$
$$\frac{1}{2N^d}\sum_x\Big(\nabla J\Big)\Big(\frac{x}{N}\Big)N\mathbf{\Psi}_x,$$

where

$$\begin{aligned}\mathbf{\Psi}_x &= \Big[\eta(x)\sum_z(1-\eta(x+z))z\,\pi(z)+(1-\eta(x))\sum_z\eta(x-z)z\,\pi(z)\Big]\\ &= \Big[-\eta(x)\sum_z\eta(x+z)z\,\pi(z)+(1-\eta(x))\sum_z\eta(x-z)z\,\pi(z)\Big]\\ &= \Big[\sum_z\eta(x-z)z\,\pi(z)-\eta(x)\sum_z(\eta(x+z)+\eta(x-z))z\,\pi(z)\Big]\\ &= \tau_x\mathbf{\Psi}_0\end{aligned}$$

with τ_x being the shift by x. The second sum is 0 in the symmetric case, and $\mathbf{\Psi}_0$ can then be written as a "gradient"

$$\mathbf{\Psi}_0=\sum_j\tau_{e_j}\xi_j-\xi_j$$

where τ_{e_j} are shifts in the coordinate directions. This allows us to do summation by parts and gain a factor of N^{-1}. When this is not the case, we have a "nongradient" model and the hydrodynamic limit can no longer be established by simple averaging.

Finally, when $m=\sum_z z\pi(z)\neq 0$, time has to be scaled by the same factor as space, namely by N. If we take $d=1$, we have

$$Nd\frac{1}{N}\sum_x J\Big(\frac{x}{N}\Big)\eta_t(x)\simeq\frac{1}{N}\sum_{x,z}J'\Big(\frac{x}{N}\Big)\eta_t(x)(1-\eta_t(x+z))z\pi(z).$$

The equation in the limit is hyperbolic and is the Burgers equation

$$\rho_t+m(\rho(1-\rho))_x=0.$$

This has been well studied. But, we will stay close to the symmetric case.

5.3. Symmetric Simple Exclusion

We will assume that we are dealing with a periodic lattice $\mathbf{Z}_N^d$ of N^d sites. We have a symmetric probability distribution $\pi(z)$ on $\mathbb{Z}^d$ that is symmetric and compactly supported. We will also assume that its support generates entire $\mathbb{Z}^d$. The covariance matrix C is defined by

$$\langle C\ell,\ell\rangle=\sum_z\langle z,\ell\rangle^2\pi(z)$$

and is positive definite. The simple exclusion process is defined by the generator

$$(\mathcal{L}f)(\eta)=\sum_{x,y}\pi(y-x)\eta(x)(1-\eta(y))[f(\eta^{x,y})-f(\eta)].$$

Here, the state space consists of Ω consisting of $\eta:\mathbf{Z}_N^d\to\{0,1\}$. The convention is that $\eta(x)=1$ if there is particle at x, and 0 otherwise. η represents the configuration of particles with at most one particle in any site. $\eta^{x,y}$ represents the state obtained by exchanging the situation in sites x and y:

$$\eta^{x,y}(z) = \begin{cases} \eta(y) & \text{if } z = x, \\ \eta(x) & \text{if } z = y, \\ \eta(z) & \text{otherwise.} \end{cases}$$

The initial configuration is a state $\{\eta(0, x)\}$, and we assume that for some density $\rho_0(u)$ on the torus $\mathbb{T}^d$, $0 \le \rho_0(u) \le 1$, we have

$$\lim_{N\to\infty} \frac{1}{N^d} \sum_{x \in \mathbf{Z}_\mathbf{N}^\mathbf{d}} J\Big(\frac{x}{N}\Big)\eta(x) \to \int_{\mathbb{T}^d} J(u)\rho_0(u)du$$

for every continuous function $J : \mathbb{T}^d \to \mathbb{R}$. $\rho_0(\cdot)$ should be thought of as the macroscopic density profile. We speed time up by a factor of N^2, and let P_N be the probability measure on the space $D[[0, T]; \Omega_N]$. We can map the configuration η to the measure $\lambda_N(du) = (1/N^d)\sum_x \delta_{x/N}$ on $\mathbb{T}^d$. This induces a measure Q_N on the space $D[[0, T]; \mathcal{M}(\mathbb{T}^d)]$. We will prove the following theorem, which is quite elementary.

THEOREM 5.1. *As $N \to \infty$, the distributions Q_N converge weakly to the degenerate distribution concentrated on the trajectory $\rho(t, u)$, that is, the unique solution $\rho(t, u)$ of the heat equation*

$$\frac{\partial \rho}{\partial t} = \frac{1}{2} \sum_{i,j=1}^{d} C_{i,j} D_i D_j \rho$$

with the initial condition $\rho(0, u) = \rho_0(u)$.

PROOF. Consider the function

$$F_J(\eta) = \frac{1}{N^d} \sum_{x \in \mathbf{Z}_\mathbf{N}^\mathbf{d}} J\Big(\frac{x}{N}\Big)\eta(x).$$

We can compute

$$\begin{aligned}
(N^2\mathcal{L}_N F_J)(\eta) &= N^{2-d} \sum_{x,y} \Big[J\Big(\frac{y}{N}\Big) - J\Big(\frac{x}{N}\Big)\Big]\eta(x)(1 - \eta(y))\pi(y - x) \\
&= \frac{N^{2-d}}{2} \sum_{x,y} [\eta(x)(1 - \eta(y)) - \eta(x)(1 - \eta(y))]\Big[J\Big(\frac{y}{N}\Big) - J\Big(\frac{x}{N}\Big)\Big]\pi(y - x) \\
&= \frac{N^{2-d}}{2} \sum_{x,y} [\eta(x) - \eta(y)]\Big[J\Big(\frac{y}{N}\Big) - J\Big(\frac{x}{N}\Big)\Big]\pi(y - x) \\
&= \frac{N^{2-d}}{2} \sum_{x,z} [\eta(x) - \eta(x + z)]\Big[J\Big(\frac{x + z}{N}\Big) - J\Big(\frac{x}{N}\Big)\Big]\pi(z) \\
&= \frac{N^{2-d}}{2} \sum_{x,z} \eta(x)\Big[J\Big(\frac{x + z}{N}\Big) - 2J\Big(\frac{x}{N}\Big) + J\Big(\frac{x - z}{N}\Big)\Big]\pi(z) \\
&\simeq \frac{1}{2N^d} \sum_{x} \eta(x) \sum_{i,j} C_{i,j}(D_i D_j J)\Big(\frac{x}{N}\Big) \\
&= \frac{1}{2} F_{J'}(\eta)
\end{aligned}$$

where

$$J'(u) = \Big(\sum_{i,j} C_{i,j} D_i D_j J\Big)(u).$$

We now establish two things: the sequence Q_N is compact as probability distributions on $D[[0,T];\Omega]$, and any limit Q is concentrated on the set of weak solutions of the heat equation with the correct initial condition. Since there is only one solution, we have weak convergence to the degenerate distribution concentrated at that solution, as claimed.

We will show compactness in $D[[0,T];\mathcal{M}(\mathbb{T}^d)]$. The topology on $\mathcal{M}(\mathbb{T}^d)$ is weak convergence. The space is compact. If we use a continuous test function J, the jump sizes are at most $\sup_{|x-y|\le C/N} |J(x)-J(y)|$, where C is the size of the support of $\pi(\cdot)$. It is therefore sufficient to get a uniform estimate on

$$\theta_{N,J}(\delta,\epsilon) = \sup_N P_N\Big[\sup_{0\le t\le\delta} |F_J(t)-F_J(0)| \ge \epsilon\Big]$$

and show that for any smooth J and $\epsilon > 0$,

$$\lim_{\delta\to 0}\ \sup_N \theta_{N,J}(\delta,\epsilon) \to 0.$$

From our computation of $(N^2\mathcal{L}_N F)(\eta)$, it follows that if J has two bounded derivatives, then

$$\sup_\eta \sup_N N^2|(\mathcal{L}_N F_J)(\eta)| \le C(J).$$

From the generator $N^2\mathcal{L}_N$, we conclude that

$$F_J(\eta(t)) - F_J(\eta(0)) - N^2\int_0^t (\mathcal{L}_N F_J)(\eta(s))ds = M_{J,N}(t)$$

is a martingale, $N^2|(\mathcal{L}_N F_J)(\eta)| \le C(J)$, and the quadratic variation of the martingale $M_{J,N}(t)$ is easily estimated. The jumps are of size $C(J)/N^{d+1}$. The total jump rate is N^{d+2}. The quadratic variation is then bounded by

$$C(J)N^{-2d-2}N^{d+2} = C(J)N^{-d}.$$

This is sufficient to prove the compactness of Q_N, and that any limit point Q will have the property that for any J,

$$\int_{\mathbb{T}^d} J(u)\lambda(t,u)du - \int_{\mathbb{T}^d} J(u)\rho_0(u)du - \frac{1}{2}\int_0^t\int_{\mathbb{T}^d}\sum_{i,j}(D_iD_jJ)(u)\lambda(s,u)du\,ds \equiv 0$$

for a.e. Q. □

REMARK 5.2. The symmetry of $\pi(\cdot)$ played a crucial part. $\eta(x)(1-\eta(y))-\eta(y)(1-\eta(x)) = \eta(x)-\eta(y)$ cancelled the nonlinearity. If we assumed only that $\sum_z z\pi(z) = 0$, we would not have arrived at the second difference, $N^2(\mathcal{L}_N F_j)(\eta)$ would be of size N, and the rapid fluctuation in the speeded up time scale would have to be used to get that term to be order 1.

5.4. Weak Asymmetry

Now that we have established the law of large numbers, if we want to investigate the large deviation properties, first we need to consider a class of perturbations of the symmetric jump rates $\pi(\cdot)$ by a small antisymmetric term. This "tilting" depends on the choice of a function $q(z)$ that is odd; i.e., $q(z) = -q(z)$ and $q(z) = 0$ unless $\pi(z) > 0$. Then, we perturb $\pi(z)$ by $\pi(z) + (1/N)q(z)$. That introduces a weak asymmetry. The choice of $q(\cdot) = q(s, u, \cdot)$ can depend smoothly on u and s. The generator is modified accordingly,

$$\begin{aligned}(N^2 \mathcal{L}_{N,s} F_J)(\eta) &= N^2 \sum_{x,y} \Big[J\Big(\frac{y}{N}\Big) - J\Big(\frac{x}{N}\Big)\Big] \eta(x)(1-\eta(y)) \Big[\pi(y-x) + \frac{1}{N} q\Big(s, \frac{x}{N}, y-x\Big)\Big] \\ &\simeq \frac{1}{2} \sum_x \eta(x) \sum_{i,j} C_{i,j} (D_i D_j J)\Big(\frac{x}{N}\Big) \\ &\quad + \Big\langle \sum_z z\, q\Big(s, \frac{x}{N}, z\Big) \eta(x)(1-\eta(x+z)), (\nabla J)\Big(\frac{x}{N}\Big)\Big\rangle.\end{aligned}$$

The first term can be replaced, as before, by

$$\frac{1}{2} \int_{\mathbb{T}^d} \sum_{i,j} (D_i D_j J)(u) \rho(s,u)\, du.$$

It is not clear what to do with the second term. In the limit, we would like this term to be

$$\int_{\mathbb{T}^d} \rho(s,u)(1-\rho(s,u)) \langle m(s,u), (\nabla J)(u)\rangle\, du$$

with

$$m(s,u) = \sum_z z q(s,u,z),$$

because we expect locally the statistics to reflect the Bernoulli distribution with the correct density.

THEOREM 5.3. *The sequence of measures $Q_{N,q}$ converges as $N \to \infty$ to the distribution concentrated on the single trajectory, that is, the unique weak solution of*

$$\frac{\partial \rho(t,u)}{\partial t} = \frac{1}{2} \sum_{i,j} C_{i,j} D_i D_j \rho(t,u) - \nabla \cdot (\rho(t,u)(1-\rho(t,u)) m(t,u)) \tag{5.8}$$

with $\rho(0,u) = \rho_0(u)$.

But this requires justification. The route is complicated. There are various approximations needed. There are several measures, P_N, Q_N, and the perturbed ones, $P_{N,q}$ and $Q_{N,q}$. Computations with P_N are easier because it is a reversible Markov process. Direct computations with $P_{N,q}$ are harder. Even while examining P_N, it is easier if we start with the uniform distribution of ρN^d particles among the N^d sites, which is a reversible invariant measure for the Markov process. One can then use the Feynman-Kac formula and some variational methods to obtain estimates. We can then transfer the estimates from P_N^{eq} to P_N and $P_{N,q}$ using an entropy inequality.

We will deal with averages of all kinds. Let us set up some notation. The configuration space consists of $\Omega_N = \{\eta : \mathbf{Z}_N^d \to \{0,1\}\}$ or $\Omega = \{\eta : \mathbb{Z}^d \to \{0,1\}\}$. η_s is the configuration at time s and $\eta_s(x)$ is 1 if there is a particle at site x at time s and 0 otherwise. If $\eta \in \Omega$ or Ω_N, we denote by $\tau_x \eta$ the translate defined by $(\tau_x \eta)(y) = \eta(x+y)$, and if

$f = f(\eta)$ is a local function on the configuration space, we denote by $\tau_x f$ the function $(\tau_x f)(\eta) = f(\tau_x \eta)$. Their averages will be denoted by

$$\bar{f}_{\ell,x} = \bar{f}_{\ell,x}(\eta) = \frac{1}{(2\ell+1)^d} \sum_{y:|y-x|\le\ell} f(\tau_y \eta).$$

A special case is when $f(\eta) = \eta(0)$, in which case we denote

$$\bar{\eta}_{\ell,x} = \frac{1}{(2\ell+1)^d} \sum_{y:|y-x|\le\ell} \eta(y).$$

$\eta(x)$ can be $\eta_s(x)$ representing the configuration at some time s, and in that case we will denote the averages by $\bar{f}_{s,\ell,x}$ or $\bar{\eta}_{s,\ell,x}$. The object we need to estimate is

$$\int_0^T e_N(\epsilon, s)ds$$

where

$$e_N(\epsilon, s) = E^{P_{N,q}}\left[\left|\frac{1}{N^d}\sum J\left(\frac{x}{N}\right)\left[f_x(\eta_s) - \hat{f}(\bar{\eta}_{N\epsilon,s,x})\right]\right|\right]$$

and

$$\hat{f}(\rho) = E^{\mu_\rho}[f(\eta)],$$

the expectation of the local function f with respect to Bernoulli with density ρ.

Let $\mu_N(s, d\eta)$ be the marginal distribution on Ω_N of the configuration η_s at time s, and $\bar{\mu}_N$ its space and time average. More precisely,

$$\int f(\eta) d\bar{\mu}_N(\eta) = \frac{1}{T}\frac{1}{N^d}\int_0^T \sum_x E^{P_N}[f(\tau_x \eta_s)]ds.$$

We need the following theorem to prove Theorem 5.3.

THEOREM 5.4.

$$\lim_{\epsilon\to 0}\limsup_{N\to\infty}\int_0^T e_N(\epsilon, s)ds = 0.$$

This is proved in two steps:

LEMMA 5.5.

$$\lim_{k\to\infty}\limsup_{N\to\infty} E^{\bar{\mu}_N}\left[\left|\bar{f}_k(\eta) - \hat{f}(\bar{\eta}_k)\right|\right] = 0.$$

LEMMA 5.6.

$$\limsup_{\substack{\epsilon\to 0,\\ k\to\infty}}\limsup_{N\to\infty} E^{\bar{\mu}_N}\left[|\bar{\eta}_{N\epsilon} - \bar{\eta}_k|\right] = 0.$$

We will prove Theorem 5.4, assuming Lemma 5.5 and Lemma 5.6. The proof proceeds along the following lines:

- We can always replace, for fixed k and large N, the sum

  $$\frac{1}{N^d}\sum_x J\left(\frac{x}{N}\right) f(\tau_x \eta)$$

  by

  $$\frac{1}{N^d}\sum_x f(\tau_x \eta)\frac{1}{(2k+1)^d}\sum_{y:|y-x|\le k} J\left(\frac{y}{N}\right)$$

or

$$\frac{1}{N^d}\sum_x J\Big(\frac{x}{N}\Big)\bar{f}_{k,x}.$$

So long as $k \ll N$, this is not a problem.

- Lemma 5.5 allows us to replace
$$\int_0^t \frac{1}{N^d}\sum_x J\Big(\frac{x}{N}\Big)\bar{f}_{k,x}(\eta_s)ds$$
by
$$\int_0^t \frac{1}{N^d}\sum_x J\Big(\frac{x}{N}\Big)\hat{f}(\bar{\eta}_{s,k,x})ds$$
with a fixed, large k.
- Lemma 5.6 allows us to replace $\bar{\eta}_{k,x}$ with large k by $\bar{\eta}_{N\epsilon,x}$ with a small ϵ.
- Now we have the expression
$$\int_0^t \frac{1}{N^d}\sum_x J\Big(\frac{x}{N}\Big)\hat{f}(\bar{\eta}_{s,N\epsilon,x})ds.$$
- This term is continuous in the weak topology and can be replaced, as $N \to \infty$, by
$$\int_0^t\int_{\mathbb{T}^d} J(u)\hat{f}(\bar{\rho}_\epsilon(s,u))ds\,du$$
where
$$\bar{\rho}_\epsilon(s,u) = (2\epsilon)^{-d}\int_{B(u,\epsilon)} \rho(s,v)dv.$$
- Finally, we let $\epsilon \to 0$ to arrive at
$$\int_0^t\int_{\mathbb{T}^d} J(u)\hat{f}(\rho(s,u))ds\,du.$$

We now concentrate on the proof of the two lemmas. They depend on the following observations:

- The function $f(a,b) = (\sqrt{a}-\sqrt{b})^2$ is a convex function of $(a,b) \in \mathbb{R}^2_+$. It is checked by computing the Hessian
$$\frac{1}{2}\begin{pmatrix} \frac{b}{a^{3/2}} & -\frac{1}{\sqrt{a}\sqrt{b}} \\ -\frac{1}{\sqrt{a}\sqrt{b}} & \frac{a}{b^{3/2}} \end{pmatrix}$$
of f, which is seen to be positive semidefinite.
- In general, if L is the generator of a Markov process on a finite state space $\mathcal{X}$ given by
$$(Lf)(x) = \sum_y [f(y) - f(x)]a(x,y),$$
and μ is a reversible invariant measure for L, the detailed balance condition
$$\mu(x)a(x,y) = a(y,x)\mu(y)$$

is satisfied. Then, the Dirichlet form is given by

$$\begin{aligned}
\mathcal{D}(f) =& \frac{1}{2}\sum_{x,y}|f(y)-f(x)|^2 a(x,y)\mu(x) \\
=& \frac{1}{2}\sum_{x,y}[f(x)^2+f(y)^2-2f(x)f(y)]a(x,y)\mu(x) \\
=& -\frac{1}{2}\Big[\sum_x f(x)\mu(x)\sum_y a(x,y)[f(y)-f(x)] \\
&\qquad + \sum_y f(y)\mu(y)\sum_x a(y,x)[f(x)-f(y)]\Big] \\
=& -\Big[\sum_x f(x)\mu(x)\sum_y a(x,y)[f(y)-f(x)]\Big] \\
=& -\langle Lf, f\rangle_\mu.
\end{aligned}$$

- If $\mu_N = \{\mu_N(\eta)\}$ is a probability distribution on Ω_N, its Dirichlet form defined by

$$\mathcal{D}(\mu_N) = \frac{1}{2}\sum_{x,y\in\mathbf{Z}_N^d}\left(\sqrt{\mu_N(\eta^{x,y})}-\sqrt{\mu_N(\eta)}\right)^2\pi(y-x)\eta(x)(1-\eta(y))$$

  is equivalent to

$$\mathcal{D}(\mu_N) = \frac{1}{4}\sum_{\substack{x,y\in\mathbf{Z}_\mathbf{N}^\mathbf{d},\\ |x-y|=1}}\left(\sqrt{\mu_N(\eta^{x,y})}-\sqrt{\mu_N(\eta)}\right)^2,$$

  if $\pi(\cdot)$ is irreducible. An estimate of the form $\mathcal{D}(\mu_N) \le CN^{d-2}$ has many consequences.
- If $\mu_N(s)$ are such that $\int_0^T D(\mu_N(s))ds \le CN^{d-2}$, then the average $\bar{\mu}_N$ over space and time satisfies

$$\mathcal{D}(\bar{\mu}_N) \le \frac{C}{T}N^{d-2}.$$

  Since there is spacial homogeneity, the quantity

$$\mathcal{D}(x) = \sum_{y:|y-x|=1}\left(\sqrt{\bar{\mu}_N(\eta^{x,y})}-\sqrt{\bar{\mu}_N(\eta)}\right)^2$$

  is independent of x and satisfies

$$\mathcal{D}(x) \le \frac{C}{T}N^{-2}.$$

- If we restrict the distribution to a block $\mathbf{B}_k$ of size k, the marginal distribution has the Dirichlet form from bonds internal to $\mathbf{B}_k$, that is, at most $(2k)^d\frac{C}{T}N^{-2}$. As $N \to \infty$, the Dirichlet form goes to 0. Any limiting distribution is then permutation invariant. So, it is uniform over all possible choices of locations for the given number of particles in the block. This shows that any limit as $k \to \infty$ of any limit as $N \to \infty$ is Bernoulli, provided the density is specified. This is precisely Lemma 5.5.

