RARE EVENT ANALYSIS FOR BRANCHING PROCESSES AND INTERACTING PARTICLE SYSTEMS

Michael Evan Conroy

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research in the College of Arts and Sciences.

Chapel Hill
2021

Approved by:

Nilay Argon

Sayan Banerjee

Shankar Bhamidi

Amarjit Budhiraja

Mariana Olvera-Cravioto

ABSTRACT

Michael Evan Conroy: Rare Event Analysis For Branching Processes and Interacting Particle Systems
(Under the direction of Amarjit Budhiraja and Mariana Olvera-Cravioto)

This dissertation is concerned with the asymptotic analysis and simulation of rare events for two types of stochastic models, each studied with a different mode of analysis. Roughly the first half is a look at tail events for solutions to max-type stochastic fixed point equations that are constructed on weighted Galton-Watson processes, which are also describable as the all-time maximum of a branching random walk with perturbation. The asymptotic analysis of these tail events is approached from the lens of direct renewal-theoretic arguments after a suitable change of measure in the tradition of Cramér-Lundberg theory. In particular, an asymptotic expression for the tail events, known from implicit renewal theoretic arguments, is re-established with the intention of elucidating its connection with the underlying branching process. Also in the same spirit, the representation after change of measure derived in the analysis allows for an importance sampling approach to efficiently estimate the rare events. Related to spine changes of measure often used in the branching process literature, the new measure induces a structure on the underlying branching process that suggests even more efficient algorithms to approximate tail events for branching random walks, which are also pursued herein.

The second half concerns large deviation asymptotics for several weakly interacting particle models described through systems of stochastic differential equations. In particular, large deviation principles are established for empirical distributions of particle sates in the infinite particle limit. The underlying models include interaction between the particles that is either both weak and through small common driving noise of Freidlin-Wentzell type, or where the noise is small, but the only interaction is mean-field. In the case of interaction through correlated noise, the rate at which the common Brownian motion term becomes small determines a bifurcation in the form of the large deviation rate functions, with each regime corresponding to whether the correlated or uncorrelated driving noise dominates in the limit. Particle approximations to Feynman-Kac functionals are also considered in the context of these models, as are certain Sobolev-space valued maps called stochastic currents that are functionals of the particle states and are convenient to describe the asymptotic behavior of the corresponding empirical measures. The proofs of the large deviation principles employ weak convergence methods, and they are based on control representations of Laplace functionals of Brownian motion, which correspondingly lead to control representations of the large deviation rate functions.

To Allison

ACKNOWLEDGEMENTS

I can only take partial credit for this work, insofar as the support I received from others was crucial to its completion. Not the least of this support was provided by Christine, Alison, Sam, Ayla, and Shayna, who kept the department running and us graduate students on track. I particularly acknowledge Christine's help in applying to, and eventually securing, a job after graduation. Her help was integral to my navigating that daunting process. Of course, none of this would be possible without the mentorship of my advisors, Amarjit and Mariana. I benefited Amarjit's patient approach, and I enjoyed candid conversations about the state of academia with Mariana. Both are the source not only of the words herein, but also my future research directions and confidence as a mathematician. Interactions with other members of my committee—Nilay, Sayan, and Shankar—have also been an important part of my maturation as a researcher.

I also need to acknowledge my collaborators on the work presented here. Mariana and I worked with Bojan Basrak and Zbyszek Palmowski on the material of Chapter 2. I thank them for their support and their fullfilling collaboration. Chapter 3 is work done with Mariana, and the material of Chapters 4 and 5 is joint work between Amarjit and myself.

What is not present in these pages is the context of the friendships that inspired and motivated me day-to-day. I need to thank my peers Eric, Adam, Jack, Mark, Samopriya, and Brendan for their tolerance of my bothering math and statistics questions that I probably should have known the answers to. More importantly, however, I thank them for their camaraderie both inside and outside of Hanes Hall in navigating the doctoral program and always being around to hang out. I learned the meaning of solidarity from Lindsay, Jen, Julia, Danielle, James, Maja, Alyssa, Sam, Tim, April, Ben, Willa, Kelly, Mark, Emil, Abigail, Trisha, Nick, Katelyn, Heather and others, who were fiercely determined to make their community and campus a better place. These friends had an enormous impact on my personal growth during my five years in Chapel Hill, and I will forever admire their commitment to antiracism and antifascism. For so many of them and others in the UNC community, they achieved what they have in spite of the failings of university leadership.

Due to the COVID-19 pandemic, much was different than it should have been during my final year as a doctoral student. I was only able to get through due to the unconditional support and companionship of my spouse Allie and our pup Sadie. So many of us owe a debt to a few who showed grace during this time, whether in the classroom, at work, or at home. Apart from UNC's harmful determination to open in a pandemic, I was extremely lucky in this regard. I am also very thankful for the support of my family during my many years of education and for believing in me throughout my life. Lastly, I thank my students, who entrusted me with their statistics education, and who helped me grow as an educator.

TABLE OF CONTENTS

LIST OF TABLES . . . x

LIST OF FIGURES . . . xi

LIST OF ABBREVIATIONS . . . xii

LIST OF SYMBOLS . . . xiii

1 Introduction . . . 1

1.1 Tail Events for Stochastic Fixed Point Equations . . . 1

1.1.1 Some Renewal Theory . . . 3

1.1.2 The Cramér-Lundberg Asymptotic . . . 5

1.1.3 Siegmund's Algorithm . . . 8

1.1.4 Branching Recursions . . . 9

1.2 Large Deviations For Interacting Particle Models . . . 14

1.2.1 Large Deviation and Laplace Principles . . . 14

1.2.2 Weakly Interacting Diffusions . . . 16

1.2.3 Control Representations . . . 19

1.3 Contributions . . . 24

2 Importance Sampling For Maxima on Trees . . . 27

2.1 Change of Measure Along a Path . . . 29

2.2 The Markov Renewal Theorem . . . 35

2.3 Main Result . . . 36

2.4 An Importance Sampling Algorithm . . . 37

2.4.1 Examples . . . 40

2.4.2 Numerical Experiments . . . 44

2.5 Proofs . . . 47

2.5.1 The Distribution of $\mathcal{T}$ Under $\widetilde{P}$ 47
2.5.2 Proof of Theorem 2.3.1 50
2.5.3 Proof of Corollary 2.3.1 62
2.5.4 The Importance Sampling Estimator $Z(t)$ 64
2.5.5 The Bounded Perturbations in the Non-Branching Case 67
3 Efficient Hybrid Estimation for Tail Events on Trees 69
3.1 A Hybrid Importance Sampler 73
3.1.1 Properties 76
3.1.2 The Population Dynamics Algorithm 78
3.2 Proofs 80
3.2.1 Auxiliary Results 81
3.2.2 Proof of Theorem 3.1.5 86
3.2.3 The Population Dynamics Case 97
4 Large Deviations for Interacting Diffusions with Common Small Noise 99
4.1 Main Results 103
4.1.1 Diffusions Interacting Through the Empirical Distribution 103
4.1.2 Interacting Particle Systems for Feynman-Kac Functionals 106
4.1.3 Intensity of the Common Noise 110
4.2 Proof of Theorem 4.1.1. 112
4.2.1 Laplace Upper Bound 114
4.2.2 Laplace Lower Bound 122
4.2.3 Rate Function Property 125
4.3 Proof of Theorem 4.1.2. 126
4.3.1 Laplace Upper Bound 127
4.3.2 Laplace Lower Bound 134
4.3.3 Rate Function Property of I_2 138
4.4 Proof Sketch of Theorem 4.1.3. 138
4.4.1 The Case $\sqrt{n}\kappa(n) \to 0$ 139
4.4.2 The Case $\sqrt{n}\kappa(n) \to \infty$ 140