- We want to estimate $(\sqrt{\mu_N(\eta^{x,y})} - \sqrt{\mu_N(\eta)})^2$ when $|x-y| = \ell$ instead of $|x-y| = 1$. Any interchange of x, y at a distance ℓ can be achieved by 2ℓ successive interchanges of nearest neighbors, and the simple inequality

$$(a_1 + \cdots + a_\ell)^2 \le \ell \sum_{j=1}^{\ell} a_j^2$$

allows us to estimate

$$\big(\sqrt{\mu_N(\eta^{x,x+z})} - \sqrt{\mu_N(\eta)}\big)^2 \le |z|^2 \frac{C}{T} N^{-2}.$$

In particular, if $|z| \le N\epsilon$, then

$$\big(\sqrt{\mu_N(\eta^{x,x+z})} - \sqrt{\mu_N(\eta)}\big)^2 \le |z|^2 \frac{C}{T} N^{-2} \le \epsilon^2 \frac{C}{T}.$$

- In the limit, we have two blocks of size q with a single bond between them connecting two sites, one from each block. The Dirichlet form of any limiting distribution is 0. The Dirichlet form has three terms, one for each block and one connecting the two. It is easy to check that the invariant distribution in each block, as before, is the uniform distribution, and that the two blocks have a bond connecting them makes it uniform over both blocks.
- This shows that if two microscopically large blocks are macroscopically close, then their empirical densities are close, with probability nearly 1. This implies Lemma 5.6.

All we need to do now is control the Dirichlet form. We defined the Dirichlet form as

$$\sum_{x,y \in \mathbf{Z}_N^d} \pi(y-x)\big(\sqrt{\mu_N(\eta^{x,y})} - \sqrt{\mu_N(\eta)}\big)^2.$$

If the group generated by the set $\{z : \pi(z) > 0\}$ is all of $\mathbb{Z}^d$, then the two Dirichlet forms are comparable. Our generator is of the form

$$\mathcal{L}_{N,q} = \mathcal{L}_N + \frac{1}{N}\mathcal{A}_{N,q},$$

the sum of the symmetric part and the weak perturbation. The entropy is defined by

$$\mathcal{H}(\mu) = \sum_\eta \mu(\eta) \log \frac{\mu(\eta)}{\alpha(\eta)}$$

where $\alpha(\eta) = c = \binom{N^d}{k_N - 1}$ is the uniform distribution of $k_N = \rho N^d$ particles on $\mathbb{Z}_N^d$. It is invariant for the evolution under $\mathcal{L}_N$.

Let us note that μ_N^t evolves according to the forward equation. Equivalently,

$$\frac{d}{dt}\sum_\eta F(\eta) d\mu_N^t = \sum_\eta \mathcal{L}_{N,q} F d\mu_N^t.$$

Let us compute the rate of change of the entropy in the perturbed system with weak asymmetry. In the unperturbed case, the relative entropy with respect to equilibrium, i.e., uniform distribution, decreases monotonically, and the rate of decrease dominates the

Dirichlet form. We will show that weak asymmetry does not change the situation considerably:

$$\begin{aligned}
\frac{\partial \mathcal{H}(\mu_N^t)}{\partial t} &= \frac{d}{dt}\sum_\eta \log\frac{\mu_N^t(\eta)}{c}\mu_N^t(\eta)\\
&= N^2\sum_\eta\big(\mathcal{L}_{N,q}\log\mu_N^t(\eta)\big)\mu_N^t(\eta) + (\log c)\frac{d}{dt}\sum_\eta\mu_N^t(\eta)\\
&= N^2\sum_\eta\big(\mathcal{L}_{N}\log\mu_N^t(\eta)\big)\mu_N^t(\eta) + N\sum_\eta\big(\mathcal{A}_{N,q}\log\mu_N^t(\eta)\big)\mu_N^t(\eta)\\
&= (N^2-CN)\sum_\eta\Big[\sum_{x,y}\pi(y-x)\eta(x)(1-\eta(y))\log\frac{\mu_N^t(\eta^{x,y})}{\mu_N^t(\eta)}\Big]\mu_N^t(\eta)\\
&\quad + N\sum_\eta\Big[\sum_{x,y}\Big[C\pi(y-x)+q\Big(t,\frac{x}{N},y-x\Big)\Big]\eta(x)(1-\eta(y))\log\frac{\mu_N^t(\eta^{x,y})}{\mu_N^t(\eta)}\Big]\mu_N^t(\eta)\\
&= -\frac{N^2-CN}{2}\sum_{x,y}\pi(y-x)\sum_\eta\Big[\log\frac{\mu_N^t(\eta^{x,y})}{\mu_N^t(\eta)}\Big]\big[\mu_N^t(\eta^{x,y})-\mu_N^t(\eta)\big]\\
&\quad + N\sum_{x,y}\Big[C\pi(y-x)+q\Big(t,\frac{x}{N},y-x\Big)\Big]\sum_\eta\eta(x)(1-\eta(y))\Big[\log\frac{\mu_N^t(\eta^{x,y})}{\mu_N^t(\eta)}\Big]\mu_N^t(\eta).
\end{aligned}$$

We can change $\eta\to\eta^{x,y}$ in the summation. It just maps Ω_N onto itself. The first term is odd when we interchange $x\leftrightarrow y$ and the sum can be rewritten as

$$\begin{aligned}
&-\frac{N^2-CN}{2}\sum_{x,y}\pi(y-x)\sum_\eta\big[\psi\big(\mu_N^t(\eta^{x,y}),\mu_N^t(\eta)\big)\big]\\
&+N\sum_{x,y}\eta(x)(1-\eta(y))\Big[C\pi(y-x)+q\Big(t,\frac{x}{N},y-x\Big)\Big]\sum_\eta\big[\phi\big(\mu_N^t(\eta^{x,y}),\mu_N^t(\eta)\big)\big]
\end{aligned}$$

where $\psi(a,b)=[\log a-\log b][a-b]$ and $\phi(a,b)=a\log\frac{b}{a}$. It is easy to check that $\psi(a,b)\geq(\sqrt{a}-\sqrt{b})^2$, and for $b>a$, $\phi(a,b)\leq 2(\sqrt{b}-\sqrt{a})\sqrt{a}$ or $\phi(a,b)\leq|\sqrt{b}-\sqrt{a}|\sqrt{a}$. Since $q(t,u,z)+C\pi(z)\geq 0$ for our choice of C, it follows that the second term is dominated by

$$\begin{aligned}
&4CN\sum_{x,y,\eta}\pi(y-x)\Big|\sqrt{\mu_N^t(\eta^{x,y})}-\sqrt{\mu_N^t(\eta)}\Big|\sqrt{\mu_N^t(\eta)}\leq\\
&\qquad\qquad 4CN\Big[\sum_{x,y}\pi(y-x)\sum_\eta\Big(\sqrt{\mu_N^t(\eta^{x,y})}-\sqrt{\mu_N^t(\eta)}\Big)^2\Big]^{\frac12}.
\end{aligned}$$

This provides an estimate

$$\frac{\partial\mathcal{H}(\mu_N^t)}{\partial t}\leq-\frac{N^2-CN}{2}\mathcal{D}(\mu_N^t)+4CN\sqrt{\mathcal{D}(\mu_N^t)}.$$

Integrating t from 0 to T, we obtain for large N with $I_N(T)=\int_0^T\mathcal{D}(\mu_N^t)dt$

$$-CN^d\leq\mathcal{H}(\mu_N^T)-\mathcal{H}(\mu_n^0)\leq-\frac{N^2}{3}I_N(T)+4C(T)N\sqrt{I_N(T)}.$$

We have used the bound on the initial entropy $\mathcal{H}(\mu_N) \le CN^d$ coupled with the obvious lower bound $\mathcal{H}(\mu_N^T) \ge 0$. It follows now that

$$I_N(t) \le C(T)N^{d-2}.$$

This completes the proof of Theorem 5.4 and therefore Theorem 5.3. But we need to prove uniqueness.

Since $q(s,x,z)$ is a nice function, $m(s,u)$ will be bounded. The difference r between two solutions ρ_1 and ρ_2 satisfies, in the weak sense,

$$r_t = \frac{1}{2}\sum_{i,j} C_{i,j} D_i D_j\, r - \nabla\cdot[\rho_1(1-\rho_1)-\rho_2(1-\rho_2)]m = \frac{1}{2}\Delta_C\, r - \nabla\cdot r(1-\rho_1-\rho_2)m.$$

The key to establishing uniqueness is the estimate

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{T}^d} |r|^2\,du = 2\int_{\mathbb{T}^d} r r_t\,du &= \int [r\Delta_C r - \nabla\cdot rH]du\\ &= -\int_{\mathbb{T}^d}\langle\nabla r, C\nabla r\rangle du + \int_{\mathbb{T}^d} r\,H\cdot\nabla r\,du\\ &\le -c\int_{\mathbb{T}^d}|\nabla r|^2\,du + \int_{\mathbb{T}^d} r\,H\cdot\nabla r\,du\\ &\le C(H)\int_{\mathbb{T}^d}|r|^2\,du\end{aligned}$$

where $H = 2(1-\rho_1-\rho_2)m$ and $C(H) = c^{-1}\sup_{s,u}|m(s,u)| \le c^{-1}\sup_{s,u}\sum_z |z||q(z,s,u)|$. In the nonlinear case, we need some a priori regularity estimates. The following lemma will suffice.

LEMMA 5.7. *Let $|m(t,u)| \le M$ and $\rho(t,u)$ be a weak solution of*

$$\rho_t = \frac{1}{2}\Delta_C\,\rho - \nabla^{\cdot}\rho(1-\rho)m(t,u).$$

Then there are smooth solutions ρ^ϵ of

$$\rho^\epsilon_t = \frac{1}{2}\Delta_C\,\rho^\epsilon - \nabla^{\cdot}\rho^\epsilon(1-\rho^\epsilon)m_\epsilon(t,u)$$

with $|m_\epsilon(t,u)| \le M$, and there is a uniform bound on

$$\int_0^T\int_{\mathbb{T}^d}\frac{|\nabla\rho^\epsilon|^2}{\rho^\epsilon(1-\rho^\epsilon)}\,du\,dt.$$

In particular, $\int_0^T\int_{\mathbb{T}^d}\frac{|\nabla\rho|^2}{\rho(1-\rho)}\,du\,dt < \infty$.

PROOF. We take ρ_ϵ to be the convolution of ρ with an approximate identity ϕ_ϵ,

$$\rho^\epsilon_t = \frac{1}{2}\Delta_C\,\rho^\epsilon - \nabla[(m\rho(1-\rho)) * \phi_n].$$

The entropy $h(\rho) = \int_{\mathbb{T}^d}[\rho\log\rho + (1-\rho)\log(1-\rho)]du \le 0$ satisfies

$$h(\rho_n(T,\cdot)) - h(\rho_n(0,\cdot)) \ge h(\rho_n(T,\cdot)) \ge -c.$$

We have, therefore, an upper bound on

$$\begin{aligned}
-\int_0^T \frac{dh(\rho^\epsilon(t))}{dt}\,dt &= -\int_0^T \int_{\mathbb{T}^d} \log\left(\frac{\rho^\epsilon}{1-\rho\epsilon}\right)\rho_t^\epsilon\, dt\, du \\
&= -\int_0^T \int_{\mathbb{T}^d} \log\left(\frac{\rho^\epsilon}{1-\rho^\epsilon}\right)\left[\frac{1}{2}\Delta_C \rho^\epsilon - \nabla\cdot[(m\rho(1-\rho))*\phi_\epsilon]\right] dt\, du \\
&= \frac{1}{2}\int_0^T \int_{\mathbb{T}^d} \frac{\langle \nabla\rho^\epsilon, C\nabla\rho^\epsilon\rangle}{\rho^\epsilon(1-\rho^\epsilon)}\,dt\,du - \int_0^T \int_{\mathbb{T}^d} \frac{\langle \nabla\rho^\epsilon, (m\rho(1-\rho))*\phi_\epsilon\rangle}{\rho^\epsilon(1-\rho^\epsilon)}\,dt\,du.
\end{aligned}$$

It is enough to get a control on

$$\frac{\langle (m\rho(1-\rho))*\phi_\epsilon, C^{-1}(m\rho(1-\rho))*\phi_\epsilon\rangle}{\rho^\epsilon(1-\rho^\epsilon)}.$$

By the concavity of the function $\rho(1-\rho)$, we have

$$\rho^\epsilon(1-\rho^\epsilon) \geq (\rho(1-\rho))^\epsilon.$$

It is now easy to estimate

$$\frac{\langle (m\rho(1-\rho))*\phi_\epsilon, C^{-1}(m\rho(1-\rho))*\phi_\epsilon\rangle}{(\rho(1-\rho))^\epsilon}$$

in terms of a bound on m. $\square$

5.5. Large Deviations

Theorem 5.3 actually provides a large deviation lower bound. If on some state space $\mathcal{X}$, P_1 and P_2 are two jump Markov processes in the interval $[0,T]$, with jump rates $\lambda_1(t,x,y)$ and $\lambda_2(t,x,y)$, and $\lambda_2(t,x,y) \leq C\lambda_1(t,x,y)$, then $P_2 \ll P_1$ and the relative entropy is given by

$$\mathcal{H}(P_2;P_1) = E_2^P\left[\int_0^T \left[\lambda_2(t,x(s),y)\log\left(\frac{\lambda_2(t,x(s),y)}{\lambda_1(t,x(s),y)}\right) - \lambda_2(t,x(s),y) + \lambda_1(t,x(s),y)ds\right]\right].$$

In our context, P_2 is the process with generator $\mathcal{L}_{N,q}$, while P_1 has generator $\mathcal{L}_N$. The relative entropy can be explicitly calculated.

$$\begin{aligned}
[\pi + N^{-1}q]\log\left(\frac{\pi+N^{-1}q}{\pi}\right) - N^{-1}q &= (\pi+N^{-1}q)\log(1+N^{-1}\pi^{-1}q) - N^{-1}q \\
&\simeq \frac{1}{2}N^{-2}\pi^{-1}q^2.
\end{aligned}$$

We have speeded up the time scale by a factor of N^2. So, the relative entropy is given by

$$\mathcal{H}(N,q) = \frac{1}{2}E^{P_{N,q}}\left[\int_0^T \sum_{x,y} \frac{q^2\left(s,\frac{x}{N},y-x\right)}{\pi(y-x)}\eta_s(x)(1-\eta_s(y))ds\right].$$

We can see that $N^{-d}\mathcal{H}(N,q)$ has a limit as $N\to\infty$ and is given by

$$\frac{1}{2}\int_0^T \int_{\mathbb{T}^d} \sum_z \frac{q^2(s,u,z)}{\pi(z)}\rho(s,u)(1-\rho(s,u))ds\,du$$

where $\rho(s,u)$ is the solution of (5.3). If we want to make a perturbation that produces a profile $\rho(s,u)$ in the limit, we need to find m that satisfies (5.3) for a given ρ. Since we want the best possible lower bound, and the lower bound is provided by entropy of the tilt required to produce the desired limit, we need to minimize the entropy term $\sum_z q^2(z)/\pi(z)$ over q such that $\sum_z zq(z) = m$. This can be easily carried out and yields

$$q(z) = \pi(z) \sum_j c_j z_j$$

with

$$\sum_z \sum_j \pi(z) c_j z_i z_j = m$$

or

$$c = C^{-1} m$$

where C is the covariance matrix of π. Finally,

$$\sum_z \frac{q(z)^2}{\pi(z)} = \sum_z \pi(z) \Big[\sum c_j z_j\Big]^2 = \langle c, C\, c\rangle = \langle m, C^{-1} m\rangle,$$

and the large deviation lower bound is given by

$$J(m) = \frac{1}{2} \int_0^T \int_{\mathbb{T}^d} \langle m(s,u), C^{-1} m(s,u)\rangle \rho(s,u)(1-\rho(s,u)) ds\, du.$$

Finally, we need to minimize the above over all m that lead to the given ρ, i.e., m such that

$$\rho_t - \frac{1}{2} \sum_{i,j=1}^{d} C_{i,j} D_i D_j \rho = \nabla \cdot m(t,u)\rho(t,u)(1-\rho(t,u)). \tag{5.9}$$

If we denote the set of solutions m by $M(\rho(\cdot))$, then the large deviation lower bound is given by

$$I(\rho(\cdot)) = \inf_{m(\cdot)\in M(\rho(\cdot))} J(m(\cdot)).$$

One can also rewrite this variational problem in terms of the dual problem

$$\begin{aligned} I(\rho(\cdot)) = \sup_{\psi(t,u)} \Bigg[\int \psi(t,u) \Bigg[\rho_t - \frac{1}{2} \sum_{i,j=1}^{d} C_{i,j} D_i D_j \rho \Bigg] dt\, du \\ - \frac{1}{2} \int \langle \nabla\psi, C \nabla\psi\rangle \rho(t,u)(1-\rho(t,u) dt\, du \Bigg], \end{aligned}$$

which is suitable for proving the large deviation upper bound.

The upper bound depends on two steps: a local estimate and an exponential tightness estimate. Since the topology is weak convergence as measures on $D[[0,T], \mathcal{M}(\mathbb{T}^d)]$, to prove exponential tightness, it is enough to prove for any $\epsilon > 0$ an estimate of the form

$$\limsup_{\delta\to 0} \limsup_{N\to\infty} \sup_{\eta} \frac{1}{N^d} \log P_\eta \Bigg[\sup_{0\le s\le \delta} \frac{1}{N^d} \sum_x J\Big(\frac{x}{N}\Big)[\eta_s(x) - \eta_0(x)] \ge \epsilon \Bigg] = -\infty.$$

The proof uses the exponential martingales $M_J(t)$ (relative to P_N) given by

$$\begin{aligned} M_J(t) = \exp\Bigg[\sum_x J\Big(\frac{x}{N}\Big)[\eta_t(x) - \eta_0(x)] \\ - \int_0^t N^2 \sum_{x,y} \pi(y-x)\eta_s(x)(1-\eta_s(y)) \Big[e^{[J(\frac{y}{N}) - J(\frac{x}{N})]} - 1\Big] ds \Bigg]. \end{aligned}$$

For smooth J, if we use the symmetry of π, the integrand in the second term in the exponent can be estimated uniformly on η_s by CN^d, where $C = C(J)$ may depend on J. Therefore,

$$\exp\Big[\sum_x J\Big(\frac{x}{N}\Big)[\eta_s(x) - \eta_0(x)] - C(J)tN^d\Big]$$

is a nonnegative supermartingale, and it follows that for any constant $k > 0$, replacing J by kJ,

$$\frac{1}{N^d}\log P_\eta\Big[\sup_{0\le s\le\delta}\sum_x J\Big(\frac{x}{N}\Big)[\eta_s(x) - \eta_0(x)] \ge N^d\epsilon\Big] \le -k\epsilon + C(kJ)\delta.$$

We can let $N \to \infty$, $\delta \to 0$, to get

$$\limsup_{\delta\to 0}\limsup_{N\to\infty}\sup_\eta \frac{1}{N^d}\log P_\eta\Big[\sup_{0\le s\le\delta}\frac{1}{N^d}\sum_x J\Big(\frac{x}{N}\Big)[\eta_s(x) - \eta_0(x)] \ge \epsilon\Big] = -k\epsilon.$$

Since $k > 0$ is arbitrary, we can let $k \to \infty$.

To prove the upper bound with the rate function $I(\rho(\cdot))$, we observe that we have martingales $M_J(t)$ with time-dependent J of the form

$$\begin{aligned} M_J(t) = \exp\Big[\Big[\sum_x J\Big(t, \frac{x}{N}\Big)\eta_t(x) - \sum_x J\Big(0, \frac{x}{N}\Big)\eta_0(x)\Big] - \int_0^t J_s\Big(s, \frac{x}{N}\Big)\eta_s(x)ds \\ - \int_0^t N^2\sum_{x,y}\pi(y-x)\eta_s(x)(1-\eta_s(y))\big[e^{[J(s,\frac{y}{N})-J(s,\frac{x}{N})]} - 1\big]ds\Big] \end{aligned}$$

with

$$E^{P_\eta}[M_J(T)] = 1,$$

and look carefully at

$$\frac{1}{N^d}\int_0^T N^2\sum_{x,y}\pi(y-x)\big[e^{[J(s,\frac{y}{N})-J(s,\frac{x}{N})]} - 1\big]\eta_s(x)(1-\eta_s(y))ds,$$

which can be uniformly approximated by

$$\begin{aligned} &\frac{1}{N^d}\int_0^T N^2\sum_{x,y}\pi(y-x)\Big[\Big[J\Big(s, \frac{y}{N}\Big) - J\Big(s, \frac{x}{N}\Big)\Big] \\ &\qquad + \frac{1}{2}\Big[J\Big(s, \frac{y}{N}\Big) - J\Big(s, \frac{x}{N}\Big)\Big]^2\Big]\eta_s(x)(1-\eta_s(y))ds \\ &= \frac{1}{N^d}\int_0^T N^2\sum_{x,y}\pi(y-x)\Big[J\Big(s, \frac{y}{N}\Big) - J\Big(s, \frac{x}{N}\Big)\Big]\eta_s(x)(1-\eta_s(y))ds \\ &\quad + \frac{1}{2}\frac{1}{N^d}\int_0^T N^2\Big[J\Big(s, \frac{y}{N}\Big) - J\Big(s, \frac{x}{N}\Big)\Big]^2\Big]\eta_s(x)(1-\eta_s(y))ds \\ &= \frac{1}{N^d}\int_0^T N^2\sum_{x,y}\pi(y-x)\Big[J\Big(s, \frac{y}{N}\Big) - J\Big(s, \frac{x}{N}\Big)\Big]\eta_s(x) \\ &\quad + \frac{1}{2}\frac{1}{N^d}\int_0^T N^2\Big[J\Big(s, \frac{y}{N}\Big) - J\Big(s, \frac{x}{N}\Big)\Big]^2\Big]\eta_s(x)(1-\eta_s(y))ds \end{aligned}$$

$$\simeq \frac{1}{N^d}\int_0^T \frac{1}{2}\sum_x \sum_{i,j=1}^d C_{i,j}\Big(\frac{\partial^2}{\partial u_i \partial u_j}J\Big)\Big(s,\frac{x}{N}\Big)\eta_s(x)$$
$$+\frac{1}{2}\frac{1}{N^d}\int_0^T \sum_x \Big\langle(\nabla J)\Big(s,\frac{x}{N}\Big), C\,(\nabla J)\Big(s,\frac{x}{N}\Big)\Big\rangle \eta_s(x)(1-\eta_s(y))ds.$$

If $(1/N^d)\sum \delta_{\frac{x}{N}}\eta_s(x)$ is in a neighborhood of $\rho(\,\cdot\,,\,\cdot\,)$ on $D[[0,T];\mathcal{M}(\mathbb{T}^d)]$, the first term is close to

$$\frac{1}{2}\int_0^T \sum_{i,j=1}^d C_{i,j}\bigg(\frac{\partial^2}{\partial u_i \partial u_j}J\bigg)(s,u)\rho(s,u)ds\,du,$$

and we would like to be able to replace the second term by

$$\frac{1}{2}\int_0^T \int_{\mathbb{T}^d} \langle(\nabla J)(t,u), C\,(\nabla J)(s,u)\rangle\rho(s,u)(1-\rho(s,u))ds\,du.$$

To do that, in this context, we need stronger versions of Lemma 5.5 and Lemma 5.6.

LEMMA 5.8. *For any $\epsilon > 0$,*

$$\lim_{k\to\infty}\limsup_{N\to\infty}\log P_N\bigg[\bigg[\int_0^T \sum_x \frac{1}{N^d}|\bar{f}_{k,s}(\eta_s)-\hat{f}(\bar{\eta}_{k,x,s})|ds\bigg] > \epsilon\bigg] = -\infty.$$

The counterpart is the following:

LEMMA 5.9. *For any $\theta > 0$,*

$$\limsup_{\substack{\epsilon\to 0,\\ k\to\infty}}\limsup_{N\to\infty}\frac{1}{N^d}\log P_N\bigg[\bigg[\int_0^T \frac{1}{N^d}\sum_x |\bar{\eta}_{N\epsilon,x,s}-\bar{\eta}_{k,x,s}|ds\bigg] > \theta\bigg] = -\infty.$$

Actually, in the way we proved our Lemma 5.5 and Lemma 5.6, the stronger version is already there. To obtain superexponential estimates on

$$P_N\bigg[\int_0^T V_{N,\epsilon}(\eta_s)ds \geq \theta\bigg],$$

it is enough if we show that for any $\lambda > 0$,

$$\limsup_{\epsilon\to 0}\limsup_{N\to\infty}\frac{1}{N^d}\log E^{P_N}\bigg[\exp\bigg[\lambda N^d\int_0^T V_{N,\epsilon}(\eta_s)ds\bigg]\bigg] = 0.$$

Since P_N is a reversible stationary Markov process by the Feynman-Kac formula, one can establish the inequality

$$\log E^{P_N}\bigg[\exp\bigg[\lambda N^d\int_0^T V_{N,\epsilon}(\eta_s)ds\bigg]\bigg] \leq \sup_{\substack{f\geq 0,\\ \int f\,d\mu=1}}\bigg[\lambda\int V_{N,\epsilon}(\eta)f(\eta)d\mu(\eta) - N^2\mathcal{D}\big(\sqrt{f}\big)\bigg]. \tag{5.10}$$

We showed that, if $\mathcal{D}(\sqrt{f}) \leq CN^{-2}$, then we could control the errors, and for the two functions of interest in Lemma 5.5 and Lemma 5.6,

$$\limsup_{\epsilon\to 0}\limsup_{N\to\infty}\sup_{\substack{f\geq 0,\int f\,d\mu=1,\\ \mathcal{D}(\sqrt{f})\leq cN^{-2}}}\int V_{N,\epsilon}(\eta)f(\eta)d\mu(\eta) = 0.$$

Since $V_{N,\epsilon}$ is uniformly bounded, we can restrict f in the variational formula (5.10) to those that satisfy $\mathcal{D}(\sqrt{f}) \le cN^{-2}$ for some c.

We have essentially established the large deviation property for the distribution $Q_{N,\eta}$ of the random measure-valued function

$$\frac{1}{N^d}\sum_x \delta_{\frac{x}{N}}\eta_s(x)$$

starting from a nonrandom initial configuration η, under the assumption that

$$\frac{1}{N^d}\sum_x \delta_{\frac{x}{N}} \to \rho_0(u)du.$$

Earlier, we proved convergence in probability to $\rho(t,u)du$; that is, the solution of the heat equation

$$\frac{\partial\rho}{\partial t} = \frac{1}{2}\Delta_C\rho$$

with $\rho(0,u) = \rho_0(u)$. We have, through martingales, controlled the exponential moments of functionals that approximate with a multiple of N^d:

$$\begin{aligned}\Psi_J(\rho(\cdot,\cdot)) = &\int_{\mathbb{T}^d} J(T,u)\rho(T,u)du - \int_{\mathbb{T}^d} J(0,u)\rho(0,u)du\\ &- \int_0^T\int_{\mathbb{T}^d} J_s(s,u)\rho(s,u)ds - \frac{1}{2}\int_0^T\int_{\mathbb{T}^d}(\Delta_C J)(s,u)\rho(s,u)ds\,du\\ &- \frac{1}{2}\int_0^T\int_{\mathbb{T}^d}\langle(\nabla J)(s,u), C(\nabla J)(s,u)\rangle\rho(s,u)(1-\rho(s,u))ds\,du.\end{aligned}$$

We then have a large deviation principle with rate function

$$I(\rho(\cdot)) = \sup_J \Psi(\rho(\cdot,\cdot)).$$

Integrating by parts, we get

$$\begin{aligned}I(\rho(\cdot)) = \sup_J\Bigg[&\int_0^T\int_{\mathbb{T}^d} J(s,u)\Big[\rho_s(s,u)ds - \frac{1}{2}\Delta_C\rho(s,u)\Big]ds\,du\\ &- \frac{1}{2}\int_0^T\int_{\mathbb{T}^d}\langle(\nabla J)(s,u), C(\nabla J)(s,u)\rangle\rho(s,u)(1-\rho(s,u))ds\,du\Bigg],\end{aligned}\tag{5.11}$$

which was the lower bound we derived.

THEOREM 5.10. *Let $\{\eta_0(x)\}$ be deterministic initial conditions such that*

$$\frac{1}{N^d}\sum_x \delta_{\frac{x}{N}}\eta(x) \to \rho_0(x)dx.$$

Let the process evolve up to time T. Then, denoting by $Q_{N,\eta}$ the distribution of

$$\gamma(s,dx) = \frac{1}{N^d}\sum_x \delta_{\frac{x}{N}}\eta_s(x),$$

a large deviation property holds for $Q_{N,\eta}$ on $C[[0,T];\mathcal{M}(\mathbb{T}^d)]$, with rate function given by (5.11) *on $\rho(s,x)dx$, satisfying $\rho(0,\cdot) = \rho_0(\cdot)$.*

The lower bound needs some tweaking. To establish it, we needed a uniqueness for the weak solution

$$\rho_t = \frac{1}{2}\Delta_C\, \rho - \nabla\cdot(\rho(1-\rho)m).$$

We established it under the condition that m was bounded. Standard calculus of variations provides a formula for m, the minimizer of

$$J(m) = \frac{1}{2}\int_0^T \int_{\mathbb{T}^d} \langle m(s,u), C^{-1}m(s,u)\rangle \rho(s,u)(1-\rho(s,u))ds\, du$$

subject to (5.9), which is

$$m = C\nabla[\nabla\cdot\rho(1-\rho)C\nabla]^{-1}\Big[\rho_t - \frac{1}{2}\Delta_C\,\rho\Big].$$

If ρ is smooth and bounded away from 0 and 1, m is nice. Otherwise, ρ has to be approximated by ρ^ϵ that converges to ρ in such a way that, at the same time, the rate functions $I(\rho_n) \to I(\rho)$. The difficulty is that we cannot change the initial value $\rho(0,u)$. This needs an approximation lemma.

LEMMA 5.11. *Let P_N be the simple exclusion process in equilibrium at density $\frac{1}{2}$ on $\mathbb{Z}_N^d$. Then,*

$$E^{P_N}\Big[\exp\Big[\sum_x J\Big(T,\frac{x}{N}\Big)\eta_T(x) - \sum_x J\Big(0,\frac{x}{N}\Big)\eta_0(x) - \int_0^T \sum_x J_t\Big(t,\frac{x}{N}\Big)\eta_t(x)dt$$
$$-N^2\int_0^T \sum_{x,z}\Big[\exp\Big[J\Big(t,\frac{x+z}{N}\Big) - J\Big(t,\frac{x}{N}\Big)\Big] - 1\Big]\eta_t(x)(1-\eta_t(x+z))\pi(z)dt\Big]\Big] = 1.$$

If Q_N is such that $H(Q_N;P_N)\le cN^d$, and if $\widehat{Q}$ is a limit point of Q_N on $C[[0,T];\mathcal{M}(\mathbb{T}^d)]$,

$$E^{\widehat{Q}}\Big[\int_0^T \frac{1}{2}\Big\|\rho_t - \frac{1}{2}\Delta_C\rho\Big\|^2_{-1,C\rho(1-\rho)} dt\Big] \le c.$$

PROOF. By Jensen's inequality,

$$E^{Q_N}\Big[\frac{1}{N^d}\sum_x J\Big(T,\frac{x}{N}\Big)\eta_T(x) - \sum_x J\Big(0,\frac{x}{N}\Big)\eta_0(x) - \int_0^T \sum_x J_t\Big(t,\frac{x}{N}\Big)\eta_t(x)dt$$
$$- N^2\int_0^T \sum_{x,z}\Big[\exp\Big[J\Big(t,\frac{x+z}{N}\Big) - J\Big(t,\frac{x}{N}\Big)\Big] - 1\Big]\eta_t(x)(1-\eta_t(x+z))\pi(z)dt\Big] \le c. \qquad\square$$

Letting $N\to\infty$ and using the superexponential approximations,

$$E^{\widehat{Q}}\Big[\int_{\mathbb{T}^d} J(T,u)\rho(T,u)du - \int_{\mathbb{T}^d} J(0,u)\rho(0,u)du$$
$$- \int_0^T J_t(t,u)\rho(t,u)du - \frac{1}{2}\int_0^T\int_{\mathbb{T}^d}(\Delta_C J)(t,x)\rho(t,u)dt\, du$$
$$- \frac{1}{2}\int_0^T\int_{\mathbb{T}^d}\langle(\nabla J)(t,u), C(\nabla J)(t,u)\rangle\rho(t,u)(1-\rho(t,u))dt\, du\Big] \le c.$$

Since the sum of two exponentials does not grow any faster than the faster of the two exponentials and the maximum is bounded by the sum, it follows that

$$
\begin{aligned}
E^{\widehat{Q}} \sup_J \Bigg[\int_{\mathbb{T}^d} & J(T,u)\rho(T,u)du - \int_{\mathbb{T}^d} J(0,u)\rho(0,u)du \\
& - \int_0^T J_t(t,u)\rho(t,u)du - \frac{1}{2}\int_0^T \int_{\mathbb{T}^d} (\Delta_C J)(t,x)\rho(t,u)dt\,du \\
& - \frac{1}{2}\int_0^T \int_{\mathbb{T}^d} \langle (\nabla J)(t,u), C(\nabla J)(t,u)\rangle \rho(t,u)(1-\rho(t,u))dt\,du \Bigg] \le c.
\end{aligned}
$$

This implies that

$$\frac{1}{2}\int_0^T \left\| \rho_t - \frac{1}{2}\Delta_C \rho \right\|^2_{-1,\rho(1-\rho)} dt \le c.$$

In particular,

$$\int_0^T \int_{\mathbb{T}^d} \frac{\|\nabla \rho\|^2}{\rho(1-\rho)} dt\,du < \infty,$$

and, in the weak sense,

$$\rho_t - \frac{1}{2}\Delta_C \rho = \nabla m \rho(1-\rho)$$

for some m such that

$$I(\rho) = \int_0^T \int_{\mathbb{T}^d} \langle m, C^{-1} m\rangle \rho(1-\rho)dt\,du < \infty. \tag{5.12}$$

We still have the task of having to perturb the density $\rho(t,u)$ without changing the initial value and not increasing the integral so that we have a bounded m. We proceed as follows. Let ρ be given. We pick δ small enough so that $\rho_\delta(t,u) = \rho(t-2\delta,u)$ is close to $\rho(t,u)$. We pick a function $f(t)$ which is 0 on $[0,\delta']$ for some $\delta' > \delta$, 1 on $[\delta'',T]$ for some $\delta'' < 2\delta$, and varies smoothly in between. We define the approximation that depends on δ and another small parameter ϵ. Let $\rho(t,u)$ be the given profile. We want to prove the lower bound for $\bar{\rho}(t,u)$, the solution of the heat equation with the same initial data. Then, we define $\rho_{\delta,\epsilon}(t,u)$ by

$$\rho_{\delta,\epsilon}(t,u) = \begin{cases} \bar{\rho}(t,u) & \text{if } 0 \le t \le \delta', \\ \bar{\rho}(2\delta - t,u) * \phi_{\epsilon f(t)} & \text{if } t \ge \delta'. \end{cases}$$

The construction involves first running the heat equation for time δ, then reversing the path for δ to 2δ, and continuing with ρ_δ for $t \ge 2\delta$. Convolute by the heat kernel with $\epsilon f(t)$. The change in $f(t)$ occurs where ρ is smooth. The reversed solution of the heat equation does not contribute much for small δ. $\rho_\delta(t) * \phi_{\epsilon f(t)}$ works nicely, as $\epsilon \to 0$. It is not difficult to check that

$$\lim_{\delta\to 0}\lim_{\epsilon\to 0} I(\rho_{\delta,\epsilon}) = I(\rho).$$

It is clear from the variational formula for $I(\rho)$ that it is convex and translation invariant. Therefore,

$$I^T_{2\delta}(\rho_{\epsilon,\delta}) \le I^T_{2\delta}(\rho).$$

Moreover, $I_0^\delta(\rho_{\epsilon,\delta}) = I(\bar\rho) = 0$. By considering the relation

$$\frac{d}{dt}\int_{\mathbb{T}^d} h(\bar\rho(t,u))du = \frac{1}{2}\int_{\mathbb{T}^d} \frac{\langle \nabla\bar\rho, C\nabla\bar\rho\rangle}{\bar\rho(1-\bar\rho)}\, du$$

where $h(\rho) = -[\rho\log\rho + (1-\rho)\log(1-\rho)]$, we can control $I_\delta^{2\delta}(\rho_{\epsilon,\delta})$.

https://doi.org/10.1090//cln/027/06

CHAPTER 6

Self-Diffusion

6.1. Motion of a Tagged Particle

Let us look at the simple exclusion process in equilibrium on $\mathbb{Z}^d$ at density ρ. The distribution is the Bernoulli distribution μ_ρ defined by $\mu_\rho[\eta(x) = 1] = \rho$, with $\{\eta(x) : x \in \mathbb{Z}^d\}$ being independent. Let us suppose that at time 0 there is a particle at 0 which is tagged and observed. It is convenient to move the origin with that particle. The simple exclusion process now acts only on $\mathbb{Z}^d - \{0\}$ and is the environment as seen by the particle. The environment changes in two different ways. When one of the other particles currently at x moves to y, the generator for this part is

$$\mathcal{A}_1 = \frac{1}{2} \sum_{x,y \neq 0} \pi(y-x)[f(\eta^{x,y}) - f(\eta)]. \tag{6.1}$$

Or, the tagged particle moves from 0 to z, and then the origin is shifted to z. This is a transformation T_z that acts when $\eta(z) = 0$, and the new configuration on $\mathbb{Z}^d - \{0\}$ is given by

$$(T_z\eta)(x) = \eta(x+z) \text{ if } x \neq -z, 0, \quad (T_z\eta)(-z) = 0,$$

contributing to the generator the term

$$\mathcal{A}_2 = \sum_z \pi(z)(1-\eta(z))[f(T_z\eta) - f(\eta)]. \tag{6.2}$$

The full generator is, therefore,

$$\begin{aligned}(\mathcal{A}f)(\eta) &= \frac{1}{2} \sum_{x,y \neq 0} \pi(y-x)[f(\eta^{x,y}) - f(\eta)] \\ &\quad + \sum_z \pi(z)(1-\eta(z))[f(T_z\eta) - f(\eta)] \\ &= (\mathcal{A}_1 f)(\eta) + (\mathcal{A}_2 f)(\eta).\end{aligned} \tag{6.3}$$

It is not difficult to check that the probability distribution μ_ρ on $\mathbb{Z}^d - \{0\}$ is a reversible invariant distribution for $\mathcal{A}$, given by (6.3). The jumps $x \to y$ and $y \to x$, as well as T_z and T_{-z}, provide pairs with detail balance. The rates are the same in either direction, and μ_ρ is invariant under the transitions.

Our main tool is a central limit theorem for additive functions of a reversible Markov process. Given a real-valued function f on a space $\mathcal{X}$, a Markov process on that space with generator $\mathcal{A}$, and a reversible ergodic invariant measure μ for $\mathcal{A}$ satisfying $E^\mu[f(x)] = 0$, under suitable conditions, we will show that

$$\int_0^t f(x(s))ds = M(t) + a(t)$$

where $M(t)$ is a square-integrable martingale with stationary increments, and $a(t)$ is negligible. If $\mathcal{A}$ is the self-adjoint generator of the process, $-\mathcal{A}$ has a spectral resolution $-\mathcal{A} = \int_0^\infty \sigma E(d\sigma)$. We have the Dirichlet form $\mathcal{D}(f) = \langle -\mathcal{A}f, f\rangle_{L_2(\mu)}$ associated with $\mathcal{A}$. The space $\mathcal{H}_1$ is the abstract Hilbert space obtained by completing the space of square-integral functions with respect to the Dirichlet inner product. One might start with functions u in the domain of $\mathcal{A}$, ensuring the finiteness of $\mathcal{D}(f)$. The completion will be an abstract space $\mathcal{H}_1$. There will be a dual $\mathcal{H}_{-1}$ to $\mathcal{H}_1$ relative to the inner product of $\mathcal{H}_0 = L_2(\mu)$. Formally, $\|u\|_{-1} = \langle (-\mathcal{A})^{-1}u, u\rangle$, and

$$\|u\|_{-1}^2 = \sup_f 2\langle u, f\rangle - \mathcal{D}(f).$$

We have the following theorem.

THEOREM 6.1. *If f is in L_2 with spectral resolution $\langle E(d\sigma)f, f\rangle$, and*

$$\langle (-\mathcal{A})^{-1}f, f\rangle = \int_0^\infty \sigma^{-1}\langle E(d\sigma)f, f\rangle = \sigma^2 < \infty,$$

there is a square-integrable martingale $M(t)$ with stationary increments such that $E^P[M(t)^2] = 2\sigma^2 t$ and

$$\int_0^t f(x(s))ds = M(t) + a(t)$$

with $E[|a(t)|^2] = o(t)$ as $t \to \infty$. The central limit theorem follows. Moreover,

$$P\Big[\sup_{0\le t\le T} |a(t)| \ge c\sqrt{T}\Big] \to 0$$

for every $c > 0$, implying the functional central limit theorem .

For the proof of the theorem, we need two lemmas.

LEMMA 6.2. *Let P be a reversible stationary Markov process with invariant measure μ and generator $\mathcal{A}$. Let $u \in L_2[\mu]$ with $\mathcal{D}(u) = \langle -\mathcal{A}u, u\rangle < \infty$. Then,*

$$P\Big[\sup_{0\le t\le T} |u(x(t))| \ge \ell\Big] \le \frac{e}{\ell}\sqrt{T\mathcal{D}(u) + \|u\|_2^2}.$$

PROOF. Since $\mathcal{D}(|u|) \le \mathcal{D}(u)$, we can assume that $u \ge 0$. If $x(t)$ is a Markov process and τ is the exit time from G, then $E_x[e^{-\sigma\tau}] = v(\sigma, x)$ is the solution of

$$\sigma v(x) - (\mathcal{A}v)(x) = 0 \text{ for } x \in G, \quad v = 1 \text{ on } G^c.$$

The function v is also the minimizer of

$$\sigma\|v\|_2^2 + \mathcal{D}(v)$$

over v such that $v = 1$ on G^c. Therefore, the solution v_σ satisfies

$$\sigma\|v_\sigma\|_2^2 \le \inf_{v:v=1 \text{ on } G^c}\big[\sigma\|v\|_2^2 + \mathcal{D}(v)\big].$$

If we take for G the set where $u(x) < \ell$, the function $v = (u\wedge\ell)/\ell$ is an admissible choice for v. Therefore, with $\sigma = T^{-1}$,

$$\|v_\sigma\|_1 \le \frac{1}{\ell}\sqrt{\|u\|_2^2 + T\mathcal{D}(u)}.$$

We obtain the estimate

$$\int P_x[\tau < T]d\mu \le e^{\sigma T}\int E_x[e^{-\sigma\tau}]d\mu = e\|v_\sigma\|_1 \le \frac{e}{\ell}\sqrt{\|u\|_2^2 + T\mathcal{D}(u)}. \qquad \square$$

This lemma quantifies the statement that the set of singularities of a function u on $\mathbb{R}^d$ that is in the Sobolev space $W_2^1(\mathbb{R}^d)$ has capacity 0. In other words, even if u has singularities, a Brownian path will not see it; i.e., $u(\beta(t))$ is almost surely continuous.