4.5 Proof Sketch of Theorem 4.1.1(i) 142
4.6 Proof Sketch of Theorem 4.1.2(i) 143
5 Large Deviations For Stochastic Currents with Mean Field Interaction and Vanishing Noise 146
5.1 Preliminaries and Main Result 150
5.1.1 Main Conditions. 151
5.1.2 Stochastic Currents 153
5.1.3 Rate Function 154
5.1.4 Main Results 156
5.2 Laplace Asymptotics and Variational Representation 158
5.2.1 Variational Representation 159
5.2.2 Tightness Properties. 160
5.2.3 Proof of the Upper Bound 161
5.2.4 Proof of the Lower Bound 162
5.2.5 Rate Function Property 166
5.2.6 Law of Large Numbers 166
5.2.7 Equivalent Formulation of the Rate Function 167
5.3 Proofs of Key Lemmas 170
5.3.1 Proof of Lemma 5.2.3 172
5.3.1.1 Proof of Lemma 5.2.3(i) 176
5.3.1.2 Proof of Lemma 5.2.3(ii) 177
5.3.2 Proof of Lemma 5.2.5 183
5.3.3 Proof of Lemma 5.2.4 186
5.4 Proofs of Sobolev Space Results 187
5.4.1 Proof of Lemma 5.2.1 188
5.4.2 Proof of Lemma 5.3.7 190
5.4.3 Proof of Lemma 5.3.3 191
5.4.4 Proof of Lemma 5.3.4 193
APPENDIX 195
A.1 Proofs of Background Theorems 196

A.2 Useful Lemmas 201
A.3 The Space $\mathbf{H^s}$ 203
REFERENCES 209

LIST OF TABLES

2.1 Importance sampling algorithm 40

2.2 Numerical results for the branching M/M/1 queue 45

2.3 Numerical results for C's on the N-simplex 46

3.1 Hybrid importance sampling algorithm 76

3.2 Population dynamics algorithm 79

LIST OF FIGURES

1.1 A typical moment generating function 6

1.2 A weighted branching tree 11

1.3 The branching Laplace transform 14

2.1 The spine of $\mathcal{T}$ and the Markov chain consisting of subtrees. 36

2.2 A branching random walk under $\widetilde{P}$ 45

2.3 Log probability plots for numerical experiments 47

3.1 The spine decomposition of $\mathcal{T}$ 71

LIST OF ABBREVIATIONS

SFPE	Stochastic fixed point equation
SDE	Stochastic differential equation
i.i.d.	Independent and identically distributed
d.R.i.	Directly Riemann integrable
LDP	Large deviation principle
LLN	Law of large numbers
PDE	Partial differential equation
M/M/1	Memoryless arrival times/memoryless service times/1 server
w.p.	With probability
a.e.	Almost everywhere
a.s.	Almost surely

LIST OF SYMBOLS

$\mathbb{R}$	Set of real numbers
$\mathbb{Z}$	Set of integers
$\mathbb{N}$	Set of natural numbers, $\{0, 1, 2, \ldots\}$
$\mathbb{R}_+$	Set of nonnegative real numbers, $[0, \infty)$
$\mathbb{N}_+$	Set of positive natural numbers, $\{1, 2, 3, \ldots\}$
$\doteq$	Definition
$\stackrel{\mathcal{D}}{=}$	Equality in distribution
$\stackrel{P}{\to}$	Convergence in P-probability
$\Rightarrow$	Convergence in distribution
$\varnothing$	Empty set
$\emptyset$	Empty string or root node
$\prec$	Length-lexicographic ordering
$\vee$	Maximum
$\wedge$	Minimum
$1(A), 1_A$	Indicator function at the set A
$\lVert\cdot\rVert, \lvert\cdot\rvert$	Euclidean norm
$\langle F, \varphi\rangle$	Evaluation of a distribution F with respect to a function φ
$\langle \mu, \varphi\rangle$	$\int \varphi \, d\mu$ for a measure μ
$\mathcal{P}(S)$	Set of probability measures on the space S
$\mathcal{B}(S)$	Borel σ-field on S
$\sigma(X)$	σ-field generated by the random variable X
$\mu \otimes \nu$	Product measure of μ and ν
$\mathcal{F} \otimes \mathcal{G}$	Product σ-field of $\mathcal{F}$ and $\mathcal{G}$
$\mathcal{R}$	Space of finite measures r on $[0, T] \times \mathbb{R}^m$ such that $r([0, t] \times \mathbb{R}^m) = t$
$\mathcal{R}_1$	Subset of $\mathcal{R}$ with $\int_{[0,T]\times\mathbb{R}^m} \lvert x\rvert \, r(dt, dx) < \infty$
r_t	Derivative measure of $r \in \mathcal{R}$, where $r(dt, dx) = r_t(dx) \, dt$
$[\theta]_k, \theta_{(k)}$	kth marginal distribution of the measure θ
E_θ	Expectation with respect to the probability measure θ
$\mathcal{C}(R, S), \mathcal{C}(R)$	Space of continuous functions $f : R \to S$ [resp. $f : R \to \mathbb{R}$]
$\mathcal{C}_b(R, S)$	Continuous and bounded functions from R to S
$\mathcal{C}_b^{\text{a.e.}}(R, S)$	Functions from R to S that are continuous and bounded a.e.

$\mathcal{C}_c(R,S)$	Compactly supported continuous functions from R to S
$\mathcal{C}^k(R,S)$	k-times differentiable functions from R to S, $k \in \mathbb{N} \cup \{\infty\}$
$\mathcal{C}_c^k(R,S)$	$\mathcal{C}_c(R,S) \cap \mathcal{C}^k(R,S)$
$L^2(\mu,R,S)$, $L^2(R,S)$	μ-square integrable [resp. Lebesgue-square integrable] functions from R to S
$\mathcal{P}_1(S)$	Subset of $\mathcal{P}(S)$ with finite means
d_1	Wasserstein-1 distance on $\mathcal{P}_1(S)$
$\mathcal{M}_+(S)$	Space of positive measures on S
d_{BL}, d_{bl}	Bounded Lipschitz metric on $\mathcal{P}(S)$
$f = O(g)$	There exists a constant $c < \infty$ such that $f \leq cg$
$\|\cdot\|_\infty$	Supremum norm, $\|f\|_\infty = \sup_{x\in S} |f(x)|$ when the domain of f is S
∇f	Gradient of f, $(\partial f/\partial x_1, \ldots, \partial f/\partial x_d)^\mathsf{T}$ when $f : \mathbb{R}^d \to \mathbb{R}$
$\hat{f}$	The Fourier transform of a function f
S'	Topological dual of the space S
$H^s(R,S)$	Fractional Sobolev space of $L^2(R,S)$ functions with index $s \in \mathbb{R}$
$\|\cdot\|_s$	Norm on the space $H^s(R,S)$
$\mathbf{H}^\mathbf{s}$	Sobolev space $H^{s_1}(U, H^{s_2}(\mathbb{R}^d,\mathbb{R}^d))$ for $\mathbf{s} = (s_1,s_2) \in \mathbb{R}^2$ and some $U \subset \mathbb{R}$
D^α	Distributional derivative of multi-index order $\alpha = (\alpha_1,\ldots,\alpha_n)$
$|\alpha|$	$|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$
$\alpha!$	$\alpha_1! \cdot \alpha_2! \cdots \alpha_n!$
$\binom{\alpha}{\beta}$	$\binom{\alpha_1}{\beta_1} \cdot \binom{\alpha_2}{\beta_2} \cdots \binom{\alpha_n}{\beta_n}$

CHAPTER 1

Introduction

This dissertation has two distinct parts, differentiated by the models considered, the types of results proved, and also the analytical techniques used. The first two chapters deal with branching process solutions to certain stochastic fixed point equations (SFPEs), and the final two concern stochastic differential equation (SDE) models of weakly interacting particle systems. However, a common theme between the chapters of this dissertation is that the results involve the approximation, simulation, and asymptotic analysis of *rare events*—events that occur with small probability. In particular, a fundamental question is precisely how small the chance of a rare event occuring is for different stochastic models. The approaches used to answer this question herein lie under the historical umbrellas of two well-formulated areas related to the analysis of rare events: Cramér-Lundberg theory and the unified formulation of *large deviations* due to Varadhan [98].

Included here is an overview of the models and problems studied both in historical and modern contexts, as well as a description of the contributions of the work presented in this dissertation. We begin with some relevant background in Camér-Lundberg theory, renewal theory, and large deviation theory. A more comprehensive background on these topics can be found in [7, 35, 21].

1.1 Tail Events for Stochastic Fixed Point Equations

The *Lindley process*, or reflected random walk, is one of the most fundamental stochastic models in queueing theory and applied probability. For $\{X_n : n \geq 1\}$ i.i.d., it is defined recursively by

$$W_{n+1} = (W_n + X_n)^+, \qquad n \geq 1, \tag{1.1.1}$$

with some starting value W_1. As a canonical example of such a process, consider i.i.d. arrival times $\{\tau_n : n \geq 1\}$ independent of i.i.d. service times $\{\chi_n : n \geq 1\}$ for the customers in a single server queue with a first-in-first-out service criterion. When $X_n = \chi_n - \tau_n$, then W_n represents the waiting time for the nth customer to arrive in the queue.

Suppose $W_1 \equiv 0$ and define $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ for $n \geq 1$, so that S_n is the associated random walk on $\mathbb{R}$ with increment distribution given by X_1. Note that for each n,

$$\begin{aligned} W_n + X_n &= (W_{n-1} + X_{n-1})^+ + X_n \\ &= \max\{W_{n-1} + X_{n-1} + X_n, X_n\}. \end{aligned}$$

Consequently, $W_{n+1} = \max\{W_{n-1} + S_n - S_{n-2}, S_n - S_{n-1}, 0\}$. The calculation above can be iterated to yield that

$$W_n = \max\{S_n, S_n - S_1, \ldots, S_n - S_{n-1}, 0\} \stackrel{\mathcal{D}}{=} \max_{k \leq n} S_k \doteq M_n,$$

for all $n \geq 1$, where $\stackrel{\mathcal{D}}{=}$ indicates equality in distribution. Here, $W_n \stackrel{\mathcal{D}}{=} M_n$ in the sense of one-dimensional distributions, *not* as processes. Indeed, M_n is nondecreasing in n while W_n is not. Because M_n is nondecreasing, it has an a.s. limit M, which is the all-time maximum of the random walk S_n. When $E[|X_1|] < \infty$ and $E[X_1] < 0$, the law of large numbers $n^{-1}S_n \to E[X_1]$ ensures that $S_n \to -\infty$ and hence $M < \infty$ a.s. Clearly, $W_n \Rightarrow M$, where $\Rightarrow$ denotes convergence in distribution, and hence M is a finite solution to the stochastic fixed point equation (SFPE)

$$W \stackrel{\mathcal{D}}{=} (W + X)^+, \tag{1.1.2}$$

where X is a copy of X_1 independent of W. (1.1.2) is the limiting version of (1.1.1) known as the *Lindley equation*, and it is an equation of probability measures. That is, W satisfies the Lindley equation when $P(W \in A) = P\left((W + X)^+ \in A\right)$ for all Borel sets $A \subset \mathbb{R}_+ = [0, \infty)$.

The solution to (1.1.2) describes the stationary waiting time distribution of the queue described above, and it also shows up in the theory of insurance risk. In the latter context, it is of practical interest to know the values of so-called *ruin probabilities* $P(W > t)$, which for large $t > 0$ are rare events for the distributional solution to the Lindley equation. These probabilities can be described by the *Cramér-Lundberg* asymptotic for the maximum of a random walk, namely

$$P(W > t) \sim He^{-\alpha t} \qquad \text{as } \ t \to \infty, \tag{1.1.3}$$

for some $H \geq 0$, when α exists such that $m(\alpha) \doteq E\left[e^{\alpha X}\right] = 1$ and $E\left[Xe^{\alpha X}\right] > 0$. Here and throughout, $f(x) \sim g(x)$ as $x \to \infty$ means that $\lim_{x\to\infty} f(x)/g(x) = 1$. This result is straightforward to prove with some basic renewal theory.

1.1.1 Some Renewal Theory

Here we introduce some basic ideas in renewal theory that show up in subsequent chapters. First is the concept of a directly Riemann integrable function.

Definition 1.1.1. *A nonnegative function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is said to be directly Riemann integrable (d.R.i.) if*

$$\lim_{\varepsilon \downarrow 0} \varepsilon \sum_{n \in \mathbb{N}} \sup_{n\varepsilon < x \le (n+1)\varepsilon} f(x) = \lim_{\varepsilon \downarrow 0} \varepsilon \sum_{n \in \mathbb{N}} \inf_{n\varepsilon < x \le (n+1)\varepsilon} f(x) \in (-\infty, \infty).$$

A function $f : \mathbb{R} \to [0, \infty)$ is d.R.i. if the above holds with $\mathbb{N}$ replaced by $\mathbb{Z}$.