LEMMA 6.3. *Let $\|u\|_2$ and $\mathcal{D}(u)$ be finite. Then, for any $c > 0$*

$$\limsup_{T\to\infty} P\Big[\sup_{0\le t\le T} |u(x(t))| \ge c\sqrt{T}\Big] = 0.$$

PROOF. For any given $\delta > 0$, find $u' \in L_\infty$ such that $\|u-u'\|_2^2 \le \delta$ and $\mathcal{D}(u-u') \le \delta$. Clearly,

$$\limsup_{T\to\infty} P\Big[\sup_{0\le t\le T} |u'(x(t))| \ge c\sqrt{T}\Big] = 0,$$

and

$$\limsup_{T\to\infty} P\Big[\sup_{0\le t\le T} |(u-u')(x(t))| \ge c\sqrt{T}\Big] \le \frac{e\sqrt{\delta}}{c},$$

and δ can be made arbitrarily small. The lemma is proved. $\square$

PROOF OF THEOREM 6.1. Now, we return to complete the proof of Theorem 6.1. First, let us note that the condition is natural. An elementary calculation shows that

$$\begin{aligned}
\frac{1}{t}E^P\Big[\Big|\int_0^t f(x(s))ds\Big|^2\Big] &= \frac{1}{t}E^P\Big[\int_0^t\int_0^t f(x(s))f(x(s'))ds\,ds'\Big] \\
&= \frac{2}{t}\int_{0\le s\le s'\le t} \langle T_{s'-s}f, f\rangle ds\,ds' \\
&= 2\int_0^t \Big(1-\frac{s}{t}\Big)\langle T_s f, f\rangle ds \\
&\simeq 2\int_0^\infty \langle T_s f, f\rangle ds \\
&= 2\langle(-\mathcal{A})^{-1} f, f\rangle \\
&= 2\sigma^2.
\end{aligned}$$

Since $\langle T_t f, f\rangle \ge 0$, the convergence has to be absolute. Let us solve the resolvent equation

$$\lambda u_\lambda - \mathcal{A}u_\lambda = f.$$

Then, $\mathcal{A}u_\lambda = \lambda u_\lambda - f$, and

$$u_\lambda(x(t)) - u_\lambda(x(0)) - \int_0^t \lambda u_\lambda(x(s))ds + \int_0^t f(x(s))ds = M_\lambda(t)$$

where $M_\lambda(t)$ is a martingale with

$$\frac{1}{t}E\big[M_\lambda(t)^2\big] = 2\mathcal{D}(u_\lambda) = 2\langle -\mathcal{A}u_\lambda, u_\lambda\rangle = 2\int_0^\infty \frac{2\sigma}{(\lambda+\sigma)^2}\langle E(d\sigma)f, f\rangle.$$

An easy computation shows that $(\sigma+\lambda)^{-1} \to \sigma^{-1}$ and is dominated by σ^{-1}, which is integrable with respect to $\langle E(d\sigma)f, f\rangle$. The martingales $M_\lambda(t)$ have a limit in $L_2(P)$, $\lambda u_\lambda \to 0$ in $L_2(\mu)$. Therefore, $a_\lambda(t) = u_\lambda(x(0)) - u_\lambda(x(t))$ has a limit $a(t)$ and

$$\int_0^t f(x(s))ds = M(t) + a(t).$$

We will show that $E[|a(t)|^2] = o(t)$. Then, the martingale central limit theorem will imply our result. This is again a spectral calculation.

$$E^P\big[|a(t)|^2\big] = 2\lim_{\lambda\to 0}\int_0^\infty \frac{1-e^{-t\sigma}}{(\lambda+\sigma)^2}\langle E(d\sigma)f, f\rangle = 2\int_0^\infty \frac{1-e^{-t\sigma}}{\sigma^2}\langle E(d\sigma)f, f\rangle.$$

Since $(1-e^{-t\sigma})/t \le \sigma$ and $\int_0^\infty \frac{1}{\sigma}\langle E(d\sigma)f, f\rangle < \infty$, the dominated convergence theorem implies that

$$\lim_{t\to\infty}\frac{1}{t}E^P\big[|a(t)|^2\big] = 0.$$

To prove the functional central limit theorem, we need to consider

$$\frac{1}{\sqrt{T}}\int_0^{tT} f(x(s))ds = \frac{1}{\sqrt{T}}M_\lambda(Tt) + \frac{1}{\sqrt{T}}\int_0^{tT}\lambda u_\lambda(s)ds - \frac{1}{\sqrt{T}}[u_\lambda(x(tT)) - u_\lambda(x(0))].$$

The functional central limit theorem holds for $(1/\sqrt{T})M_\lambda(tT)$, and uniformly so as $\lambda\to 0$, because $M_\lambda(t)\to M(t)$ in mean-square. We note that, with the help of the dominated convergence theorem,

$$\lambda\|u_\lambda\|^2 = \int_0^\infty \frac{\lambda}{(\lambda+\sigma)^2}\langle E(d\sigma)f, f\rangle \to 0$$

as $\lambda\to 0$. Clearly, with the choice of $\lambda = T^{-1}$,

$$\xi_T = \sup_{0\le t\le 1}\left|\frac{1}{\sqrt{T}}\int_0^{tT}\lambda u_\lambda(s)ds\right| \le \frac{1}{T}\int_0^T \sqrt{\lambda}|u_\lambda(x(s))|ds,$$

and $E\big[|\xi_T|^2\big]\to 0$ as $t\to\infty$. To complete the proof, we need to show that, with $\lambda = T^{-1}$,

$$P\Big[\sup_{0\le t\le T}|u_{1/T}(x(s))| \ge c\sqrt{T}\Big] \to 0.$$

We can represent $u_{1/T}$ as $u_\delta + (u_{1/T} - u_\delta)$. By Lemma 6.3, we have for any $\delta > 0$,

$$\limsup_{T\to\infty} P\Big[\sup_{0\le s\le T}|u_\delta(x(s))| \ge c\sqrt{T}\Big] = 0.$$

Moreover,

$$\lim_{T\to\infty}\frac{1}{T}\|u_{1/T} - u_\delta\|_2^2 = 0$$

and

$$\lim_{\substack{T\to\infty,\\ \delta\to 0}} D(u_{1/T} - u_\delta) = 0.$$

They imply that

$$\limsup_{\delta\to 0}\limsup_{T\to\infty} P\Big[\sup_{0\le t\le T}|u_{1/T}(x(s)) - u_\delta(x(s))| \ge c\sqrt{T}\Big] = 0. \qquad \square$$

We now return to the motion of the tagged particle. We need to keep track of its motion, as well as the changing environment seen by it. If $w\in\mathbb{Z}^d$ is the location of the tagged particle in the original reference frame, then jointly the generator for $w(t)\in\mathbb{Z}^d$ and $\eta(\cdot)\in\{0,1\}^{\mathbb{Z}^d-\{0\}}$ is

$$(\widetilde{\mathcal{A}}f)(\eta) = \sum_z \pi(z)(1-\eta(z))[f(w+z, T_z\eta) - f(w,\eta)] + (\mathcal{A}f)(w,\eta), \tag{6.4}$$

with $\mathcal{A}$ acting only on η for each w. The $\eta(t, \cdot)$ part is a Markov process by itself and is in equilibrium at density ρ, the distribution being μ_ρ. We are interested in establishing a central limit theorem for $w(t)$. We note that

$$w(t) - w(0) = \int_0^t \sum_z z\,\pi(z)(1 - \eta(s,z))ds + M(t)$$

where $M(t)$ is a martingale with the decomposition

$$M(t) = \sum_z z M_z(t)$$

and

$$M_z(t) = N_z(t) - \pi(z)\int_0^t (1 - \eta(s,z))ds.$$

The quantity

$$V(\eta(\cdot)) = \sum_z z\,\pi(z)(1 - \eta(z))$$

has mean 0 in equilibrium, and one may expect a central limit theorem for

$$\int_0^t V(\eta(s, \cdot))ds.$$

We will prove a decomposition of the form

$$\int_0^t V(\eta(s, \cdot))ds = N(t) + a(t),$$

where $N(t)$ is a martingale and $a(t)$ is negligible. Then,

$$w(t) - w(0) = M(t) + N(t) + a(t),$$

and, since the central limit theorem for martingales is automatic, the result will follow. The quantities here are vectors and the equations are for each component, or they are interpreted as

$$\langle w(t) - w(0), \xi\rangle = \langle M(t), \xi\rangle + \langle N(t), \xi\rangle + \langle a(t), \xi\rangle$$

for $\xi \in \mathbb{R}^d$. We have now the main theorem.

THEOREM 6.4. *The position $w(t)$ of the tagged particle satisfies a functional central limit theorem , with positive definite covariance matrix $S(\rho)$ given by*

$$\langle S(\rho)\xi, \xi\rangle = \inf_f \Big[\int \Big[\sum_z \pi(z)(1 - \eta(z))\big(\tau_z f - f - \langle \xi, z\rangle\big)^2 + \frac{1}{2}\sum_{x,y}\pi(y - x)(f(\eta^{x,y}) - f(\eta))^2\Big]d\mu_\rho\Big]. \tag{6.5}$$

First, we need to prove that for each vector $\xi \in \mathbb{R}^d$ there exists a bound of the form

$$\Big|\int \sum_z \langle z, \xi\rangle(1 - \eta(z))\pi(z) f(\eta) d\mu_\rho\Big| \le \sqrt{C(\xi)}\sqrt{D_\rho(f)}.$$

We can rewrite, after combining the z and $-z$ terms and symmetrizing,

$$\begin{aligned}
E^{\mu_\rho}\Big[\sum_z \langle z,\xi\rangle(1-\eta(z))\pi(z)f(\eta)\Big] &= \frac{1}{2}E^{\mu_\rho}\Big[\sum_z \langle z,\xi\rangle\big[(1-\eta(z))-(1-\eta(-z))\big]\pi(z)f(\eta)\Big]\\
&= \frac{1}{2}E^{\mu_\rho}\Big[\sum_z \langle z,\xi\rangle[1-\eta(z))]\pi(z)[f(\eta)-f(T_z\eta)]\Big]\\
&\le \frac{1}{2}\Big[E^{\mu_\rho}\Big[\sum_z |\langle z,\xi\rangle|^2[1-\eta(z))]\pi(z)\Big]\Big]^{\frac{1}{2}}\\
&\quad\cdot\Big[E^{\mu_\rho}\Big[\sum_z[1-\eta(z))]\pi(z)[f(\eta)-f(T_z\eta)]^2\Big]\Big]^{\frac{1}{2}}\\
&\le \sqrt{C(\xi)}\sqrt{\mathcal{D}_\rho(f)}
\end{aligned}$$

with

$$C(\xi)=\frac{1-\rho}{4}\sum_z |\langle z,\xi\rangle|^2\pi(z).$$

This proves the validity of a functional central limit theorem for $w(t)$ with an upper bound on the variance.

The next step is to establish the formula and a lower bound for it. Let us compute $\langle S(\rho)\xi,\xi\rangle$. The minimizer $f = f_\xi$ may not exist. The space $\mathcal{H}_1$ of functions $u \in L_2$, with the Dirichlet inner product, when completed, will admit objects that are not in $L_2(\mu_\rho)$. There is no Poincaré inequality available. Abstractly, the space consists of collections of functions $\{g^{x,y}(\eta)\}, \{g_z\}$, that are the limits in H_1 of $\{f(\eta^{x,y})-f(\eta)\}$, $(1-\eta(z))[f(T_z\eta)-f(\eta)]$. The functions $g^{x,y}(\eta), g_z$ satisfy identities. $g^{x,y}$ is 0 unless $\eta(x)\neq\eta(y)$ and satisfies $g^{x,y}(\eta)+g^{x,y}(\eta^{x,y}) = 0$. Similarly, g_z is nonzero only when $\eta(z) = 0$ and $(1-\eta(z))g_z(\eta)+(1-\eta(-z))g_{-z}(T_{-z}\eta) = 0$. The Euler equation for the variational problem is

$$\begin{aligned}
&E^{\mu_\rho}\Big[\sum_z \pi(z)(1-\eta(z))[g_z(\eta)-\langle\xi,z\rangle][f(\tau_z\eta)-f(\eta)]\Big]\\
&\qquad\qquad +\frac{1}{2}E^{\mu_\rho}\Big[\sum_{x,y}\pi(y-x)g^{x,y}(\eta)[f(\eta^{x,y})-f(\eta)]\Big]=0
\end{aligned}$$

for all f, which, after a bit of calculation, takes the form

$$V(\eta)+\frac{1}{2}\sum\pi(y-x)g^{x,y}+\sum_z\pi(z)(1-\eta(z))g_z=0.$$

$w(t)$ now has the representation

$$\langle\xi,w(t)\rangle=\int_0^t\langle\xi,V(\eta(s))\rangle ds+\sum_z\int_0^t\langle\xi,z\rangle(1-\eta(z))dM_z(t)$$

with

$$\int_0^t\big\langle\xi,V(\eta_s))\big\rangle ds=a(t)+N(t)$$

and

$$N(t)=\sum_z\int_0^t g_z(\eta_s)dM_z(s)+\sum_{x,y}\int_0^t g^{x,y}(\eta_s)M_{x,y}(t)$$

with

$$M_{x,y}(t) = N_{x,y}(t) - \int_0^t \pi(y-x)\eta_s(x)(1-\eta_s(y))ds.$$

Therefore,

$$\begin{aligned}\langle \xi, w(t)\rangle &= \int_0^t \sum_z [\langle z,\xi\rangle - g_z(\eta_s)]dM_z(s) - \int_0^t \sum_{x,y} g_{x,y}(\eta_s)dN_{x,y}(s) + a(t)\\ &= M(t) + a(t).\end{aligned}$$

Computing the quadratic variation of the martingale $M(t)$ proves the formula (6.5).

Finally, we will prove the nondegeneracy of the quadratic form $\langle S(\rho)\xi, \xi\rangle$. We have to exclude the one-dimensional nearest-neighbor case, where $S(\rho) \equiv 0$. The proof depends on the following fact. We can obtain an estimate of the form

$$E^{\mu_\rho}[(\eta(z) - \eta(-z))f(\eta)] \le C\sqrt{\mathcal{D}_1(f)}$$

in terms of the Dirichlet form

$$\mathcal{D}_1(u) = \langle -\mathcal{A}_1 u, u\rangle = \frac{1}{4}E^{\mu_\rho}\Big[\sum_{x,y}\pi(y-x)[f(\eta^{x,y}) - f(\eta)]^2\Big].$$

It is possible to shift a particle from z to $-z$ without touching the tagged particle at 0. Jump over it or go around it. This provides an estimate of the form

$$E^{\mu_\rho}[(\eta(z) - \eta(-z))f(\eta)] \le C(z)\Big[E^{\mu_\rho}\sum_{x,y}\big[\pi(y-x)[f(\eta^{x,y}) - f(\eta)]^2\big]\Big]^{\frac{1}{2}}. \tag{6.6}$$

We can estimate for any $a > 0$

$$\begin{aligned}\big\langle(\lambda I - \mathcal{A})^{-1}\langle V,\xi\rangle, \langle V,\xi\rangle\big\rangle &\le \sqrt{\big\langle -\mathcal{A}_1(\lambda I - \mathcal{A})^{-1}\langle V,\xi\rangle, (\lambda I - \mathcal{A})^{-1}\langle V,\xi\rangle\big\rangle}\\ &\quad\cdot\sqrt{\big\langle -\mathcal{A}_1)^{-1}\langle V,\xi\rangle, \langle V,\xi\rangle\big\rangle}\\ &\le \frac{a}{2}\big\langle(-\mathcal{A}_1(\lambda I - \mathcal{A})^{-1}\langle V,\xi\rangle, (\lambda I - \mathcal{A})^{-1}\langle V,\xi\rangle\big\rangle\\ &\quad + \frac{1}{2a}\big\langle(-\mathcal{A}_1)^{-1}\langle V,\xi\rangle, \langle V,\xi\rangle\big\rangle.\end{aligned}$$

Letting $\lambda \to 0$,

$$\begin{aligned}\langle S(\rho)\xi,\xi\rangle &\ge \big\langle -\mathcal{A}_1(-\mathcal{A})^{-1}\langle V,\xi\rangle, (-\mathcal{A})^{-1}\langle V,\xi\rangle\big\rangle\\ &\ge \frac{2}{a}\big\langle(-\mathcal{A})^{-1}\langle V,\xi\rangle, \langle V,\xi\rangle\big\rangle - \frac{1}{a^2}\big\langle(-\mathcal{A}_1)^{-1}\langle V,\xi\rangle, \langle V,\xi\rangle\big\rangle.\end{aligned}$$

We can obtain a lower bound for the first quadratic form on the right from the variational formula

$$\langle(-\mathcal{A})^{-1}g, g\rangle = \sup_f[2\langle g, f\rangle - \langle -\mathcal{A}f, f\rangle],$$

and an upper bound for the second one from (6.6). Picking a large will do it.

For later use, we will need some properties of $S(\rho)$ as a function of ρ. It is known that it is infinitely differentiable on $0 \le \rho \le 1$ and satisfies $S(0) = C$, $S(1) = 0$, and $S(\rho) \ge c(1-\rho)I$.

https://doi.org/10.1090//cln/027/07

CHAPTER 7

Nongradient Systems

7.1. Multicolor Systems

Let us look at the situation where there are K types of particles. For convenience, we will take the types to be K colors. The state space is $\Omega_N = \{F\}^{N^d}$ where F is the finite set of $K+1$ points representing the presence of a particle of color type $\{1, \dots, K\}$ or no particle at any given site of $\mathbf{Z}_N^d$. There can be at most one particle at any site. We define

$$
\begin{aligned}
\zeta_i(x) &= 1 && \text{if there is a particle of type } i \text{ at } x \text{ and } 0 \text{ otherwise,}\\
\eta(x) &= \textstyle\sum_{i=1}^K \zeta_i(x) = 1 && \text{if there is a particle (of any type) at } x,\\
\eta(x) &= 0 && \text{if there is no particle at } x,\\
\zeta(x) &= \{\zeta_i(x)\},\\
\zeta &= \{\zeta(x)\},\\
\eta &= \{\eta(x)\}.
\end{aligned}
$$

The state evolves with time. We have $K+1$ empirical measures. For $i = 1, \dots, K$

$$
\lambda_i(s, d\theta) = \frac{1}{N^d} \sum_x \delta_{x/N} \zeta_i(s, x) \tag{7.1}
$$

and

$$
\lambda(s, d\theta) = \frac{1}{N^d} \sum_x \delta_{x/N} \eta(s, x) = \sum_{i=1}^K \lambda_i(s, d\theta). \tag{7.2}
$$

The evolution is specified by the generator of the Markov process quite similar to the old one of a single type.

$$
\begin{aligned}
\big(N^2 \mathcal{A}_N F\big)(\zeta) &= N^2 \sum_{x,y \in \mathbf{Z}_N^d} \pi(y-x)\eta(x)(1-\eta(y))[F(\zeta^{x,y}) - F(\zeta)]\\
&= \frac{N^2}{2} \sum_{x,y \in \mathbf{Z}_N^d} \pi(y-x) a_{x,y}(\zeta)[F(\zeta^{x,y}) - F(\zeta)] \qquad (7.3)
\end{aligned}
$$

where

$$
a_{x,y}(\zeta) = a_{x,y}(\eta) = [\eta(x)(1-\eta(y)) + \eta(y)(1-\eta(x))], \tag{7.4}
$$

and $\zeta^{x,y}$ interchanges the situations at x and y. $a_{x,y}(\zeta)$ is either 0 or 1. When it is 1, one of the two sites is empty, and either a jump from x to y, or one from y to x, can occur with equal rate $\pi(y-x) = \pi(y-x)$. We will assume for simplicity that $\pi(\pm e_r) > 0$ in any coordinate direction; i.e., the rate for jumping to a nearest neighbor in any coordinate direction is positive. We also need either $d \geq 2$ or, if $d = 1$, $\pi(z) > 0$ for some z with

$|z| \geq 2$. Otherwise, if $d = 1$, relative positions of the particles do not change and the colors do not mix.

The particles evolve as before and are not affected by their type. But we keep track of their types. Let $k_i = k_i(N)$ be the number of particles of type i and $k(N) = \sum_{i=1}^{K} k_i(N)$ be the total number of particles. If $k(N) \leq N^d - 2$, i.e., if there are at least two empty sites, then the only invariant distribution for the Markov process is the uniform distribution $\mu_{N,\mathbf{k}(N)}$ over all possible configurations. Here $\mathbf{k}(N)$ stands for the collection $\{k_i(N)\}$. In the limit as $N \to \infty$, assuming $k_i(N)N^{-d} \to \rho_i$ for $i = 1, \dots, K$, one has a product measure $\mu_{\tilde{\rho}}$ on $\mathbb{Z}^d$ where $\tilde{\rho} = \{\rho_i\}$ and $\rho = \sum_{i=1}^{K} \rho_i = \rho \leq 1$. Each site x has a particle of type i with probability ρ_i and is empty with probability $1-\rho$. Different sites are independent. The situation with $\rho = 1$ is the extreme case, where there is no movement and every configuration is static.