Loosely, a function is d.R.i. if it does not "fluctuate" too much at infinity. As a consequence of $f : \mathbb{R}_+ \to \mathbb{R}_+$ being d.R.i., it is also integrable, and

$$\int_0^\infty f(x)\, dx = \lim_{\varepsilon \downarrow 0} \varepsilon \sum_{n \in \mathbb{N}} \sup_{n\varepsilon < x \le (n+1)\varepsilon} f(x) = \lim_{\varepsilon \downarrow 0} \varepsilon \sum_{n \in \mathbb{N}} \inf_{n\varepsilon < x \le (n+1)\varepsilon} f(x).$$

Direct Riemann integrability is often a condition that is needed to be checked to apply theorems in renewal theory, however checking the definition directly can be tedious. The following lemma provides some useful alternate criteria (see [7, Proposition V.4.1] for a proof). Let $\mathcal{C}_b(\mathbb{R}_+)$ denote the set of bounded and continuous functions on $\mathbb{R}_+$, and let $\mathcal{C}_b^{\text{a.e.}}(\mathbb{R}_+)$ be the set of functions on $\mathbb{R}_+$ that are continuous and bounded a.e. with respect to Lebesgue measure. If a function f is d.R.i. then necessarily $f \in \mathcal{C}_b^{\text{a.e.}}(\mathbb{R}_+)$.

Lemma 1.1.1. *Sufficient conditions for $f : \mathbb{R}_+ \to \mathbb{R}_+$ to be d.R.i. are*

(i) $f \in \mathcal{C}_b^{a.e.}(\mathbb{R}_+)$ *and*

$$\sum_{n \in \mathbb{N}} \sup_{n\varepsilon < x \le (n+1)\varepsilon} f(x) < \infty$$

for some $\varepsilon > 0$,

(ii) $f \in \mathcal{C}_b^{a.e.}(\mathbb{R}_+)$ *and $f \le g$ where g is d.R.i.,*

(iii) f is nonincreasing and $\int_0^\infty f(x)\, dx < \infty$.

A sufficient condition for $f : \mathbb{R} \to \mathbb{R}_+$ to be d.R.i. is for both $f(x)$ and $f(-x)$ to be d.R.i. on $\mathbb{R}_+$.

Here and in the sequel, it will be useful to have results about so-called *renewal equations*, which are convolution equations of the form

$$K(t) = g(t) + F * K(t) = g(t) + \int_0^t K(t-s)\, F(ds), \qquad (1.1.4)$$

for a function $F : \mathbb{R}_+ \to \mathbb{R}_+$ that is nondecreasing and right continuous, usually a cumulative distribution function (CDF) of some random variable. We understand integration against $F(ds)$ as integration against the Borel measure μ_F induced by F in the usual way, namely

$$\mu_F((a,b]) = F(b) - F(a), \qquad (a,b] \subset \mathbb{R}.$$

Given such an F, we define the *renewal measure* U by

$$U(dx) = \sum_{n=0}^{\infty} F^{*n}(dx),$$

where F^{*n} denotes the nth-fold convolution of F, i.e.

$$F^{*n}(t) = \int_0^t F^{*(n-1)}(t-s)\,F(ds) = \int_0^t F(t-s)\,F^{*(n-1)}(ds)$$

for all $n \geq 1$. Note that by iterating (1.1.4), we can write

$$K(t) = U * g(t) = \int_0^t g(t-s)\,U(ds).$$

For the following theorem and throughout, the measure induced by F is said to be *lattice* if it is concentrated on a set $\{n\theta : n \in \mathbb{N}\}$ for some $\theta > 0$.

Theorem 1.1.2 (Key Renewal Theorem)**.** *Suppose that $F : \mathbb{R}_+ \to \mathbb{R}_+$ is nonlattice, nondecreasing, right continous function with $\int_0^\infty F(dx) = 1$. Let $\mu = \int_0^\infty x\,F(dx)$. If* (1.1.4) *holds for g that is d.R.i., then*

$$\lim_{t\to\infty} K(t) = \lim_{t\to\infty} U * g(t) = \frac{1}{\mu}\int_0^\infty g(s)\,ds.$$

In the case that F is the CDF of nonnegative X_i's with $E[X_1] = \mu$, the theorem says that

$$\lim_{t\to\infty} E\left[\sum_{n=0}^{\infty} g(t-S_n)\right] = \frac{1}{\mu}\int_0^\infty g(s)\,ds$$

when g is d.R.i.

The idea of proving limit theorems for renewal processes is to formulate a renewal equation to which a renewal theorem can be applied. In later chapters, it will be useful to employ the following version of the renewal theorem for the whole real line due to Athreya, McDonald, and Ney [9, Theorem 4.2].

Corollary 1.1.3. *If $\{S_n : n \geq 0\}$ is a nonlattice random walk on $\mathbb{R}$ (i.e. its increment distribution is nonlattice) with $\mu = E[S_1] > 0$ and $g : \mathbb{R} \to \mathbb{R}_+$ is d.R.i., then*

$$\lim_{t\to\infty} E\left[\sum_{n=0}^{\infty} g(t - S_n)\right] = \frac{1}{\mu}\int_{-\infty}^{\infty} g(s)\, ds.$$

1.1.2 The Cramér-Lundberg Asymptotic

The Cramér-Lundberg Asymptotic (1.1.3) can be seen as a direct consequence of the key renewal theorem applied to a random walk after a suitable change of measure. Much of Chapters 1 and 2 employs a branching generalization of *exponential tilting* in its analysis, and this is also the basis for Cramér-Lundberg theory. If $S_n = \sum_{i=1}^{n} X_i$ is a random walk with increment moment generating function $m(\theta) = E\left[e^{\theta X_1}\right] < \infty$, then let $\kappa(\theta) = \log m(\theta)$ and note that $\{e^{\theta S_n - n\kappa(\theta)} : n \geq 1\}$ defines a positive, mean-one martingale for all θ for which $\kappa(\theta) < \infty$. For such θ, $m'(\theta)$ exists and it is equal to $E\left[X_1 e^{\theta X_1}\right]$, a fact which follows from standard results about convergent sums once $m(\theta)$ is expanded as a Taylor series. Letting $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ be the filtration generated by the increments $\{X_n\}$, we can define a new probability measure P_θ on $\sigma(X_i : i \geq 1)$ by

$$P_\theta(A) = E\left[1(A) e^{\theta S_n - n\kappa(\theta)}\right], \qquad A \in \mathcal{F}_n. \tag{1.1.5}$$

Under this new measure, the law of the random walk is "tilted" in the sense that its new mean is

$$E_\theta[X_1] = E\left[X_1 e^{\theta X_1 - \kappa(\theta)}\right] = \frac{m'(\theta)}{m(\theta)} = \kappa'(\theta).$$

The amount that the tilting changes the drift of the random walk can be selected for various applications by choose an appropriate value of θ.

One needs to be careful that (1.1.5) indeed defines a legitimate (unique) measure on $\sigma(X_i : i \geq 1)$. Consider, for example, the events $A = \{n^{-1}S_n \to E[X_1]\}$ and $A_\theta = \{n^{-1}S_n \to \kappa'(\theta)\}$ for $\theta > 0$. Clearly A and A_θ are disjoint, however $P(A) = 1$ and $P_\theta(A_\theta) = 1$ by the strong law of large numbers, ensuring the nonexistence of the Radon-Nikodym derivative dP_θ/dP. However, it is clear from the formula (1.1.5) that absolute continuity is present up to each finite time step n. Indeed, as the next theorem shows, the time step can also be random. Its proof is standard and can be found in the appendix.

Theorem 1.1.4. *Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $\{\mathcal{F}_n : n \in \mathbb{N}\}$, and let $\mathcal{F}_\infty = \sigma\left(\bigcup_{n=0}^{\infty} \mathcal{F}_n\right)$.*

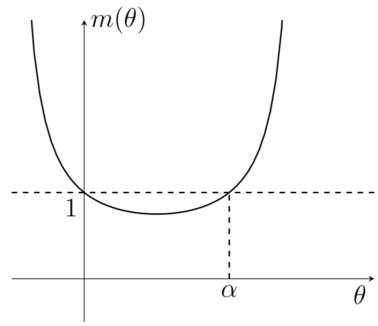


Figure 1.1: A typical moment generating function

(i) If $\{L_n : n \in \mathbb{N}\}$ *is a nonnegative martingale with respect to* $\mathcal{F}_n$ *such that* $E[L_n] = 1$ *for all* n*, then there exists a unique probability measure* P^* *on* $(\Omega, \mathcal{F}_\infty)$ *such that*

$$P^*(A) = E[1(A)L_n] \qquad \text{for} \;\; A \in \mathcal{F}_n.$$

(ii) If τ *is an* $\mathcal{F}_n$*-stopping time and* $A \in \mathcal{F}_\tau$ *such that* $A \subset \{\tau < \infty\}$*, then*

$$P(A) = E^*\left[1(A)L_\tau^{-1}\right],$$

where E^* *denotes expectation on* $(\Omega, \mathcal{F}_\infty, P^*)$.

The main assumption for (1.1.3) to hold, namely that there exists $\alpha > 0$ such that $E\left[e^{\alpha X}\right] = 1$ and $E\left[Xe^{\alpha X}\right] > 0$, is known as the *Cramér condition*, and while limiting the scope to light-tailed distributions it is fairly general in that context. Indeed, let $m(\theta) = E\left[e^{\theta X}\right]$ denote the moment generating function of X, and so that we have a finite solution to the Lindley equation, suppose $E[X] < 0$. From observing the Lindley equation, it is clear that for its solution, $W \geq 0$ a.s., and so to avoid triviality, let us assume that $P(X > 0) > 0$, which in turn implies that there is some $\varepsilon > 0$ such that $P(X > \varepsilon) > 0$. For θ such that $m(\theta) < \infty$, we have that $m'(\theta) = E\left[Xe^{\theta X}\right]$, and in particular $m(0) = 1$ and $m'(0) = E[X] < 0$. Also, by Jensen's inequality,

$$\liminf_{\theta \to \infty} m(\theta) \geq \liminf_{\theta \to \infty} E\left[e^{\theta X} 1(X > \varepsilon)\right] \geq \liminf_{\theta \to \infty} e^{\theta \varepsilon} P(X > \varepsilon) = \infty.$$

This together with the continuity and convexity of $m(\theta)$ ensures the existence of $\alpha > 0$ such that $m(\alpha) = 1$ and $m'(\alpha) > 0$. This typical situation is illustrated in Figure 1.1.

If we choose this α to induce the change of measure (1.1.5), then in particular

$$E_\alpha[X_1] = \frac{m'(\alpha)}{m(\alpha)} = E\left[Xe^{\alpha X}\right] > 0, \tag{1.1.6}$$

and analysis of the rare event $P(W > t)$ can now be reformulated into the analysis of a random walk with *positive* drift, for which the tail event is not rare. In fact, $P_\alpha(W > t) = 1$. Because of the importance of this particular tilted measure, we will denote it $\widetilde{P} = P_\alpha$ when α is fixed. Letting $\tau(t) = \inf\{n \geq 0 : S_n > t\}$, we note that $\{\sup_{n\geq 0} S_n > t\} = \{\tau(t) < \infty\}$. Theorem 1.1.4(ii) then allows us to write

$$P(W > t) = P(\tau(t) < \infty) = \widetilde{E}\left[1(\tau(t) < \infty)e^{-\alpha S_{\tau(t)}}\right].$$

Furthermore, the law of large numbers and (1.1.6) imply that $\lim_{n\to\infty} n^{-1}S_n > 0$ $\widetilde{P}$-a.s. and hence $\widetilde{P}(\tau(t) < \infty) = \widetilde{P}(\sup_{n\geq 0} S_n > t) = 1$. Hence,

$$P(W > t) = \widetilde{E}\left[e^{-\alpha S_{\tau(t)}}\right]. \tag{1.1.7}$$

This is the starting point for the following theorem. Again, see the appendix for its proof.

Theorem 1.1.5 (Cramér-Lundberg Asymptotic)**.** *Let W be the (unique) solution to* (1.1.2) *given as the maximum of a random walk $\{S_n : n \geq 0\}$ starting from* 0 *with nonlattice increment distribution X with $E[X] < 0$. If there exists $\alpha > 0$ such that $m(\alpha) = 1$ and $m'(\alpha) > 0$, then*

$$P(W > t) \sim \frac{\widetilde{E}\left[1 - e^{-\alpha S_{\tau_+}}\right]}{\alpha m'(\alpha)\widetilde{E}\left[\tau_+\right]} e^{-\alpha t} \qquad \text{as } t \to \infty.$$

where $\tau_+ = \inf\{n \geq 0 : S_n > 0\}$ and $\widetilde{E}$ denotes expectation wth respect to $\widetilde{P}$.

Often the asymptotic approximation above is established by noting that the *overshoot* $B(t) = S_{\tau(t)} - t$ has a finite weak limit point

$$B(t) \Rightarrow B(\infty) \tag{1.1.8}$$

as $t \to \infty$ (see for example [7, Theorem VIII.2.1]), and since $B(t) \geq 0$, therefore $e^{\alpha t}P(W > t) = \widetilde{E}\left[e^{-\alpha B(t)}\right] \to \widetilde{E}\left[e^{-\alpha B(\infty)}\right]$, giving an alternate description of the asymptotic constant.