There are Dirichlet forms associated with these processes given by

$$
\begin{aligned}
\langle -\mathcal{A}_N f, f\rangle_{\mathbf{k}(N)} &= \mathcal{D}^N_{\mathbf{k}(N)}(f) \\
&= \frac{1}{4} E^{\mu_{N,\mathbf{k}(N)}}\Big[\sum_{x,y\in\mathbf{Z}_N^d} a_{x,y}(\zeta)\pi(y-x)[f(\zeta^{x,y}) - f(\zeta)]^2\Big] \\
&= \frac{1}{2} E^{\mu_{N,\mathbf{k}(N)}}\Big[\sum_{x,y\in\mathbf{Z}_N^d} \eta(x)(1-\eta(y))\pi(y-x)[f(\zeta^{x,y}) - f(\zeta)]^2\Big] \qquad (7.5)
\end{aligned}
$$

and the similar form

$$
\begin{aligned}
\mathcal{D}_{\tilde{\rho}}(f) &= \frac{1}{4} E^{\mu_{\tilde{\rho}}}\Big[\sum_{x,y\in\mathbb{Z}^d} a_{x,y}(\zeta)\pi(y-x)[f(\zeta^{x,y}) - f(\zeta)]^2\Big] \\
&= \frac{1}{2} E^{\mu_{\tilde{\rho}}}\Big[\sum_{x,y\in\mathbb{Z}^d} \eta(x)(1-\eta(y))\pi(y-x)[f(\zeta^{x,y}) - f(\zeta)]^2\Big] \qquad (7.6)
\end{aligned}
$$

on $\mathbb{Z}^d$.

We can also consider our process in a box $\mathbb{B}_q^d$ of size $(2q+1)^d$, without assuming a periodic boundary. Jumping outside the box is not allowed. In this case, a minimal number n_0 of empty sites that depends only on $\pi(\cdot)$ and d are needed to ensure uniqueness of the uniform distribution $\mu_{q,\mathbf{k}}$ as the only invariant distribution. If $\pi(\pm e_i) > 0$ for all i and $d \geq 2$, then n_0 can be taken to be 2. The operator and the Dirichlet form look exactly the same as before, except x and y are now restricted to $\mathbb{B}_q^d$. It is easy to verify that with a fixed number K of colors, the space of configurations in any finite box is irreducible. For irreducibility, rates are not relevant, only the set of possible moves. Basically, the empty sites can change positions with any neighbor. Any two particles at neighboring sites can change positions if they have two empty sites directly above, below, or on any adjacent side. Denoting their successive positions by

$$
\begin{matrix} 1 & 2 \\ e & e \end{matrix} \qquad \begin{matrix} e & 2 \\ 1 & e \end{matrix} \qquad \begin{matrix} e & 2 \\ e & 1 \end{matrix} \qquad \begin{matrix} 2 & e \\ e & 1 \end{matrix} \qquad \begin{matrix} 2 & 1 \\ e & e \end{matrix}
$$

we see that two neighboring particles can exchange positions assisted by two empty spaces. Empty spaces can move freely to come, help the exchange, and move back to where they came from, undoing the local dislocation they have created. This only requires that they return by exactly reversing their steps. When $d = 1$ and $\pi(z) > 0$ for some $z > 2$, and the group generated by $\{z : \pi(z) > 0\}$ contains ± 1, it is possible to design a more complex system of allowable exchanges that effects the exchange of two neighboring particles. We have to bring in a sufficient number of empty sites and place them next to the two sites where the particles to exchange positions are located. Then effect the exchange and go back by reversing the steps in the exact reverse order.

We take K smooth test functions $\widetilde{J}(\theta) = \{J^{(i)}(\theta)\}$ on the torus $\mathbb{T}^d$ and consider

$$F_{\widetilde{J}}(\zeta) = \frac{1}{N^d} \sum_{x \in \mathbf{Z}_N^d} \sum_{i=1}^{K} J^{(i)}\Big(\frac{x}{N}\Big)\zeta_i(x). \tag{7.7}$$

Next, we compute $N^2 \mathcal{A}_N F_{\widetilde{J}}(\zeta)$ as

$$N^{2-d} \sum_{x,y} \pi(y-x) \sum_{i=1}^{K} \zeta_i(x)(1-\eta(y)) \Big[J^{(i)}\Big(\frac{y}{N}\Big) - J^{(i)}\Big(\frac{x}{N}\Big)\Big] \tag{7.8}$$

and can approximate it using the symmetry of $\pi(z)$ with an error of at most $o(1)$ by

$$\frac{N^{1-d}}{2} \sum_{x,y} \pi(y-x) \sum_{i=1}^{K} \Big[\zeta_i(x)(1-\eta(y)) \Big(\nabla J^{(i)}\Big(\frac{y}{N}\Big) + \nabla J^{(i)}\Big(\frac{x}{N}\Big)\Big) \cdot (y-x)\Big]. \tag{7.9}$$

Or, by using the symmetry of $\pi(z)$ and interchanging x and y,

$$\begin{aligned} &\frac{N}{4N^d} \sum_{x,y} \pi(y-x) \\ &\cdot \sum_{i=1}^{K} \big[\zeta_i(x)(1-\eta(y)) - \zeta_i(y)(1-\eta(x))\big] \Big(\nabla J^{(i)}\Big(\frac{y}{N}\Big) + \nabla J^{(i)}\Big(\frac{x}{N}\Big)\Big) \cdot (y-x). \end{aligned} \tag{7.10}$$

For $i = 1, \dots, K$, the d components $\mathbf{f}_r^i$ of the current for the K types are given by

$$\mathbf{f}_r^i(\zeta) = \frac{1}{2} \sum_z \pi(z) \langle z, e_r \rangle \big[\zeta_i(0)(1-\eta(z)) - \zeta_i(z)(1-\eta(0))\big]. \tag{7.11}$$

The extra factor of N in (7.10) is a problem that will not go away. We cannot do a summation by parts to bring in the second difference. $\{\mathbf{f}_r^i\}$ are not gradients. Note that their sum,

$$\begin{aligned} \sum_{i=1}^{K} \mathbf{f}_r^i &= \sum_z \pi(z) \langle z, e_r \rangle [\eta(0)(1-\eta(z)) - \eta(z)(1-\eta(0))] \\ &= \frac{1}{2} \sum_z \pi(z) \langle z, e_r \rangle [\eta(0) - \eta(z)] \\ &= \frac{1}{2} \sum_z \pi(z) \langle z, e_r \rangle [\eta(0) - (\tau_z \eta)(0)], \end{aligned}$$

is a "gradient" that allows for another summation by parts that gets rid of the unwelcome factor of an extra N when we do not distinguish between the K types of particles.

If we decompose

$$d \frac{1}{N^d} \sum J\Big(\frac{x}{N}\Big) \zeta_i(t,x) = A_N(t)dt + dM_N(t)$$

into a function of bounded variation and a martingale, the martingale has jumps of size $1/N^{d+1}$ with a total intensity of $O(N^{d+2})$. This makes

$$E\Big[\sup_{0\le t\le T}|M_N(t)|^2\Big]\le CN^{-d}$$

and therefore negligible. Estimating $A_N(t)$ is the problem. While $|A_N(t)|\le C$ in the gradient case of a single color, here $|A_N(t)|=O(N)$ and is bounded only when integrated over time. We need to estimate quantities of the type

$$E^{Q_N}\left[\left|N\int_s^t V(\sigma,\zeta_\sigma)d\sigma\right|\right] \tag{7.12}$$

where Q_N is a perturbation of the process P_N that is in equilibrium with some stationary marginal μ, $H(Q_N;P_N)\le CN^d$, and

$$V(s,\zeta)=\frac{1}{N^d}\sum_{x\in\mathbb{Z}^d}a\Big(s,\frac{x}{N}\Big)g(s,\tau_x\zeta)$$

with $g(s,\zeta)$ being a local function with mean 0 under every equilibrium. If we can estimate

$$\frac{1}{N^d}\log E^{P_N}\left[\exp\left[N^{d+1}\int_s^t V(\sigma,\zeta_\sigma)d\sigma\right]\right]$$

by the Feynman-Kac formula, then we can use Jensen's inequality to estimate (7.12). The absolute value is not a problem because $e^{|x|}\le e^x+e^{-x}$, and $-V$ is as good as V.

We need the following lemma to make uniform estimates in the Feynman-Kac formula.

LEMMA 7.1. *Let L be an infinitesimal generator of a Markov semigroup T_t on X, and V multiplication by a bonded function $V(x)$. Then, if $\mathcal{D}(f)$ is the Dirichlet form of L with respect to the reversible invariant measure μ,*

$$\begin{aligned}E^\mu\left[\exp\left[\int_0^T V(x(s))ds\right]\right]&\le\exp\left[T\left[\sup_{f:\|f\|_2=1}\int V(x)[f(x)]^2d\mu-\mathcal{D}(f)\right]\right]\\&=\exp\left[T\left[\sup_{\substack{f\ge0\\ \|f\|_1=1}}\int V(x)f(x)d\mu-\mathcal{D}\big(\sqrt{f}\big)\right]\right].\end{aligned}$$

PROOF. Let us denote by $g(t)=e^{t(L+V)}1$.

$$\frac{d}{dt}\|g(t)\|^2=2\langle(L+V)g(t),g(t)\rangle\le2\|g(t)\|^2\sup_{g:\|g\|_2=1}\langle(L+V)g,g\rangle=2\lambda\|g(t)\|^2$$

where

$$\lambda=\sup_{g:\|g\|_2=1}\left[\int V(x)[g(x)]^2d\mu-\mathcal{D}(g)\right].$$

This provides the estimate $\|g(t)\|_1\le\|g(t)\|_2\le e^{\lambda t}$. This works just as well if V depends on s explicitly. Then, if $\lambda(s)$ is the bound for $V(s,x)$ in the variational formula, we obtain the estimate

$$\log E^\mu\left[\exp\left[\int_0^T V(x(s))ds\right]\right]\le\int_0^T\lambda(s)ds$$

where

$$\lambda(s)=\sup_{f:\|f\|_2=1}\left[\int V(s,x)[f(x)]^2d\mu-\mathcal{D}(f)]\right]. \qquad \square$$

We can rewrite the Dirichlet form as

$$\mathcal{D}(f) = \frac{1}{4}\sum_x \sum_z E^{\mu}\big[a_0(x, x+z)\pi(z)[f(\zeta^{x,x+z}) - f(\zeta)]^2\big]$$
$$= \sum_x \mathcal{D}_x(f) = \frac{1}{|B|}\sum_x \sum_{y\in B+x} \mathcal{D}_y(f)$$

where μ is an invariant measure on $\mathbb{Z}^d$,

$$\mathcal{D}_x(f) = \frac{1}{4}\sum_z E^{\mu}\big[a_0(x, x+z)\pi(z)[f(\zeta^{x,x+z}) - f(\zeta)]^2\big],$$

and B is some fixed box around 0 of size $|B|$. If we have a local function $g(\zeta)$ around 0 that satisfies a bound of the type

$$|E^{\mu}[g(\zeta)f]| \le C\Big[\sum_{y\in D_q} \mathcal{D}_y\big(\sqrt{f}\big)\Big]^{\frac{1}{2}}$$

for all f that are densities with respect to μ and some cube D_q, then

$$E^{\mu}\Big[\sum_x a(x)g(\tau_x\zeta)f\Big] - \sum_x \mathcal{D}_x(f)$$
$$\le C\sum_x |a(x)|\Big[\sum_{y\in D_q+x} \mathcal{D}_y\big(\sqrt{f}\big)\Big]^{\frac{1}{2}} - \sum_x \mathcal{D}_x(f)$$
$$= C\sum_x |a(x)|\Big[\sum_{y\in D_q+x} \mathcal{D}_y\big(\sqrt{f}\big)\Big]^{\frac{1}{2}} - \frac{1}{(2q+1)^d}\sum_x \sum_{y\in D_q+x} \mathcal{D}_y(f)$$
$$\le C^2(2q+1)^d \sum_x [a(x)]^2.$$

We will often have the need to estimate

$$E^{\mu}\Big[\Big[N\sum_x a\Big(\frac{x}{N}\Big)g(\tau_x\zeta)f\Big] - N^2\sum_x \mathcal{D}_x(f)\Big] =$$
$$N^2\Big[E^{\mu}\Big[\frac{1}{N}\sum_x a\Big(\frac{x}{N}\Big)g(\tau_x\zeta)f\Big] - \sum_x \mathcal{D}_x(f)\Big].$$

It is not difficult to prove that for a finite Markov chain with an ergodic, reversible, invariant measure μ and Dirichlet form $\mathcal{D}(f)$ for a potential $V(x)$ with mean 0, the eigenvalue

$$\lambda(\epsilon) = \sup_{\|f\|_2=1}\Big[\epsilon\int V(x)[f(x)]^2 d\mu - \mathcal{D}(f)\Big]$$

satisfies

$$\lim_{\epsilon\to 0}\frac{\lambda(\epsilon)}{\epsilon^2} = \frac{\sigma^2(V)}{2}.$$

Here

$$\sigma^2(V) = \lim_{t\to\infty}\frac{1}{t}E^{\mu}\Big[\Big[\int_0^t V(x(s))ds\Big]^2\Big].$$

Although it may look like we are estimating exponential moments for a large class of functions we are interested in, because of diffusive scaling we really only need to estimate variances in the central limit theorem.

The expectations of the currents $\{\mathbf{f}_r^i\}$ in any equilibrium $\mu_{\tilde\rho}$ are easily calculated and are equal to 0. We need to understand this combination of large N and currents $\mathbf{f}$ that are small on average. We will determine constants $\{c_{r,r'}^{i,j}\}$ with $r, r' = 1, \dots, d$ and $1 \le i, j \le K$ (that are actually functions of $\tilde\rho$) such that

$$\mathbf{f}_r^i + \frac{1}{2}\sum_{j,r'} c_{r,r'}^{i,j}[\zeta_j(e_{r'}) - \zeta_j(0)] = w_r^i \tag{7.13}$$

and $\{w_r^i\}$ are negligible. The sense in which they are negligible has to be specified. They will become negligible only when integrated over time. The context will be relative to the process in equilibrium under the measure $\mu_{\tilde\rho}$. This makes the constants $c_{r,r'}^{i,j}$ functions of $\tilde\rho$. We can now do another summation by parts and get rid of the extra N, replacing $\{N(\zeta_j(e_{r'}) - \zeta_j(0))\}$ by $\partial\rho_j/\partial\theta_{r'}$. We end up with a weak formulation

$$\frac{\partial}{\partial t}\sum_i \langle J_i, \rho_i\rangle - \frac{1}{2}\sum_i \left\langle J_i \sum_{j,r,r'} \frac{\partial}{\partial\theta_r} c_{r,r'}^{i,j}(\tilde\rho)\frac{\partial\rho_j}{\partial\theta_{r'}}\right\rangle = 0 \tag{7.14}$$

of the elliptic system

$$\frac{\partial\rho_i}{\partial t} = \frac{1}{2}\sum_{j,r,r'} \frac{\partial}{\partial\theta_r} c_{r,r'}^{i,j}(\tilde\rho(\theta))\frac{\partial\rho_j}{\partial\theta_{r'}}. \tag{7.15}$$

The approximations will be good enough, and it will be possible to do large deviations as well.

7.2. Tightness Estimates

We will need various tightness estimates as we proceed. Since our goal is to prove a large deviation estimate for $\mathcal{R}_N$, the distribution of

$$\gamma = \frac{1}{N^d}\sum_{i=1}^{k(N)} \frac{\delta_{x_i(N^2\cdot)}}{N} \in \mathcal{M}[[D[0,T];\mathbb{T}^d]]$$

on the space $\mathcal{M}(D[0,T])$ of stochastic processes on $D[[0,T];\mathbb{T}^d]$, we will need an exponential tightness estimate. This will establish tightness under all perturbations with relative entropy bounded by CN^d. The process, the trajectory of the i^{th} particle, can be represented as $x_i^N(t) = \xi_i^N(t) + M_i^N(t)$. $\xi_i^N(t)$ and $M_i^N(t)$ are processes with values in $\mathbb{R}^d$, and $x_i^N(t)$ is the projection of their sum to $\mathbb{T}^d$. Since the jumps are small, $x_i^N(t)$ can be lifted in a canonical manner from $\mathbb{T}^d$ to $\mathbb{R}^d$ and decomposed there as the sum of a martingale $M_i^N(t)$ and a function of bounded variation $\xi_i^N(t)$ with values in $\mathbb{R}^d$. Moreover, $\xi_i^N(t)$ has a representation

$$\xi_i^N(t) = N\int_0^t b_i(\mathbf{x}(s))ds$$

with b_i of the form

$$\begin{aligned} b_i(\mathbf{x}) &= b_i(x_1, \dots, x_{k_N}) \\ &= \sum_z (1 - \eta(x_i + z))\langle z, e\rangle\pi(z) \\ &= \sum_{x,y}(1 - \eta(y))\pi(y - x)\langle y - x, e\rangle\mathbb{1}_{x_i = x} \end{aligned}$$

$$= -\sum_{x,y}(1-\eta(x))\pi(y-x)\langle y-x, e\rangle \mathbb{1}_{x_i=y}$$
$$= \frac{1}{2}\sum_{x,y}\pi(y-x)\langle y-x, e\rangle[(1-\eta(y))\mathbb{1}_{x_i=x} - (1-\eta(x))\mathbb{1}_{x_i=y}].$$

We will appeal to a result of Garsia-Rodemich-Rumsey.

LEMMA 7.2. *There are constants C_1, C_2 such that for any continuous function ϕ on* $[0, T]$

$$S_T(\phi) = \sup_{0\le\delta\le\frac{1}{2}} \sup_{\substack{0\le s\le T,\\ |s-t|\le\delta}} \frac{|\phi(t)-\phi(s)|}{\sqrt{\delta}\log\frac{1}{\delta}}$$
$$\le C_1 + C_2\log^+\int_0^T\int_0^T \exp\left\{\frac{|\phi(t)-\phi(s)|}{\sqrt{|t-s|}}\right\}dt\,ds. \tag{7.16}$$

Their results are in terms of the choice of two functions, $\Psi(x)$ and $p(x)$. To get our estimate, we need to take $\Psi(x) = e^{|x|}-1$ and $p(x) = \sqrt{x}$. See [**23**] for details. The main step is to obtain an estimate of the form

$$\frac{1}{N^d}\log E^{P_N}\Big[\exp\Big[\alpha\sum_i S_T(\xi_i(\cdot))\Big]\Big] \le C_3$$

for some $\alpha > 0$. By Jensen's inequality, this would yield an estimate for Q_N with $H(Q_N; P_N) \le C_4N^d$,

$$E^{Q_N}\left[\frac{1}{N^d}\sum_i S_T(\xi_i(\cdot))\right] \le C_5,$$

which is enough to prove the tightness of $\mathcal{R}_N$. We need to estimate

$$E^{P_N}\left[\int_0^T\cdots\int_0^T\prod_i\exp\left[\frac{|\xi_i(t_i)-\xi_i(s_i)|}{\sqrt{|t_i-s_i|}}\right]\right]dt_i\,ds_i.$$

We can replace $e^{|x|}$ by e^x+e^{-x} and pick $s_i \le t_i$. This adds at most a factor of 2^{N^d} twice. It is enough to prove the following lemma and we can then choose $\lambda_i = (t_i-s_i)^{1/2}$.

LEMMA 7.3. *For any choice of λ_i,*

$$E^{P^N}\Big[\exp\Big[\sum\lambda_i(\xi(t_i)-\xi(s_i))\Big]\Big] \le \exp\Big[C_6\sum_i\lambda_i^2(t_i-s_i)\Big]$$

holds uniformly with some constant C_6.