More generally, Goldie [49] used *implicit* renewal theory to study the tail behavior of solutions to a variety of SFPEs, including the Lindley equation generalization

$$W \stackrel{\mathcal{D}}{=} \xi \vee (W + X),$$

where W is independent of (ξ, X), which has arbitrary joint distribution. Here, the solution W is given as the supremum of a *perturbed* random walk $S_n + \xi_n = \sum_{i=1}^n X_i + \xi_n$, where $\{(\xi_{i-1}, X_i) : i \geq 1\}$ are i.i.d. copies of (ξ, X). When $E\left[e^{\alpha\xi}\right] < \infty$ and X is nonlattice and satisfies the Cramér condition for $\alpha > 0$,

$$P(W > t) \sim \frac{E\left[e^{\alpha\xi} \vee e^{\alpha(W+X)} - e^{\alpha(W+X)}\right]}{\alpha E\left[X e^{\alpha X}\right]} e^{-\alpha t} \qquad \text{as } t \to \infty, \tag{1.1.9}$$

where in the expectation, W is independent of (ξ, X). Taking $\xi \equiv 0$ gives an alternate expression for the asymptotic constant for the random walk—one that is "implicit" in the sense that it is computed in terms of the distribution of W itself.

1.1.3 Siegmund's Algorithm

The naive Monte Carlo approach to estimating probabilities is by relative frequencies: an estimate of $p = P(A)$ is given as the average of n independent realizations of the indicator random variable $1(A)$. Under P, each single estimate has a Bernoulli distribution with $P(1(A) = 1) = p$ and

$$\operatorname{Var}(1(A)) = p(1-p).$$

This procedure is not ideal when the event A is rare (i.e. p is small). Indeed, the number of realizations needed until the first occurance of $\{1(A) = 1\}$ has a Geometric(p) distribution, and hence one expects that the very large number of p^{-1} iterations are needed to obtain a nonzero estimate. Additionally, since an accurate estimate of a small probability is desired, the relevant measure of accuracy is *relative error*. However,

$$\frac{\operatorname{Var}(1(A))}{E[1(A)]^2} = \frac{p(1-p)}{p^2} = \frac{1}{p} - 1 \to \infty$$

as $p \to 0$. This approach to estimating p may require a prohibitively large number of iterations n, a number which depends on how small p is, which is generally unknown a priori.

If there is some probability measure Q and a likelihood ratio L such that $P(A) = \int_A L\, dQ$, then *importance sampling* is a technique that can be used to skirt these issues. Motivated by estimating error probabilities in sequential hypothesis testing, Siegmund [92] developed an importance sampler for tail events of the maximum of a random walk based on the representation (1.1.7), namely

$$P(W > t) = P(\tau(t) < \infty) = \int_{\{\tau(t) < \infty\}} e^{-\alpha S_{\tau(t)}}\, d\widetilde{P} = \widetilde{E}\left[e^{-\alpha S_{\tau(t)}}\right].$$

This suggests performing Monte Carlo simulation using the likelihood ratio estimator $Z(t) = e^{-\alpha S_{\tau(t)}}$ sampled under $\widetilde{P}$. A single copy of $Z(t)$ is generated by simulating the positive drift random walk $\{S_n\}$ until the first n such that $S_n > t$, then setting $Z(t) = e^{-\alpha S_n}$. Note that on each iteration, one is guaranteed to see a nonzero estimate. Furthermore, using Theorem 1.1.5,

$$\frac{\widetilde{\mathrm{Var}}(Z(t))}{P(W>t)^2} \leq \frac{\widetilde{E}\left[e^{-2\alpha S_{\tau(t)}}\right]}{P(W>t)^2} \sim \frac{\widetilde{E}\left[e^{-2\alpha(S_{\tau(t)}-t)}\right]}{\widetilde{E}\left[e^{-\alpha B(\infty)}\right]^2} \to \frac{\widetilde{E}\left[e^{-2\alpha B(\infty)}\right]}{\widetilde{E}\left[e^{-\alpha B(\infty)}\right]^2} < \infty$$

as $t \to \infty$, where $\widetilde{\mathrm{Var}}$ denotes variance with respect to $\widetilde{P}$ and $B(\infty)$ is the limiting overshoot defined in (1.1.8). Hence the estimator $Z(t)$ has *bounded relative error*, ensuring that a smaller number of iterations is suitable in the Monte Carlo procedure.

1.1.4 Branching Recursions

Chapters 2 and 3 concern the tail behavior of solutions to the branching generalization of (1.1.2), known as the high order Lindley equation:

$$W \stackrel{\mathcal{D}}{=} \max\left\{Y, \max_{1\leq i\leq N}(X_i + W_i)\right\}, \tag{1.1.10}$$

where $\{W_i : i \geq 1\}$ are independent copies of W independent of the vector $(Y, N, X_1, X_2, \ldots)$, which takes values in $\mathbb{R} \times \mathbb{N} \times \mathbb{R}^\infty$. Note that when we replace Y with 0 and N with 1 in (1.1.10), we obtain the Lindley equation (1.1.2). Although Lindley's equation has a unique solution whenever $E[X_1] < 0$, there is no uniqueness in the branching case, as shown in [68]. As the work in [12] shows, the solutions to (1.1.10) can be constructed using one special solution, known as the endogenous solution [3]. The endogenous solution can be explicitly constructed on a structure known as a weighted branching process [11, 90, 62], and other solutions can be obtained by adding different "terminal" values to the leaves of a finite tree (see [12] and Section 2.1 for more details). From an applications point of view (e.g., the models in [68, 87]), it is usually the special endogenous solution that is of interest. If $W = \log R$, $X_i = \log C_i$, $Y = \log Q$, equation (2.0.1) is equivalent to the random extremal equation

$$R \stackrel{\mathcal{D}}{=} Q \vee \left(\bigvee_{i=1}^{N} C_i R_i\right), \tag{1.1.11}$$

where the $\{R_i\}$ are i.i.d. copies of R, independent of the vector $(Q, N, \{C_i\})$, where $N \in \mathbb{N} \cup \{\infty\}$, $Q, \{C_i\} \geq 0$ and $P(Q > 0) > 0$. These types of distributional equations and their simulation have received considerable attention in the recent literature, although most of it has centered around the affine version of the equation considered here; see [23, 27, 28, 40, 49, 55, 62, 63, 64]. We refer to the overview on this topic given in [31, 61].