PROOF. For $t\in[0,T]$, let us define $k(t) = \{i : t\in[s_i,t_i]\}$. Then, with $\mathbf{x} = \{x_i\}$,

$$V(t,\mathbf{x}) = \sum_{i\in k(t)}\lambda_i\sum_{x,y}(y-x)\pi(y-x)[(1-\eta(y))\mathbb{1}_{x_i=x} - (1-\eta(x))\mathbb{1}_{x_i=y}].$$

For any density f with respect to μ,

$$|E^\mu[V(t,\mathbf{x})f]|$$
$$= \left|E^\mu\Big[-\frac{1}{2}\sum_{i\in k(t)}\lambda_i\sum_{x,y}(y-x)\pi(y-x)[(1-\eta(y))\mathbb{1}_{x_i=x}-(1-\eta(x))\mathbb{1}_{x_i=y}][f^{x,y}-f]\Big]\right|$$
$$\le E^\mu\Big[\sum_{i\in k(t)}|\lambda_i|\sum_x\int_{x_i=x}\sum_y(1-\eta(y))\pi(y-x)|y-x|\big|\sqrt{f^{x,y}}-\sqrt{f}\big|\big|\sqrt{f^{x,y}}+\sqrt{f}\big|\Big]$$

$$\le E^{\mu}\Big[\sum_{i\in k(t)}|\lambda_i|\sum_x\int_{x_i=x}\Big[\sum_y \pi(y-x)(1-\eta(y))\big|\sqrt{f^{x,y}}-\sqrt{f}\big|^2\Big]^{\frac12}$$

$$\times\Big[\sum_y \pi(y-x)(y-x)^2\big(\sqrt{f^{x,y}}+\sqrt{f}\big)^2\Big]^{\frac12}\Big]$$

$$\le C\Big[\sum_{i\in k(t)}\lambda_i^2\Big]^{\frac12}\Big[\sum_{i\in k(t)}\pi(z)(1-\eta(x_i+z))\big(\sqrt{f^{x_i,x_i+z}}-\sqrt{f}\big)^2\Big]^{\frac12}$$

$$\le C\Big[\sum_{i\in k(t)}\lambda_i^2\Big]^{\frac12}\big[\mathcal{D}\big(\sqrt{f}\big)\big]^{\frac12}.$$

It follows that

$$\lambda(V(t,\mathbf{x}))=\sup_{\substack{f\ge0,\\ \|f\|_1=1}}\Big[E^{\mu}[V(t,\mathbf{x})f]-\frac12\mathcal{D}\big(\sqrt{f}\big)\Big]\le C_6\sum_{i\in k(t)}\lambda_i^2,$$

and, by Fubini's theorem,

$$\int_0^T\lambda(V(t,\mathbf{x}))dt\le C_6\sum_i\lambda_i^2[t_i-s_i].\qquad\square$$

THEOREM 7.4. *The distribution $\mathcal{R}_N$ of the empirical distribution satisfies the exponential tightness estimate as probability measures on the space of measures $D[[0,T];\mathbb{T}^d]$. In particular, given any L and ϵ, there are compact sets $K_{L,\epsilon}$ in $D[[0,T];\mathbb{T}^d]$ such that*

$$\frac{1}{N^d}\log\mathcal{R}_N\big[\gamma:\gamma\big[K^{c}_{L,\epsilon}\big]\ge\epsilon\big]\le -L.$$

PROOF. We can choose B large enough so that

$$\log P_N\Big[\sum_i S_T(x_i(\cdot))\ge\epsilon BN\Big]\le -LN^d,$$

and it follows that the empiricals

$$\frac{1}{N^d}\sum\delta_{\xi_i(\cdot)}$$

satisfy exponential tightness estimates. We now have to worry about the martingales. We can decompose the martingales into ones that correspond to specific jump sizes. Let $c_N(i,z,t)$ be the number of jumps of size $\frac{z}{N}$ that the i^{th} particle had up to time t. Then,

$$c_N(i,z,t)=N^2\int_0^t\pi(z)(1-\eta(x_i(s)+z)ds+NM_N(i,z,t)$$

where $M_N(i,z,t)$ are mutually orthogonal martingales

$$M_N(i,t)=\sum_z zM_N(i,z,t).$$

Given ϵ and δ, we say that the martingale $M_N(i,z,\cdot)$ "misbehaves" if

$$\sup_{\substack{0\le s,t\le T,\\ |s-t|\le\delta}}|M_N(i,z,t)-M_N(i,z,s)|\ge\epsilon.$$

Let H be the number of misbehaving martingales. We need an estimate of the type

$$\limsup_{\delta\to0}\limsup_{N\to\infty}\frac{1}{N^d}\log P_N[H\ge\epsilon N^d]=-\infty$$

for every $\epsilon > 0$. If the martingales were independent instead of just being orthogonal, this would only require that

$$\limsup_{\delta\to 0}\limsup_{N\to\infty}\sup_{i,z} P\Big[\sup_{\substack{0\le s,t\le T,\\ |s-t|\le\delta}} |M_N(i,z,t)-M_N(i,z,s)|\ge\epsilon\Big]=0$$

for every $\epsilon > 0$. If we toss N coins independently with probability of heads $P(H)\le\eta$ for any one of them, then the probability of getting more than $k = N\epsilon$ heads is bounded by

$$\beta_{N,\eta,\epsilon}=\sum_{r\ge k}\binom{N}{r}\eta^r(1-\eta)^{N-r},$$

and the large-deviation estimate for the binomial implies that for any $\epsilon > 0$

$$\limsup_{\eta\to 0}\limsup_{N\to\infty}\frac{1}{N}\log\beta_{N,\eta,\epsilon}=-\infty.$$

We want to compare two models. P_N is the distribution of N independent, compensated Poisson processes with jump size $\frac{z}{N}N^2$ and intensity c. H is the number of processes that misbehave during $[0,T]$. The second model Q_N corresponds to N mutually orthogonal martingales $y_i(t)$, again made of jumps of size $\frac{z}{N}$ and their compensators. They need not be independent and the rates $\lambda_N(i,t,\omega)$ are at most cN^2. We claim that $Q_N[H\ge k]\le P_N[H\ge k]$ for every k. This would solve our problem. We need exponential estimates on $Q_N[H\ge N\epsilon]$ and we have it for P_N. We will proceed with proving the claim.

Let $h(t)$ be a nonincreasing function on $[0,\infty)$ with $0\le h\le 1$ and $h(\infty)=0$. For the compensated Poisson process starting from a at time t, let $q(t,a)=E_{t,a}[h(\tau)]$ where τ is the exit time from $(-m,m)$ of the space-time process $(t,a(t))$ with generator

$$\Omega f=\frac{\partial f(t,a)}{\partial t}+cN^2\Big[f\Big(t,a+\frac{z}{N}\Big)-f(t,a)-\frac{z}{N}\frac{\partial f(t,a)}{\partial a}\Big].$$

Then, q is the solution of the equation

$$\begin{aligned} q_t+\Omega q&=0 &&\text{for } |a|<m,\ t\ge 0,\\ q(t,-m)&=h(t) &&\text{for } t\ge 0,\\ q(t,a)&=h(t) &&\text{for } a\in[m,m+1),\ t\ge 0.\end{aligned}$$

It satisfies $q_t\le 0$ and $\Omega q\ge 0$. Let

$$u_k^N(t,a_1,\dots,a_N)=\sum_{A\subset\{1,2,\dots,N\}}\prod_{i\in A}q(a_i,t)\prod_{i\notin A}(1-q(a_i,t))$$

where the summation is over all subsets with cardinality at least k. It is easy to check that

$$\frac{du_k^N}{dt}+\sum_i\Omega_i u_k^N=0.$$

u_k^N is a multilinear expression in $q(t,a_i)$, and the coefficient of $q(t,a_i)$ is $u_{k-1}^{N-1}(t,\dots,a_{i-1},a_{i+1},\dots)$ and $\Omega_i u_k^N\ge 0$ for every i. Suppose we now have N, not necessarily independent processes $y_i(t)$ such that

$$f(t,y_i(t))-f(0,y_i(0))-\int_0^t c_N(i,s,\omega)\Big[f\Big(s,y_i(s)+\frac{z}{N}\Big)-f(s,y_i(s))-\frac{z}{N}f_a(s,y_i(s))\Big]ds$$

are orthogonal martingales. Moreover, $c_N(i,s,\omega) \le cN^2$ makes

$$f(t, y_i(t)) - f(0, y_i(0)) - \int_0^t cN^2\Big[f\Big(s, y_i(s) + \frac{z}{N}\Big) - f(s, y_i(s)) - \frac{z}{N} f_a(s, y_i(s)\Big]ds$$

into supermartingales. If we take $h(t) = \mathbb{1}_{[0,T]}(t)$, which is still nonincreasing, the expressions $u_k^N(t, y_1(t), \dots, y_N(t))$ are now supermartingales under Q_N. To see this, we saw that u_k^N is linear in each variable with nonnegative coefficients. If we compute modulo martingales using orthogonality,

$$d\big[u_k^N(t, y_1(t), \dots, y_N(t)\big] = \sum_i \big[-\Omega_i u_k^N dt + g^i c_N(i,t,\omega)\Omega_i q(t, a_i)\big].$$

Since $\Omega_i u_k^N = cN^2 g^i \Omega_i q(t,a_i)$ where $g_i = u_{k-1}^{N-1}$, it follows that $u_k^N(t, y_1(t), \dots, y_N(t))$ is a supermartingale under Q_N. If we stop the processes that exit from $(-m, m)$ and continue with the rest, the event in question counts the numbers that are still inside and the ones that got out.

$$Q_N(H) = E^{Q_N}\big[u^N(T, y_1(T), \dots, y_N(T))\big] \le u^N(0, \dots, 0) = P_N[H].$$

$y_i(t) = 1$ for the paths that got out and 0 for those that did not. □

7.3. Approximations

There are three versions of our basic simple exclusion process with K colors. One on all of $\mathbb{Z}^d$, one in the periodic box $\mathbf{Z}_N^d$, which is our main focus, and finally, one in a finite box $\mathbb{B}_q^d$ with no jumps allowed to the exterior. Their generators are

$$(\mathcal{A}f)(\zeta) = \sum_{x,y\in\mathbb{Z}^d} \eta(x)(1-\eta(y))\pi(y-x)[f(\zeta^{x,y}) - f(\zeta)]$$

and

$$(\mathcal{A}_N f)(\zeta) = \sum_{x,y\in\mathbf{Z}_N^d} \eta(x)(1-\eta(y))\pi(y-x)[f(\zeta^{x,y}) - f(\zeta)].$$

In a finite box $\mathbb{B}_q^d = [-q,q]^d$ of size $(2q+1)^d$, with jumps to the exterior excluded, the generator will be

$$(\mathcal{A}_q^o f)(\zeta) = \sum_{x,y\in\mathbb{B}_q^d} \eta(x)(1-\eta(y))\pi(y-x)[f(\zeta^{x,y}) - f(\zeta)].$$

In addition to the two Dirichlet forms (7.14) and (7.15), we have on $\mathbb{B}_q^d$ the form

$$\mathcal{D}_q^o(u) = \frac{1}{4} E^{\mu_{q,\mathbf{k}}}\Big[\sum_{x,y\in\mathbb{B}_q^d} a_{x,y}(\zeta)\pi(y-x)[u(\zeta^{x,y}) - u(\zeta)]^2\Big].$$

We have three types of local functions, all having the common property that they have mean zero under every $\mu_{\bar\rho}$ with $\sum_{i=1}^K \rho_i < 1$. The first type consists of functions $f = \mathcal{A}u$ for some local function u. As u varies, we get a large family $\mathcal{N}$ of local functions $\{f\}$. The second family of "currents" consists of Kd functions $\{\mathbf{f}_r^i\}$ given by

$$\mathbf{f}_r^i(\zeta) = \frac{1}{2}\sum_z \pi(z)\langle z, e_r\rangle[\zeta_i(0)(1-\eta(z)) - \zeta_i(z)(1-\eta(0))].$$

Both families have the property that their expectation is 0 under every invariant distribution, in every sufficiently large box. If u is defined in a box, then $\mathcal{A}u$ has zero mean with respect to any invariant distribution in any box $\mathbb{B}_q^d$ provided $\mathcal{A}_q^0 u = \mathcal{A}u$. Finally, we have the Kd microscopic "density gradients"

$$\mathbf{d}_r^i = \{\zeta_i(e_r) - \zeta_i(0)\}. \tag{7.17}$$

The first type will be "negligible." The goal is to express the "currents" as a linear combination of the third type, "density gradients," modulo the first type that are "negligible." The density gradients are a bit more difficult to handle because their expectation is not 0 if the density is 1. There will be problems when ρ is close to 1. But the basic object we want to approximate is 0 if the density is 1; so there is a natural decay when ρ is close to 1.

We consider a function of the form $f = \mathcal{A}u = \mathcal{A}_N u$ where u is a local function. Let $U(\zeta) = \sum_{x \in \mathbf{Z}_N^d} u(\tau_x \zeta)$ and $F(\zeta) = \sum_{x \in \mathbf{Z}_N^d} f(\tau_x \zeta)$. Then, $\mathcal{A}_N U = F$. In the speeded-up time scale with generator $N^2\mathcal{A}$,

$$\frac{1}{N^d}\int_0^t [N^2 F](\zeta(s))ds = \frac{1}{N^d}[U(\zeta(t)) - U(\zeta(0))] - M_N(t).$$

We can express

$$\frac{1}{N^d}\int_0^t [NF](\zeta(s))ds = q_N^1(t) + q_N^2(t)$$

where $|q_N^1(t)| = (1/N^{d+1})|U(\zeta(t)) - U(\zeta(0))| \le C/N$, and $q_N^2 = (1/N)M_N(t)$ is a martingale with $O(N^d)$ possible jumps of size $N^{-(d+1)}$ with rate $O(N^2)$ making its quadratic variation $O(N^{-d})$. For $f \in \mathcal{N}$, this makes

$$\frac{N}{N^d}\sup_{0\le t\le T}\left|\int_0^t \Big[\sum_{x\in\mathbf{Z}_N^d} f(\tau_x\zeta(s))\Big]ds\right|. \tag{7.18}$$

"negligible."

For any $1 \le r \le d$, $1 \le i \le K$, and smooth function A on $\mathbb{T}^d$, we need to be able to replace

$$\frac{1}{N^d}\int_0^t \Big[\sum_{x\in\mathbf{Z}_N^d} A\Big(\frac{x}{N}\Big)\big(N\mathbf{f}_r^i\big)(\tau_x\zeta(s))\Big]ds$$

by

$$-\frac{1}{2N^d}\int_0^t \Big[\sum_{x\in\mathbf{Z}_N^d} A\Big(\frac{x}{N}\Big)\sum_{r',j} c_{r,r'}^{i,j}(\bar{\zeta}_{x+N\epsilon'}(s))\frac{1}{2\epsilon}\big[\bar{\zeta}_{j,x+N\epsilon e_{r'},N\epsilon'}(s) - \bar{\zeta}_{j,x-N\epsilon e_{r'},N\epsilon'}(s)\big]\Big]ds$$

with an error that becomes negligible as $N \to \infty$, followed by $\epsilon, \epsilon' \to 0$. That would lead to

$$-\frac{1}{2}\int_0^t\int_{\mathbb{T}^d} A(u)\sum_{r',j} c_{r,r'}^{i,j}(\bar{\rho}_{\epsilon'}(s,u))\Big[\frac{1}{2\epsilon}\big[\bar{\rho}_{j,\epsilon'}(s,u+\epsilon e_{r'}) - \bar{\rho}_{j,\epsilon'}(s,u-\epsilon e_{r'})\big]\Big]ds\,du$$

where

$$\bar{\rho}_{i,\epsilon'}(s,u) = \frac{1}{(2\epsilon')^d}\int_{|u'-u|\le\epsilon'} \rho_i(s,u')du'$$

and

$$\bar{\zeta}_{r,x,N\epsilon'} = \frac{1}{(2N\epsilon')^d}\sum_{|y-x|\le N\epsilon'} \zeta_r(x).$$

As $\epsilon', \epsilon \to 0$, this becomes

$$-\frac{1}{2}\int_0^t \int_{\mathbb{T}^d} A(u) \sum_{r',j} c^{i,j}_{r,r'}(\tilde{\rho}(s,u)) \frac{\partial \rho_j(s,u)}{\partial u_{r'}} ds\, du.$$

If we consider a family of local functions $v(\tilde{\rho}, \zeta)$ and $f(\tilde{\rho}, \cdot) = \mathcal{A}v(\tilde{\rho}, \cdot)$, depending smoothly on $\tilde{\rho}$ and depending on ζ in a finite box,

$$\frac{1}{N^d}\int_0^t \sum_{x \in \mathbf{Z}_N^d} (Nf)(\tau_x \bar{\zeta}_{N\epsilon'}(t), \tau_x \zeta(t)) dt =$$
$$\frac{1}{N^{d+1}} \sum_{x \in \mathbf{Z}_N^d} \big[v(\tau_x \bar{\zeta}_{N\epsilon'}(t), \tau_x \zeta(t)) - v(\tau_x \bar{\zeta}_{N\epsilon'}(0), \tau_x \zeta(0)) \big] + M_N(t) + o(1)$$

is negligible. But the process under consideration is not necessarily in equilibrium, and we may want to use these estimates to establish large deviations as well. So, we need to show that these error probabilities are superexponentially small in equilibrium.

In other words, we need to show that for any smooth function A on $\mathbb{T}^d$,

$$\inf_{f(\cdot,\cdot,\cdot)} \limsup_{\epsilon,\epsilon' \to 0} \limsup_{N \to \infty} \frac{1}{N^d} \log E^{N,\mathbf{k}(N)}\bigg[\exp\bigg[N \int_0^t \Theta_N(f, \zeta(s)) ds \bigg]\bigg] = 0 \tag{7.19}$$

where for each i, r

$$\Theta_N(f, \zeta) = \sum_{x \in \mathbf{Z}_N^d} A\Big(\frac{x}{N}\Big) h(\tau_x \zeta),$$

with

$$h(\tau_x \zeta) = \big[\mathbf{f}^i_r(\tau_x \zeta) - f\big(\tau_x \bar{\zeta}_{N\epsilon'}, \tau_x \zeta\big) + T_\epsilon\big(\tau_x \bar{\zeta}_{N\epsilon'}, \zeta\big) \big],$$

and

$$T_\epsilon\big(\bar{\zeta}_{N\epsilon'}, \zeta\big) = \frac{1}{2} \sum_{r',j} c^{i,j}_{r,r'}\big(\bar{\zeta}_{N\epsilon'}\big) \frac{1}{2\epsilon}\big[\tau_{N\epsilon e_{r'}} \bar{\zeta}_{j,N\epsilon'} - \tau_{-N\epsilon e_{r'}} \bar{\zeta}_{j,N\epsilon'} \big].$$

By the use of the variational formula and Feynman-Kac representation,

$$\begin{aligned}
&\frac{1}{N^d} \log E^{N,\mathbf{k}(N)}\bigg[\exp\bigg[N \int_0^t \Theta_N(f, \zeta(s)) ds \bigg]\bigg] \\
&\quad \le t N^{-d} \sup_G \big[N E^{\mu_{N,\mathbf{k}(N)}}\big[\Theta_N(f,\zeta) G^2\big] - N^2 \mathcal{D}_N(G) \big] \\
&\quad = t N^{2-d} \sup_G \big[N^{-1} E^{\mu_{N,\mathbf{k}(N)}}\big[\Theta_N(f,\zeta) G^2\big] - \mathcal{D}_N(G) \big] \\
&\quad \simeq t N^{-d} \langle \Theta_N(f,\zeta), \Theta_N(f,\zeta) \rangle_{CLT} \\
&\quad = t N^{-d} \sup_G \bigg[E^{\mu_{N,\mathbf{k}(N)}}\big[G \Theta_N(f,\zeta) \big] - \frac{1}{8} \mathcal{D}_N(G) \bigg].
\end{aligned}$$

If we have an expression of the form

$$E^{\mu_{N,\mathbf{k}(N)}}\Big[G\Big[\sum_{x \in \mathbf{Z}_N^d} H(\tau_x \zeta) \Big] - D_N(G) \Big]$$

to estimate and if $H(\zeta)$ is of the form $-[F(\zeta^{a,a+z}) - F(\zeta)]$, we can do a summation by parts and rewrite the above expression as

$$E^{\mu_{N,\mathbf{k}(N)}}\Big[\frac{1}{2}\sum_{x\in\mathbf{Z}_N^d}\big[F(\tau_x\zeta^{x+a,x+a+z}) - F(\tau_x\zeta)\big]\big[G(\zeta^{x+a,x+a+z}) - G(\tau_x\zeta)\big]\Big] - D_N(G).$$

Both sides can be localized by breaking them up into sums over $\mathbb{B}_q^d$. One can replace densities over small macroscopic blocks by densities over large microscopic blocks. The problems with $\rho \simeq 1$ has to be handled. The Dirichlet form $\mathcal{D}_N$ can be thought of as $(2q+1)^{-d}\sum_{x\in\mathbf{Z}_N^d}\mathcal{D}^0_{x,\mathbf{k}}$. The problem can now be reduced to estimating the following quantities after trimming the edges by leaving a border of size v so that all the functions depend only on the configuration in $\mathbb{B}_q^d$:

$$\Lambda(q,f) = \sup_G\Big[E^{\mu_{k,\mathbf{k}}}\Big[G\sum_{x\in\mathbb{B}_{q-v}}\Big[\mathbf{f}_r^i(\tau_x\zeta) - \sum_{r',j}c_{i,j}^{i,j}[\zeta_j(x+e_{r'}) - \zeta_j(x)] - f(\tau_x\zeta)\Big]\Big] - \mathcal{D}^0_{q,\mathbf{k}}(G)\Big],$$

with $f = \mathcal{A}u$ where u varies over local functions. We need to show that, with the proper choice of constants $c_{r,r'}^{i,j}(\tilde{\rho})$,

$$\inf_u \limsup_{\substack{q\to\infty,\\ (2q+1)^{-d}k_i\to\rho_i}} (2q+1)^{-d}\Lambda(q,f) = 0. \tag{7.20}$$

The plan is to first establish the validity of the above approximation. This is at the level of the central limit theorem in equilibrium. Then use the idea of localization to make smooth choices of f and $c_{r,r'}^{i,j}$ that depend on $\bar{\zeta}$ over blocks, and use it to prove (7.20). The central limit theorem requires us to use $\bar{\zeta}$ over large microscopic blocks, whereas (7.20) will need small macroscopic blocks. We will need analogues of Lemma 5.8 and Lemma 5.9, since we want to do large deviations as well.

The plan is to establish (7.20) first.

7.4. Calculating Variances

We are given two local functions g_1, g_2, depending on configurations in a box $\mathbb{B}_q^d$. They have mean 0 under every invariant distribution μ_q under $\mathcal{A}_q^o$. We try to define an inner product $[g_1, g_2]_{\tilde{\rho}}$ in two steps.

$$\langle g_1, g_2\rangle_{q,\mathbf{k}} = \lim_{t\to\infty}\frac{1}{t}E^{\mu_{q,\mathbf{k}}}\Big[\Big(\int_0^t\sum_{|x|\le q-v}g_1(\tau_x\zeta(s))ds\Big)\Big(\int_0^t\sum_{|x|\le q-v}g_2(\tau_x\zeta(s))ds\Big)\Big].$$

We note that g_1 and g_2 continue to have mean 0 under every invariant measure $\mu_{q',\mathbf{k}'}$ on $\mathbb{B}_{q'}$ for $q' \ge q$.

$$[g_1, g_2]_{\tilde{\rho}} = \lim_{\substack{q\to\infty,\\ (2q+1)^{-d}k_i\to\rho_i}} (2q+1)^{-d}\langle g_1, g_2\rangle_{q,\mathbf{k}}. \tag{7.21}$$

We will construct for each $\tilde{\rho}$ with $\rho = \sum_{i=1}^K \rho_i < 1$, a Hilbert space $\mathcal{H} = \mathcal{H}_{\tilde{\rho}}$ and a linear map $g \to \hat{g} = \sigma(g)$ that imbeds linear combinations $g = \mathcal{A}u + \sum_{i,r}a_r^i\mathbf{f}_r^i + \sum_{i,r}b_r^i\mathbf{d}_r^i$ with inner product $[g_1, g_2]_{\tilde{\rho}}$ isometrically into $\mathcal{H}_{\tilde{\rho}}$ with inner product $\langle\cdot,\cdot\rangle$. It will turn out that $\sigma(\mathcal{A}u) \perp \sigma(\mathbf{d}_r^i)$ and $\mathcal{H}_{\tilde{\rho}}$ is spanned by them. Then the approximation is an easy consequence.

Let $\mathcal{K}$ be the space of functions $\hat{g} = \{g(z,\zeta)\}$ defined on $\{z : \pi(z) > 0\}$ with values in $L_2[\mu_{\tilde{\rho}}]$. We define an inner product

$$\langle \hat{g}_1, \hat{g}_2 \rangle = \frac{1}{2} E^{\mu_{\tilde{\rho}}}\Big[\sum_z a_{0,z}(\zeta)\pi(z)\hat{g}_1(z,\zeta)\hat{g}_2(z,\zeta)\Big].$$

Let $\mathcal{N}$ be the space of functions $f = \mathcal{A}u = \mathcal{A}_q^0 u$ for some local u depending on $\mathbb{B}_q^d$. Since $\mathcal{A}$ is linear and translation invariant,

$$\mathcal{A}_q^0\Big(\sum_{|x|\le q-v} u(\tau_x\zeta)\Big) = \sum_{|x|\le q-v} f(\tau_x\zeta)$$

and

$$\langle f, f\rangle_{q,\mathbf{k}} = 2\mathcal{D}_N\Big(\sum_{|x|\le q-v} u(\tau_x\zeta)\Big).$$

It is now easy to calculate the limit (7.21) as $q \to \infty$. If $\mathcal{A}u_i = f_i$ for $i = 1, 2$ and $(2q+1)^{-d}k_i \to \rho_i$,

$$[f_1, f_2]_{\tilde{\rho}} = \frac{1}{2} E^{\mu_{\tilde{\rho}}}\Big[\sum_z a_{0,z}(\zeta)\pi(z)\hat{f}_1(z,\zeta)\hat{f}_2(z,\zeta)\Big] = \langle \hat{f}_1, \hat{f}_2\rangle$$

where, for $i = 1, 2$,

$$\hat{f}_i(z,\zeta) = \mathbf{U}_i(\zeta^{0,z}) - \mathbf{U}_i(\zeta) \quad \text{and} \quad \mathbf{U}_i(\zeta) = \sum_{x\in\mathbb{Z}^d} u_i(\tau_x\zeta).$$

Although $\mathbf{U}_i$ are not well-defined, $\hat{f}_i^{x,y} = \mathbf{U}_i(\zeta^{x,y}) - \mathbf{U}_i(\zeta)$ are well-defined and satisfy linear identities. $\hat{f}_i^{x,y}(\zeta)$ are covariant, i.e., $\hat{f}_i^{x+z,y+z}(\zeta) = \hat{f}_i^{x,y}(\tau_z\zeta)$. If $\{\sigma_j\}$ are permutations of the form $x \leftrightarrow x + e_j$ for some x and j, and $\sigma_1\sigma_2\cdots\sigma_k = \mathrm{Id}$, then with $\sigma_j = x_j \leftrightarrow x_j + e_{\alpha(j)}$ for some $e = e_{\alpha(j)}$

$$0 = \sum_{j=1}^k \mathbf{U}_i(\sigma_j\sigma_{j-1}\cdots\sigma_1\zeta) - \mathbf{U}_i(\sigma_{j-1}\cdots\sigma_1\zeta) = \sum_{j=1}^k \hat{f}_i^{x_j, x_j+e_{\alpha(j)}}(\sigma_{j-1}\cdots\sigma_1\zeta).$$

The Hilbert space $\mathcal{H} \subset \mathcal{K}$ consists of all such maps. With $h_{0,z}(\zeta) = h(0,\zeta)$ and $h_{x,y}(\zeta) = h_{0,y-x}(\tau_x\zeta)$, they satisfy these linear identities. The inner product is given by

$$\langle h_1, h_2\rangle_{\tilde{\rho}} = \frac{1}{2} E^{\mu_{\tilde{\rho}}}\Big[\sum_z a_{0,z}(\zeta)\pi(z)h_1(z,\zeta)h_2(z,\zeta)\Big]$$

and $\mathcal{H}_0 \subset \mathcal{H}$ is the closure of the span of $\hat{f}$ as f ranges over $\mathcal{N}$. We need to consider two families of functions that are not in $\mathcal{N}$. $\{\mathbf{f}_r^i\}$ and $\{\zeta_i(e_r) - \zeta_i(0)\}$ with $1 \le i \le K$ and $1 \le r \le d$. We will show that they can be imbedded in $\mathcal{H}$ as well. Imbedding $\mathbf{f}_r^i$ is relatively easy. We can take

$$V_r^i(\zeta) = \sum_{x\in\mathbb{B}_k^d} \langle x, e_r\rangle \zeta_i(x).$$

A calculation of $\mathcal{A}_q^o V_r^i$ yields

$$\big(\mathcal{A}_q^o V_r^i\big)(\zeta) = \sum_{x,y\in\mathbb{Z}_q} \zeta_i(x)(1-\eta(y))\pi(y-x)\langle y-x, e_r\rangle \simeq \sum_x \mathbf{f}_r^i(\tau_x\zeta),$$

the error coming entirely from boundary terms. The error can be controlled and for large q becomes relatively negligible. Therefore,

$$\sigma\big(\mathbf{f}_r^i\big)(z,\zeta) = V_r^i(\zeta^{0,z}) - V_r^i(\zeta) = \langle z, e_r\rangle[\zeta_i(0) - \zeta_i(z)].$$

While we defer saying anything about $\sigma(\mathbf{d}_r^i)$ till later, we can, however, compute its inner product in $\mathcal{H}$ with objects in $\mathcal{H}_0$ and $\sigma(\mathbf{f}_r^i)$:

$$\langle f, \zeta_i(e_r) - \zeta_i(0)\rangle_{q,\mathbf{k}} = -2E^{\mu_{q,\mathbf{k}}}\Big[\Big[\sum_{|x|\le q-v} u(\tau_x\zeta)\Big]\Big[\sum_{|x|\le q-1}(\zeta_i(x+e_r) - \zeta_i(x))\Big]\Big] = O(q^{d-1}).$$

The summation $\sum_{|x|\le q-1}(\zeta_r(x+e_i) - \zeta_r(x))$ telescopes, $\mu_{q,\mathbf{k}}$ is almost a product measure, and u is local. Therefore, only the boundary contributes. This proves that $\sigma(\mathbf{d}_r^i) \perp \sigma(f)$ for all $f = \mathcal{A}u$, i.e., $\sigma(\mathbf{d}_r^i) \perp \mathcal{H}_0$.

We next compute the inner product $\langle\sigma(\mathbf{f}_r^i), \sigma(\mathbf{d}_{r'}^j)\rangle_{\tilde\rho}$. We can compute it as

$$\begin{aligned}
&\big\langle\sigma\big(\mathbf{f}_r^i\big), \sigma\big(\mathbf{d}_{r'}^j\big)\big\rangle_{\tilde\rho}\\
&\quad = -2 \lim_{\substack{q\to\infty,\\ (2q+1)^{-d}k_i\to\rho_i,\,\forall i}} \frac{1}{(2q+1)^d} E^{\mu_{q,\mathbf{k}}}\Big[\Big[\sum_x \langle x, e_r\rangle\zeta_i(x)\Big]\\
&\qquad\qquad \cdot\Big[\sum_{x_r=q,|x|\le q}\zeta_j(x) - \sum_{x_r=-q,|x|\le q}\zeta_j(x)\Big]\Big]\\
&\quad = -2[\delta_{i,j}\rho_i - \rho_i\rho_j].
\end{aligned}$$

We next do a calculation. With $f_{0,z} = U(\zeta^{0,z}) - U(\zeta)$ and $U(\zeta) = \sum_{x\in\mathbb{Z}^d} u(\tau_x\zeta)$ and denoting $\mathcal{U}$ the set of local functions u, we will show that

$$\begin{aligned}
\inf_{u\in\mathcal{U}} \frac{1}{2} E^{\mu_{\tilde\rho}}\Big[\sum_z a_{0,z}(\zeta)\pi(z)\Big[\langle z, w\rangle\Big(\sum_i a_i(\zeta_i(0) - \zeta_i(z)\Big) - f_{0,z}(\zeta)\Big]^2\Big] = \\
\langle w, S(\rho)w\rangle\langle a, R_1 a\rangle + \langle w, Dw\rangle\langle a, R_2 a\rangle
\end{aligned}$$

where D is the covariance matrix

$$\langle w, Dw\rangle = \sum \pi(z)\langle z, w\rangle^2,$$

$$R_1 = \begin{pmatrix}
\rho_1(1-\frac{\rho_1}{\rho}) & -\frac{\rho_1\rho_2}{\rho} & \cdots & -\frac{\rho_1\rho_K}{\rho}\\
-\frac{\rho_1\rho_2}{\rho} & \rho_2(1-\frac{\rho_2}{\rho}) & \cdots & -\frac{\rho_2\rho_K}{\rho}\\
\vdots & \vdots & \ddots & \vdots\\
-\frac{\rho_1\rho_K}{\rho} & -\frac{\rho_2\rho_K}{\rho} & \cdots & \rho_K(1-\frac{\rho_K}{\rho})
\end{pmatrix},$$

and

$$R_2 = \frac{1-\rho}{\rho}\begin{pmatrix}
\rho_1^2 & \rho_1\rho_2 & \cdots & \rho_1\rho_k\\
\rho_1\rho_2 & \rho_2^2 & \cdots & \rho_2\rho_K\\
\vdots & \vdots & \ddots & \vdots\\
\rho_1\rho_K & \rho_2\rho_K & \cdots & \rho_K^2
\end{pmatrix}.$$

Because the generator leaves invariant the class of expressions of the form $\sum_{x\in\mathbf{Z}_N^d}\ell(\zeta(x))u(\tau_x\eta)$ where $\ell(\zeta) = \sum_i a_i\zeta_i$ is a linear expression and $u(\eta)$ is a local function of η, we can restrict the class of functions $u(\zeta)$ to functions of the form $\ell(\zeta(0))u(\eta)$.

We need to compute

$$\inf_{u\in\mathcal{U}}\frac{1}{2}E^{\mu_{\tilde\rho}}\Big[\sum_z\big[\eta(0)(1-\eta(z))+\eta(z)(1-\eta(0))\big]\pi(z)[\langle z,w\rangle\big(\ell(\zeta(0))-\ell(\zeta(z))\big)-f_{0,z}(\zeta)]^2\Big].$$

We decompose $\sum a_i\zeta_i$ as

$$\Big[(1/\rho)\sum_{i=1}^K a_i\rho_i\Big]\eta+\sum(a_i-\bar a)\zeta_i=\bar a\eta+\hat\ell(\zeta)$$

where $\bar a = (1/\rho)\sum_{i=1}^K a_i\rho_i$. Moreover, $\hat\ell(\zeta) = \sum_{i=1}^K\hat a_i\zeta_i$ and $\sum_{i=1}^K\rho_i\hat a_i = 0$. The conditional mean of $E[\hat\ell(\zeta(x))|\eta]$ is easily calculated to be 0, and the conditional variance of $\ell(\zeta(x))$ is seen to be $[(1/\rho)\sum_{i=1}^K a_i^2\rho_i-\bar a^2]\eta(x)$.

We note that, as we saw earlier,

$$E^{\mu_{\tilde\rho}}[a_{0.z}(\eta(0)-\eta(z))f_{0,z}(\zeta)] = E^{\mu_{\tilde\rho}}[(\eta(0)-\eta(z))f_{0,z}(\zeta)] = E^{\mu_{\tilde\rho}}[(\eta(0)-\eta(z))f_{0,z}(\zeta)] = 0.$$

Therefore, we now have

$$\begin{aligned}\inf_{u\in\mathcal{U}}\frac{1}{2}E^{\mu_{\tilde\rho}}\Big[&\sum_z\big[\eta(0)(1-\eta(z))+\eta(z)(1-\eta(0))\big]\pi(z)[\langle z,w\rangle\big(\hat\ell(\zeta(0))-\hat\ell(\zeta(z))\big)-f_{0,z}(\zeta)]^2\Big]\\&+\frac{1}{2}\bar a^2E^{\mu_{\tilde\rho}}\Big[\sum_z\big[\eta(0)(1-\eta(z))+\eta(z)(1-\eta(0))\big]\pi(z)\big[\langle z,w\rangle(\eta(0)-\eta(z))\big]^2\Big].\end{aligned}$$

The second term is easily computed to be

$$\bar a^2\langle Dw,w\rangle\rho(1-\rho).$$

We examine now the first term:

$$\begin{aligned}f_{0,z}(\zeta)&=\Big[\sum_x\hat\ell(\zeta(x))u(\tau_x\eta)\Big]^{0,z}-\Big[\sum_x\hat\ell(\zeta(x))u(\tau_x\eta)\Big]\\&=\sum_{x\neq0,z}\hat\ell(\zeta(x))\big[u\big((\tau_x\eta)^{0,z}\big)-u(\tau_x\eta)\big]+\hat\ell(\zeta(z))u((\eta^{0,z}))-\hat\ell(\zeta(0))u(\eta)\\&\qquad+\hat\ell(\zeta(0))u\big((\tau_z\eta)^{0,z}\big)-\hat\ell(\zeta(z))u(\tau_z\eta).\end{aligned}$$

We remark that given η, $\zeta(x)$ are mutually independent and

$$E[\hat\ell(\zeta)|\eta]=0,\quad E[\hat\ell(\zeta)^2|\eta]=\sigma^2=\frac{1}{\rho}\sum_{i=1}^K\rho_ia_i^2-\bar a^2.$$

We can calculate the conditional expectation given η explicitly to obtain the variational formula

$$\begin{aligned}\sigma^2\inf_{u\in\mathcal{U}}\frac{1}{2}E^{\mu_\rho}\Big[&\sum_z\pi(z)\eta(0)(1-\eta(z))\big[\langle z,w\rangle-u\big((\tau_z\eta)^{0,z}\big)+u(\eta)\big]^2\\&+\sum_z\pi(z)\eta(z)(1-\eta(0))[\langle z,w\rangle-u(\eta^{0,z})+u(\tau_z\eta)]^2\\&+\sum_z\sum_{x\neq0,z}\pi(z)[\eta(0)(1-\eta(z))+\eta(z)(1-\eta(0))]\big[u\big((\tau_x\eta)^{0,z}\big)-u(\tau_x\eta)\big]^2\Big].\end{aligned}$$

Using the translation invariance of μ_ρ and the symmetry of $\pi(\,\cdot\,)$, this can be rewritten as

$$\sigma^2\rho \inf_{u\in\mathcal{U}} E^{\mu_\rho}\Big[\sum_z \pi(z)(1-\eta(z))\big[\langle z,w\rangle - u\big((\tau_z\eta)^{0,z}\big)+u(\eta)\big]^2 + \sum_{\substack{x,y\\ x\neq y}} \pi(y-x)[u(\eta^{x,y})-u(\eta)]^2\Big]$$

which by Theorem 6.4 equals $\sigma^2\rho S(\rho)$.

The last calculation is to show that $\sigma(\mathbf{f}_i^r)+\sum_{j,r'} c_{j,r'}^{i,r}\sigma(\mathbf{d}_j^{r'}) \in \mathcal{H}_0$ for a choice of $\{c_{j,r'}^{i,r}\}$ to be determined. If we denote the projections of $\sigma(\mathbf{f}_i^r)$ in the orthogonal complement of $\mathcal{H}_0$ by $\hat{f}_i^r$, then we know that

$$A_{j,r'}^{i,r}(\rho) = \frac{1}{2}\langle \hat{f}_i^r, \hat{f}_j^{r'}\rangle = S_{i,j}(\rho)\left(\rho_r\delta_{r,r'} - \frac{\rho_r\rho_{r'}}{\rho}\right) + D_{i,j}\rho_r\rho_{r'}\frac{1-\rho}{\rho}.$$

We also know

$$\frac{1}{2}\langle \hat{f}_i^r, \mathbf{d}_j^{r'}\rangle = \frac{1}{2}\langle\sigma(\mathbf{f}_i^r), \mathbf{d}_j^{r'}\rangle = -[\delta_{r,r'}\rho_r - \rho_r\rho_{r'}] = -\chi_{r,r'}^{-1}.$$

By elementary calculation,

$$C = A\chi$$

where

$$A = S\otimes R_1 + D\otimes R_2$$

and

$$\chi = \begin{pmatrix} \frac{1}{\rho_1}+\frac{1}{1-\rho} & \frac{1}{1-\rho} & \cdots & \frac{1}{1-\rho} \\ \frac{1}{1-\rho} & \frac{1}{\rho_2}+\frac{1}{1-\rho} & \cdots & \frac{1}{1-\rho} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{1-\rho} & \frac{1}{1-\rho} & \cdots & \frac{1}{\rho_K}+\frac{1}{1-\rho} \end{pmatrix}.$$

This is how the equation

$$\tilde{\rho}_t = \frac{1}{2}\nabla A\chi\nabla\tilde{\rho}$$

for the multicolor system $\tilde{\rho} = (\rho_1,\ldots,\rho_K)$ is established.

7.5. Proofs

We will not provide complete proofs here. Instead, we will indicate the main steps and provide references. The first step is to show that in any $\mu_{\tilde{\rho}}$, $\mathcal{H}$ is spanned by $\mathcal{H}_0$ and the Kd-dimensional subspace that is generated by $\sigma(d_r^i)$. Since we have already calculated the relevant inner products, it will then follow that $\sigma(\mathbf{f})+A\chi\sigma(\mathbf{d})\in\mathcal{H}_0$. Suppose we start with an $h\in\mathcal{H}$. We are trying to find a $U=\sum_x F(\tau_x\eta)$ so that

$$h(z,\zeta) = U(\zeta^{0,z}) - U(\zeta).$$

The first step is to integrate out all the variables outside a box so that we are computing

$$h^q(x,y,\zeta) = E\big[h(y-x,\tau_x\zeta)|\mathcal{F}_{\mathcal{D}_q}\big].$$

Since the measure involved is a product measure, this is plain averaging and $\{h^q(x,y)\}$ will satisfy all the linear relations except the covariance under translations. We broke the translation symmetry by conditioning over $\mathcal{D}_q$. We can now find a function U^q on the space of configurations ζ on $\mathcal{D}_q$ such that $U^q(\zeta^{x,y}) - U^q(\zeta) = h^q(x,y,\zeta)$. If there are enough empty sites, we can move any configuration to any other by making a finite number of allowed exchanges, and the result will be path independent. We have a free constant and we can choose it so that $E[U^q(\zeta)] = 0$. We can then get a translation invariant U by defining

$$U(\zeta) = \frac{1}{(2q+1)^d} \sum_x U^q(\tau_x \zeta).$$

This is almost what we want except for contributions near the border. They should be relatively negligible when q is large. But that requires knowledge about the size of U^q. That is, a Poincaré inequality or a spectral gap estimate is needed for functions linear in ζ with coefficients that are functions of η. Then it turns out that the boundary terms do not go away but remain bounded and contribute in the limit. The limit can be shown to be a linear combination of $\langle z, e_r\rangle[\zeta_i(e_r) - \zeta_i(0)]$.

Finally, we have to show that $\zeta_i(e_r) - \zeta_i(0)$ can be embedded in $\mathcal{H}$ where they have to be in the span of $\mathcal{H}_0$ and the Kd-dimensional subspace $\langle z, e_r\rangle[\zeta_i(e_r) - \zeta_i(0)]$. With this and analogues of Lemma 5.8 and Lemma 5.9, we can replace the currents **f** by appropriate linear combinations of $\zeta_i(e_r) - \zeta(0)$ to establish the limiting hydrodynamic equation in the weak form,

$$\tilde{\rho}_t = \frac{1}{2}\nabla A(\tilde{\rho})\chi(\tilde{\rho})\nabla\tilde{\rho}. \tag{7.22}$$

This system, although nonlinear, can be solved in two steps by reducing it to two linear equations. One can argue independently that a tagged particle in nonequilibrium will behave locally like one in equilibrium. So it will diffuse according to

$$\frac{1}{2}\nabla S(\rho(t,u))\nabla + b(t,u)\cdot\nabla$$

with some drift term to be determined. Collectively, the density will evolve as

$$\rho_t = \frac{1}{2}\nabla S(\rho(t,u))\nabla\rho - \nabla\cdot[b(t,u)\rho]. \tag{7.23}$$

But the density also evolves as

$$\tilde{\rho}_t = \frac{1}{2}\nabla D\nabla\tilde{\rho}. \tag{7.24}$$

Therefore we must have

$$D\nabla\rho = S(\rho)\nabla\rho - 2b\rho + m\rho$$

with a divergence free $m\rho$; i.e., $\nabla\cdot m\rho = 0$. It is most likely that $m = 0$, and so we expect

$$b(t,u) = [S(\rho(t,u)) - D]\frac{(\nabla\rho)(t,u)}{2\rho(t,u)}.$$

Solving (7.22) is carried out in two steps. Solve for the total density (7.24). Then, with the solution $\rho(t,x)$, solve for each ρ_i by the linear equation (7.23) with variable coefficients given by $\rho(t,u)$.

You can find the details in [**19, 21**]. It is useful to note that we can limit ourselves to the class of functions that depend linearly on ζ with coefficients that can be functions of η. This class is left invariant by the semigroup. We saw already how this made certain computations easier.

Analogues of Lemma 5.8 and Lemma 5.9 are not very difficult if $\rho < 1$. This will mean that there will be empty spaces far away and they can facilitate interchanges between particles of different colors that are far away. The proof is not all that different from proofs when there is a single color.