To define the endogenous solution, we adopt the notation from [63] to define a marked Galton-Watson process. To this end, let $\mathbb{N}_+ = \{1, 2, 3, \dots\}$ be the set of positive integers and let $U = \bigcup_{k=0}^{\infty} (\mathbb{N}_+)^k$ be the set of all finite sequences $\boldsymbol{i} = (i_1, i_2, \dots, i_n)$, where by convention $\mathbb{N}_+^0 = \{\emptyset\}$ contains the null sequence $\emptyset$. To ease the exposition, for a sequence $\boldsymbol{i} = (i_1, i_2, \dots, i_k) \in U$ we write

$$\boldsymbol{i}|n = (i_1, i_2, \dots, i_n),$$

provided $k \geq n$, and $\boldsymbol{i}|0 = \emptyset$ to denote the index truncation at level n, $n \geq 0$. Also, for $\boldsymbol{i} \in A_1$ we simply use the notation $\boldsymbol{i} = i_1$, skipping the parenthesis. Similarly, for $\boldsymbol{i} = (i_1, \dots, i_n)$ we will use

$$(\boldsymbol{i}, j) = (i_1, \dots, i_n, j)$$

to denote the index concatenation operation, and if $\boldsymbol{i} = \emptyset$, then write $(\boldsymbol{i}, j) = j$. Let $|\boldsymbol{i}|$ be the length of index $\boldsymbol{i}$, i.e., $|\boldsymbol{i}| = k$ if $\boldsymbol{i} = (i_1, \dots, i_k) \in \mathbb{N}_+^k$. We order U according to a length-lexicographic order $\prec$: $\boldsymbol{i} \prec \boldsymbol{j}$ if either $|\boldsymbol{i}| < |\boldsymbol{j}|$, or $|\boldsymbol{i}| = |\boldsymbol{j}|$ and $i_r = j_r$ for $r = 1, \dots, t-1$, and $i_t < j_t$ for some $t \leq |\boldsymbol{i}|$.

To iteratively construct the weighted branching tree $\mathcal{T}$, let $\{\boldsymbol{\psi}_{\boldsymbol{i}}\}_{\boldsymbol{i} \in U}$ denote a sequence of i.i.d. random elements in $\mathbb{N} \times \mathbb{R}^\infty$, where $\boldsymbol{\psi}_{\boldsymbol{i}} = (N_{\boldsymbol{i}}, Q_{\boldsymbol{i}}, C_{(\boldsymbol{i},1)}, C_{(\boldsymbol{i},2)}, \dots)$. For simplicity we denote $\boldsymbol{\psi} = (N, Q, C_1, C_2, \dots) = \boldsymbol{\psi}_\emptyset$ to represent a generic element of the sequence $\{\boldsymbol{\psi}_{\boldsymbol{i}}\}$. The random integers $\{N_{\boldsymbol{i}}\}_{\boldsymbol{i} \in U}$ herein define the structure of the tree as follows. Let $A_0 = \{\emptyset\}$,

$$\begin{aligned} A_1 &= \{i \in \mathbb{N} : 1 \leq i \leq N_\emptyset\}, \quad \text{and} \\ A_n &= \{(\boldsymbol{i}, i_n) \in U : \boldsymbol{i} \in A_{n-1}, 1 \leq i_n \leq N_{\boldsymbol{i}}\}, \quad n \geq 2, \end{aligned} \tag{1.1.12}$$

be the set of individuals in the nth generation. Thus to each node $\boldsymbol{i}$ in the tree different from the root we assign the weight $C_{\boldsymbol{i}}$, and a cumulative weight $\Pi_{\boldsymbol{i}}$ computed according to

$$\Pi_{i_1} = C_{i_1}, \qquad \Pi_{(i_1,\dots,i_n)} = C_{(i_1,\dots,i_n)} \Pi_{(i_1,\dots,i_{n-1})}, \qquad n \geq 2,$$

where $\Pi = \Pi_\emptyset \equiv 1$ is the cumulative weight of the root node. See Figure 1.2.

The random variable

$$R = \bigvee_{\boldsymbol{i} \in \mathcal{T}} \Pi_{\boldsymbol{i}} Q_{\boldsymbol{i}} \tag{1.1.13}$$

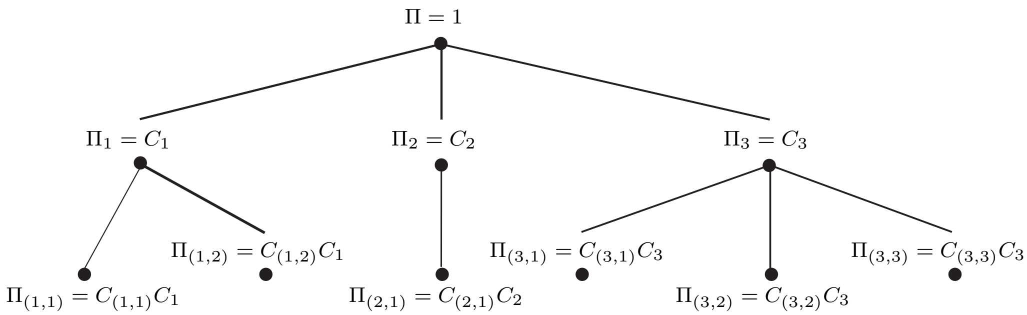


Figure 1.2: A weighted branching tree

is known as the special endogenous solution to (1.1.11). It is a relatively straight-forward exercise to check that indeed this R satisfies (1.1.11). Equivalently, the endogenous solution to the high order Lindley equation (1.1.10) is given by

$$W = \log R = \bigvee_{\boldsymbol{i} \in \mathcal{T}} (S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}), \tag{1.1.14}$$

where $Y_{\boldsymbol{i}} = \log Q_{\boldsymbol{i}}$ and

$$S_{\boldsymbol{i}} = \log \Pi_{\boldsymbol{i}} = \sum_{j=1}^{|\boldsymbol{i}|} X_{\boldsymbol{i}|j},$$

with $X_{\boldsymbol{i}} = \log C_{\boldsymbol{i}}$ for each $\boldsymbol{i} \in \mathcal{T}$. The above display is recognizable as a branching random walk, i.e. the endogenous solution to the high order Lindley equation is given as the maximum of a branching random walk with a pertubation described by Y. As mentioned earlier, the high-order Lindley equation has in general multiple solutions, but before we discuss those it is convenient to focus first on the so-called *regular* case, which corresponds to the existence of a unique $\alpha > 0$ satisfying

$$E\left[\sum_{j=1}^{N} C_j^{\alpha}\right] = E\left[\sum_{j=1}^{N} e^{\alpha X_j}\right] = 1 \quad \text{and} \quad E\left[\sum_{j=1}^{N} C_j^{\alpha} \log C_j\right] = E\left[\sum_{j=1}^{N} X_j e^{\alpha X_j}\right] \in (0, \infty). \tag{1.1.15}$$