Another point to note is that the approximation estimates are strong enough to allow us to carry out large-deviation estimates. The reason is that although the large-deviation estimates require the calculation of exponential moments, the scaling makes the constant in front of the exponential small, bringing it within the scope of the central limit theorem. The upper bound of the dynamical part of the rate function is computed easily through the exponential martingales,

$$E^{P_N}\left[\exp\left[\sum_{i=1}^{K}\sum_{x\in\mathbf{Z}_N^d}\left[J_i\Big(T,\frac{x}{N}\Big)\zeta(T,x)-J_i\Big(0,\frac{x}{N}\Big)\zeta_i(0,x)\right]-\int_0^t\exp[\Psi_{\mathbf{J}}(t,\zeta(u)]du\right]\right]=1,$$

and leads to the rate function

$$\frac{1}{2}\int_0^T\left\|\tilde{\rho}_t-\frac{1}{2}\nabla A\chi\nabla\tilde{\rho}\right\|^2_{-1,A(\tilde{\rho})}dt.$$

This is very similar to the single color case. Weak perturbations with rates that depend on location, time, and color provide lower bounds in terms of relative entropy and Girsanov's formula. One optimizes the perturbations over those that have the same end effect but least relative entropy. This will match the upper bound. Some technical details come up. If we are not considering large deviations, the solution $\rho(t,u)$ of the total density satisfies $0<\rho(t,u)<1$ for $t>0$, and there is no problem in solving the linear equations for the densities of colors. But when we consider large deviation, we need to consider what happens when $\rho(t,u)=1$ on a set of positive measure in $[0,T]\times\mathbb{T}^d$. This requires an understanding of degenerate diffusions with minimal regularity; see [**22**]. Once we understand the behavior of multicolor systems, we can go on to empirical processes where we have a simple exclusion process and every particle is tagged.

$$\gamma_N(\,\cdot\,,\omega)=\frac{1}{N^d}\sum_{j=1}^{\mathbf{k}(N)}\delta_{x_j(\cdot)}$$

is a random measure on $D[[0,T];\mathbb{T}^d]$ with total mass $\bar{\rho}=\lim_{N\to\infty}k(N)/N$. Its distribution is $\mathcal{R}_N$. We know that that the sequence is tight. We want to show that any subsequence of $\mathcal{R}_N$ that converges is concentrated on a single measure consisting of a Markov process Q with initial distribution $\rho_0(u)du$ and time-dependent generator

$$\frac{1}{2}\nabla S(\rho(t,u))\cdot\nabla+[S(\rho(t,u)-C]\frac{\nabla\rho(t,u)}{\rho(t,u)}.$$

Since we have tightness, we only need to identify uniquely the possible limit. Let $0<t_1<\cdots<t_p=1$ be p time points. Let $A_1,\dots,A_p$ be a partition of $\mathbb{T}^d$ into p sets that are continuity sets for Lebesgue measure, i.e., have boundaries that have Lebesgue measure 0. It is enough to show that the random variables $\gamma[\xi(t_i)\in A_{s(i)}$ for $i=1,\dots,p]$ converge in probability to $Q[\xi(t_i)\in A_{s(i)}$ for $i=1,\dots,p]$ for all choices of $\{s_i\}$, time points $t_1,\dots,t_p$, sets $A_1,\dots,A_p$, and $p\geq 1$.

We can at time 0 color the particles from $A_1,\dots,A_p$ by different colors and see how their distributions are at time t_1. Refine the color code to incorporate the information at time 0 and t_1, and proceed in a similar fashion. The law of large numbers for the multicolor case guarantees that the $\mathcal{R}_N$ is concentrated on those measures γ whose finite-dimensional distributions coincide with that of Q. Notice that the total mass is not 1 but some constant $\rho<1$. Having this for all p and choices of time points and partitions of $\mathcal{T}^d$ and the tightness of $\mathcal{R}_N$ proves the result; see [**23**].

One can try the same thing at the level of large deviations. One has to calculate the limit of the rate functions as time steps shrink and the cells get small. It is a challenging calculation carried out in [**21**] and the rate function is rather interesting. Let P be a stochastic process with continuous paths on $C[[0,T];\mathbb{T}^d]$. The dynamic rate function is calculated as follows. P has marginals $\rho(t,u)$. The rate function of Section 5.5 is obtained by weak perturbations.

The law of large numbers holds for the solution of

$$\rho_t = \frac{1}{2}\nabla C \nabla \rho.$$

We can perturb it to a solution of

$$\rho_t = \frac{1}{2}\nabla C \nabla \rho - \nabla \cdot \rho(1-\rho)Cb \tag{7.25}$$

with an entropy cost of

$$\mathcal{E}(b) = \frac{1}{2}\int_0^T \int_{\mathbb{T}^d} \rho(1-\rho)\langle b, Cb\rangle d\theta\, dt.$$

For an optimal rate, one minimizes $\mathcal{E}(b)$ subject to (7.25). The process will evolve as the diffusion process P_b with generator

$$\frac{1}{2}\nabla S(\rho(t,\theta))\nabla + \left[(S(\rho(t,\theta)) - C)\frac{\nabla\rho(t,\theta)}{2\rho(t,\theta)}\right]\cdot\nabla + (1-\rho(t,\theta))Cb\cdot\nabla.$$

Solutions other than minimizing b have a role to play. If $\phi(t,\theta)$ is a smooth function of t,θ with values in $\mathbb{R}^d$, one can compute the stochastic integral

$$E^{P_b}\left[\int_0^T \langle\phi(t,\theta(t))\cdot d\theta(t)\rangle\right].$$

Given a process Q, in addition to matching the marginals, one may want to match the circulation as well. Find $b = b_Q$ such that the process P_b matches Q in the sense that

$$E^{P_b}\left[\int_0^T \langle\phi(t,\theta(t))\cdot d\theta(t)\rangle\right] = E^{Q}\left[\int_0^T \langle\phi(t,\theta(t))\cdot d\theta(t)\rangle\right]$$

and

$$E^{P_b}\left[\int_0^T \psi(t,\theta(t))dt\right] = E^{Q}\left[\int_0^T \psi(t,\theta(t))dt\right].$$

If b exists, it is unique, and if does not, $I(Q)$ is infinite. With $b = b_Q$, we have P_{b_Q} and, finally, the dynamical rate function for Q is

$$I(Q) = \mathcal{E}(b_Q) + h(Q; P_{b_Q}).$$

It is interesting to notice the universal, simple form of the rate function involving only $C, S(\rho), A(\tilde{\rho})$, and $\chi(\tilde{\rho})$.

https://doi.org/10.1090//cln/027/08

CHAPTER 8

Some Comments About TASEP

TASEP (totally asymmetric simple exclusion process) is a one-dimensional simple exclusion model on $\mathbb{Z}_N$ in which the particles are allowed to jump to the neighboring site in only one direction, i.e., from site x to site $x+1$ if it is free. Formally one can easily derive the hydrodynamic limit when space and time are scaled by the same factor,

$$d\frac{1}{N}\sum J\left(\frac{1}{N}\right)\eta_t(x) = \sum_x\left[J\left(\frac{x+1}{N}\right) - J\left(\frac{x}{N}\right)\right]\eta_t(x)(1-\eta_t(x+1)) + M_N(t).$$

Again, it is easy to check that the quadratic variation of the martingale goes to 0 with N. If we believe, as before, that the distributions are locally Bernoulli at a suitable density, we obtain

$$\frac{d}{dt}\langle J,\rho\rangle = \langle J',\rho(1-\rho)\rangle,$$

a weak formulation of the Burger's equation

$$\rho_t + [\rho(1-\rho)]_x = 0 \tag{8.1}$$

with some initial condition $\rho(0,u) = \rho_0(u)$ on the circle $\mathbb{T}$. This equation may not have smooth solutions, even if ρ_0 is smooth. It can develop shocks at a later time. There are results that prove local existence of smooth solutions. While a weak solution to (8.1) exists for all times, it is not unique.

The following general facts are known. Among all weak solutions there is one that satisfies an entropy condition. A smooth solution always satisfies it. If f is a smooth function, formally,

$$\begin{aligned}\frac{d}{dt}f(\rho(t,x)) &= -f'(\rho(t,x)[\rho(t,x)(1-\rho(t,x))]_x\\ &= f'(\rho(t,x))(1-2\rho(t,x))\rho_x(t,x) = -[g(\rho(t,x))]_x\end{aligned}$$

where $g'(\rho) = f'(\rho)(1-2\rho)$. This can be written in the weak form as

$$\frac{d}{dt}\langle J, f(\rho)\rangle = \langle J', g(\rho)\rangle. \tag{8.2}$$

If ρ is not continuous, (8.2) does not follow from (8.1). The entropy condition under which uniqueness is valid requires that, if $f(\rho)$ is a convex function of ρ, then

$$\frac{d}{dt}f(\rho(t,x)) + [g(\rho(t,x))] \le 0$$

as a distribution. It is known that the validity of the entropy condition for any strictly convex function implies the validity of the condition for all convex functions. The limit of the TASEP is the unique weak solution satisfying the entropy condition. This has been studied by many; see, for instance, [**26**].

A possible weak solution $\rho(t,x)$ with a discontinuity along a line $x = ct$ on $[0,T]\times\mathbb{R}$ (rather than on $[0,T]\times\mathbb{T}$) is given by

$$\rho(t,x) = \begin{cases} \rho_- & \text{if } x < ct, \\ \rho_+ & \text{if } x > ct. \end{cases} \tag{8.3}$$

For $\rho(t,x)$ to be a weak solution of $\rho_t + (\rho(1-\rho))_x = 0$, it has to satisfy

$$\begin{aligned}\rho_-(1-\rho_-)\int_0^T dt\int_{-\infty}^{ct}\phi_x(t,x)dx + \rho_+(1-\rho_+)\int_0^T dt\int_{ct}^{\infty}\phi_x(t,x)dx \\ + \rho_-\int_{-\infty}^{\infty}dx\int_{\frac{x}{c}}^{T}\phi_x(t,x)dt + \rho_+\int_{-\infty}^{\infty}dx\int_0^{\frac{x}{c}}\phi_x(t,x)dt = 0.\end{aligned}$$

It is easy to check now that this implies $c = (1-\rho_-\rho_+)$. There is another way of checking that shocks have to travel at their specific speeds, satisfying what is known as Rankine-Hugoniot conditions. If we consider an interval around a shock, the inflow is $\rho_-(1-\rho_-)$ and the outflow is $\rho_+(1-\rho_+)$. The difference has to be compensated by the shock moving to the left or right to compensate for the increase or decrease of the number of particles in the interval. This provides the relation

$$\rho_-(1-\rho_-) - \rho_+(1-\rho_+) = c(\rho_- - \rho_+).$$

Let us consider the space $\mathbb{T}$ and the lattice $\mathbb{Z}_N$ imbedded in it. A microscopic profile is created by randomly distributing particles independently at sites x/N with probability $\rho(x/N)$. Then by the law of large numbers this will result in a macroscopic profile $\rho(\theta)$ that will evolve in time as a weak solution of $\rho_t + (\rho(1-\rho))_x = 0$. This will be the macroscopic profile of the particle system at time Nt. The initial entropy of the particle system, ignoring an additive constant, is easily computed and

$$\sum_\eta p(\eta)\log p(\eta) \simeq N\int_{\mathbb{T}}\rho(\theta)\log\rho(\theta)d\theta.$$

From the theory of Markov processes, we know that this is nondecreasing. One would expect that at a future time the microscopic entropy and the macroscopic entropy maintain their relationship and that would make the macroscopic entropy

$$h(t) = \int_{\mathbb{T}}\rho(t,\theta)\log\rho(t,\theta)d\theta$$

nondecreasing. Since $h(t)$ is conserved for smooth solutions, any change in the value of entropy for piecewise smooth solutions comes from shocks, and the effect of a shock must be to reduce the entropy. This requires that $\rho_+ > \rho_-$. Otherwise the shock would be unstable and disappear, instantly replaced by what is known as a rarefaction wave. Shocks with $\rho_- > \rho_+$ are unstable weak solutions and do not appear as the weak limits of unperturbed systems. They can arise in large deviations, and the increase in total entropy caused by them has to be paid for by entropy contribution from the Girsanov perturbation of rates near the shock.

In the final analysis the rate function for large deviations TASEP take the following form. If $\rho(t,x)$ is not a weak solution of $\rho_t+(\rho(1-\rho))_x = 0$, the rate function $I(\rho) = +\infty$. If ρ is a weak solution, we look at the convex function $h(\rho) = \rho\log\rho + (1-\rho)\log(1-\rho)$. Then

$$[h(\rho)]_t = h'(\rho)\rho_t = -h'(\rho)[\rho(1-\rho)]_x = h'(\rho)(1-2\rho)\rho_x = -[g(\rho)]_x$$

where $g'(\rho) = h'(\rho)(1 - 2\rho)$. We examine the distribution $[h(\rho)]_t + [g(\rho)]_x$. For the actual solution, this is nonpositive as a measure and there is a unique ρ satisfying this that is the true limit. Otherwise, the large-deviation rate function for such a ρ is finite only if $[h(\rho)]_t + [g(\rho)]_x$ is a measure of bounded variation on $[0, T] \times \mathbb{T}$, and the rate function is the total measure of the positive part

$$I(\rho) = \int_0^T \int_{\mathbb{T}} [[h(\rho)]_t + [g(\rho)]_x]^+ dt\, dx.$$

The first results on TASEP were obtained by Rost [**24**]. He considered the situation where every site $x \leq 0$ was initially occupied by a particle and the sites $x > 0$ were all empty. $\rho_0(\theta)$ was 1 for $\theta < 0$ and 0 for $\theta > 0$. The hydrodynamic limit was established, the limiting solution being

$$\rho(t, \theta) = \begin{cases} 1 & \text{if } \theta < t, \\ \dfrac{t - \theta}{2t} & \text{if } -t < \theta < t, \\ 0 & \text{if } \theta > t. \end{cases}$$

The equation $\rho_t + [\rho(1 - \rho)]_x = 0$ can be integrated once and $\rho = U_x$ where U staisfies a Hamilton-Jacobi equation

$$U_t + U_x(1 - U_x) = 0$$

which has a uniquely defined variational solution. It will be Lipschitz and $\rho = U_x$ will be the weak solution we will need. The large-deviation upper bounds were obtained by Jensen in his thesis and appears in [**32**]. While there are partial results on lower bounds, complete results are not yet available.

Bibliography

[1] Ciesielski, Z. Heat conduction and the principle of not feeling the boundary. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* 14: 435–440, 1966.

[2] Cramér, H. Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrialles*, 736, 5–23. Lecture at the Colloque Consecré à la Théorie des Probabilités. III. Hermann, Paris, 1938.

[3] Donsker, M. D., and Varadhan, S. R. S. Asymptotic evaluation of certain Markov process expectations for large time. I. *Comm. Pure Appl. Math.* 28: 1–47, 1975. doi:10.1002/cpa.3160280102

[4] Donsker, M. D., and Varadhan, S. R. S. Asymptotic evaluation of certain Markov process expectations for large time. II. *Comm. Pure Appl. Math.* 28: 279–301, 1975. doi:10.1002/cpa.3160280206

[5] Donsker, M. D., and Varadhan, S. R. S. Asymptotic evaluation of certain Wiener integrals for large time. *Functional integration and its applications (Proc. Internat. Conf., London, 1974)*, 15–33. Clarendon Press, Oxford, 1975.

[6] Donsker, M. D., and Varadhan, S. R. S. Asymptotics for the Wiener sausage. *Comm. Pure Appl. Math.* 28(4): 525–565, 1975. doi:10.1002/cpa.3160280406

[7] Donsker, M. D., and Varadhan, S. R. S. Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.* 29(4): 389–461, 1976. doi:10.1002/cpa.3160300603

[8] Donsker, M. D., and Varadhan, S. R. S. On laws of the iterated logarithm for local times. *Comm. Pure Appl. Math.* 30(6): 707–753, 1977. doi:10.1002/cpa.3160300603

[9] Donsker, M. D., and Varadhan, S. R. S. On the number of distinct sites visited by a random walk. *Comm. Pure Appl. Math.* 32(6): 721–747, 1979. doi:10.1002/cpa.3160320602

[10] Donsker, M. D., and Varadhan, S. R. S. Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.* 36(2): 183–212, 1983. doi:10.1002/cpa.3160360204

[11] Donsker, M. D., and Varadhan, S. R. S. Asymptotics for the polaron. *Comm. Pure Appl. Math.* 36(4): 505–528, 1983. doi:10.1002/cpa.3160360408

[12] Freidlin, M. I., and Wentzell A. D. On small random perturbations of dynamical systems. *Russian Mathematical Surveys* 25(1): 1–56, 1970. doi:10.1070/RM1970v025n01ABEH001254

[13] Freidlin, M. I., and Wentzell, A. D. Random perturbations of Hamiltonian systems. *Mem. Amer. Math. Soc.* 109(523): viii+82 pp, 1994.

[14] Gertner, J. On large deviations from an invariant measure. *Teor. Verojatnost. i Primenen.* 22(1): 27–42, 1977.

[15] Guo, M. Z., Papanicolaou, G. C., and Varadhan, S. R. S. Nonlinear diffusion limit for a system with nearest neighbor interactions. *Comm. Math. Phys.* 118(1): 31–59, 1998.

[16] Kac, M., and Luttinger, J. M. Bose-Einstein condensation in the presence of impurities. II. *J. Mathematical Phys.* 15: 183–186, 1974. doi:10.1063/1.1666617

[17] Kipnis, C., and Landim, C. *Scaling limits of interacting particle systems.* Grundlehren der mathematischen Wissenschaften, 320. Springer, Berlin, 1999.

[18] Kipnis, C., Olla, S., and Varadhan, S. R. S. Hydrodynamics and large deviation for simple exclusion processes. *Comm. Pure Appl. Math.* 42(2): 115–137, 1989.

[19] Quastel, J. Diffusion of color in the simple exclusion process. *Comm. Pure Appl. Math.* 45(6): 623–679, 1992. doi:10.1002/cpa.3160450602

[20] Quastel, J. Large deviations from a hydrodynamic scaling limit for a nongradient system. *Ann. Probab.* 23(2): 724–742, 1995.

[21] Quastel, J., Rezakhanlou, F., and Varadhan, S. R. S. Large deviations for the symmetric simple exclusion process in dimensions $d \geq 3$. *Probab. Theory Related Fields* 113(1): 1–84, 1999. doi:10.1007/s004400050202

[22] Quastel, J., and Varadhan, S. R. S. Diffusion semigroups and diffusion processes corresponding to degenerate divergence form operators. *Comm. Pure Appl. Math.* 50(7): 667–706, 1997. doi:10.1002/(SICI)1097-0312(199707)50:7<667::AID-CPA3>3.3.CO;2-T

[23] Rezakhanlou, F. Propagation of chaos for symmetric simple exclusions. *Comm. Pure Appl. Math.* 47(7): 943–957, 1994. doi:10.1002/cpa.3160470703

[24] Rost, H. Nonequilibrium behaviour of a many particle process: density profile and local equilibria. *Z. Wahrsch. Verw. Gebiete* 58(1): 41–53, 1981. doi:10.1007/BF00536194

[25] Schilder, M. Some asymptotic formulas for Wiener integrals. *Trans. Amer. Math. Soc.* 125: 63–85, 1966. doi:10.2307/1994588

[26] Seppäläinen, T. Coupling the totally asymmetric simple exclusion process with a moving interface. *Markov Process. Related Fields* 4(4): 593–628, 1998.

[27] Strassen, V. An invariance principle for the law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 3: 211–226, 1964. doi:10.1007/BF00534910

[28] Varadhan, S. R. S. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* 19: 261–286, 1966. doi:10.1002/cpa.3160190303

[29] Varadhan, S. R. S. Diffusion processes in a small time interval. *Comm. Pure Appl. Math.* 20: 659–685, 1967. doi:10.1002/cpa.3160200404

[30] Varadhan, S. R. S. On the behavior of the fundamental solution of the heat equation with variable coefficients. *Comm. Pure Appl. Math.* 20: 431–455, 1967. doi:10.1002/cpa.3160200210

[31] Varadhan, S. R. S. Nonlinear diffusion limit for a system with nearest neighbor interactions. II. *Asymptotic problems in probability theory: stochastic models and diffusions on fractals (Sanda/Kyoto, 1990)*, 75–128. Pitman Research Notes in Mathematics Series, 283. Longman Scientific & Technical, Harlow, 1993.

[32] Varadhan, S. R. S. Large deviations for the asymmetric simple exclusion process. *Stochastic analysis on large scale interacting systems*, 1–27. Advanced Studies in Pure Mathematics, 39. Mathematical Society of Japan, Tokyo, 2004.

Large Deviations

S. R. S. VARADHAN

The theory of large deviations deals with rates at which probabilities of certain events decay as a natural parameter in the problem varies. This book, which is based on a graduate course on large deviations at the Courant Institute, focuses on three concrete sets of examples: (i) diffusions with small noise and the exit problem, (ii) large time behavior of Markov processes and their connection to the Feynman-Kac formula and the related large deviation behavior of the number of distinct sites visited by a random walk, and (iii) interacting particle systems, their scaling limits, and large deviations from their expected limits. For the most part the examples are worked out in detail, and in the process the subject of large deviations is developed.

The book will give the reader a flavor of how large deviation theory can help in problems that are not posed directly in terms of large deviations. The reader is assumed to have some familiarity with probability, Markov processes, and interacting particle systems.

For additional information and updates on this book, visit

www.ams.org/bookpages/cln-27

NEW YORK UNIVERSITY

ISBN 978-0-8218-4086-3

9 780821 840863

CLN/27