As the work in [12] shows, other solutions to (1.1.11) can be constructed by using “terminal" values. More precisely, consider the finite tree $\mathcal{T}^{(n)} = \{\boldsymbol{i} \in \mathcal{T} : |\boldsymbol{i}| \leq n\}$, and construct the random variable

$$R_n(B) = \left(\bigvee_{\boldsymbol{i} \in \mathcal{T}^{(n-1)}} \Pi_{\boldsymbol{i}} Q_{\boldsymbol{i}}\right) \vee \left(\bigvee_{\boldsymbol{i} \in A_n} \Pi_{\boldsymbol{i}} B_{\boldsymbol{i}}\right),$$

where the $\{B_{\boldsymbol{i}}\}$ are i.i.d. nonnegative random variables, independent of all other branching vectors in $\mathcal{T}^{(n-1)}$. Then, provided

$$\lim_{x \to \infty} x^{\alpha} P(B > x) = \gamma \geq 0,$$

the random variable $R(B) = \lim_{n\to\infty} R_n(B)$ is a solution to (2.0.3) (see Theorem 1(ii) in [12]). Note that the special endogenous solution R given by (2.0.4) corresponds to taking the terminal values $\{B_{\boldsymbol{i}}\}$ identically equal to zero, and is known to be the minimal solution in the usual stochastic order sense (see Proposition 5 in [12]). Moreover, by Theorem 1(i) in [12], R is finite a.s. whenever

$$\sup_{x\geq 1} x^{\alpha}(\log x)^{1+\epsilon} P(Q > x) < \infty$$

for some $\epsilon > 0$.

Besides observing that in applications [68, 87] it is usually the special endogenous solution that is of interest, it is worth mentioning that it plays an important role in characterizing all the solutions defined through $R(B)$, whose distributions are given by

$$P(R(B) \leq x) = E\left[1(R \leq x)\exp(-\gamma W(\alpha)x^{-\alpha})\right], \tag{1.1.16}$$

where $W(\alpha)$ is the a.s. limit of the martingale $W_n(\alpha) := \sum_{\boldsymbol{i}\in A_n} \Pi_{\boldsymbol{i}}^{\alpha}$ (see Theorem 1(ii) in [12]). The martingale $\{W_n(\theta) : n \geq 1\}$ defined via

$$W_n(\theta) := \rho_\theta^{-n} \sum_{\boldsymbol{i}\in A_n} \Pi_{\boldsymbol{i}}^{\theta}, \tag{1.1.17}$$

where $\rho_\theta := E\left[\sum_{j=1}^{N} C_j^{\theta}\right]$, is known as the Biggins-Kyprianou martingale [11, 72], and it plays an important role in much of the weighted branching processes literature. Moreover, under additional technical conditions, all the solutions to (1.1.11) can be characterized through (1.1.16) (see Theorem 1(iii) in [12]).

As an example illustrating the multiplicity of solutions to (1.1.11), consider the case when $N \equiv 2$, $C_i \equiv \frac{1}{2}$ for $i = 1, 2$ and $Q \equiv \frac{1}{2}$, whose endogenous solution is given by

$$R = \bigvee_{\boldsymbol{i}\in\mathcal{T}} \Pi_{\boldsymbol{i}} Q_{\boldsymbol{i}} = \bigvee_{n=0}^{\infty} \left(\frac{1}{2}\right)^{n+1} = \frac{1}{2}.$$

Now note that if $R' = (T \vee 1)/2$ where T has a Frechet distribution with shape/scale parameters $(1, s)$, i.e., $P(T \leq x) = e^{-s/x}$ for $x > 0$, then R' is a (non-endogenous) solution since

$$Q \vee \bigvee_{i=1}^{N} C_i R_i' = \frac{1}{2} \vee \bigvee_{i=1}^{2} \frac{1}{2} \cdot \frac{(T_i \vee 1)}{2} = \frac{1}{2}\max\left\{1, \frac{T_1 \vee T_2}{2}\right\} \stackrel{\mathcal{D}}{=} \frac{1}{2}(1 \vee T) = R'.$$

Furthermore, by setting $B = T/2$ we can identify R' with

$$R(B) = \lim_{n\to\infty} R_n(B) = \lim_{n\to\infty} \frac{1}{2} \vee \left(\bigvee_{\boldsymbol{i}\in A_n} \frac{B_{\boldsymbol{i}}}{2^n}\right) = \lim_{n\to\infty} \frac{1}{2}\max\left\{1, \bigvee_{\boldsymbol{i}\in A_n} \frac{T_{\boldsymbol{i}}}{2^n}\right\} \stackrel{\mathcal{D}}{=} \frac{1}{2}(1 \vee T) = R'.$$

The condition (1.1.15) is the branching version of the Cramér condition on the moment generating function discussed in Section 1.1.2. Under this condition (among others), a tail asymptotic analogous to the Cramér-Lundberg asymptotic was established in [62, 64] for the endogenous solution to (1.1.10). In particular, the methods of Goldie [49] we extended to give the following generalization of (1.1.9).

Theorem 1.1.6. *Let W be the endogenous solution to* (1.1.10) *given by* (1.1.14) *for some* $(Y, N, \{X_i\})$ *with* $N \in \mathbb{N} \cup \{\infty\}$*,* $X_i \in \mathbb{R}$ *for each* i *and* $P(Y > -\infty) > 0$*. Suppose there exists* $j \geq 1$ *with* $P(N \geq j) > 0$ *such that the measure* $P(X_j \in dx, N \geq j)$ *is nonlattice, and that* (1.1.15) *holds for some* $\alpha > 0$ *such that*

(i) $E\left[e^{\alpha Y}\right] < \infty$*, and*

(ii) $E\left[\left(\sum_{i=1}^N e^{X_i}\right)^\alpha\right] < \infty$ *if* $\alpha > 1$ *or* $E\left[\left(\sum_{i=1}^N e^{\alpha X_i/(1+\varepsilon)}\right)^{1+\varepsilon}\right] < \infty$ *for some* $0 < \varepsilon < 1$ *if* $\alpha \leq 1$.

Then,

$$P(W > t) \sim He^{-\alpha t} \qquad \text{as } \ t \to \infty,$$

where $0 < H < \infty$ *is given by*

$$H = \frac{E\left[e^{\alpha Y} \vee \bigvee_{i=1}^N e^{\alpha(X_i + W_i)} - \sum_{i=1}^N e^{\alpha(X_i + W_i)}\right]}{\alpha E\left[\sum_{i=1}^N X_i e^{\alpha X_i}\right]},$$

where in the expectation in the numerator, $\{W_i\}$ *are independent copies of* W *independent of* $(Y, N, \{X_i\})$*.*

If we consider the Laplace transform

$$\rho_\theta = E\left[\sum_{i=1}^N e^{\theta X_i}\right]$$

as a generalized version of the moment generating function in the nonbranching case, then the condition (1.1.15) says that for some $\alpha > 0$, $\rho_\alpha = 1$ and $\frac{d}{d\theta}\rho_\theta\big|_{\theta=\alpha} > 0$. Note that unlike with a moment generating function, $\rho_0 = E[N] \neq 1$ in general. By convexity of ρ, the condition then says that $\rho_\theta = 1$ for exactly two values of θ, the larger of which is the α in the theorem. See Figure 1.3 for an illustration of a typical situation. In particular, there is some $0 < \beta < \alpha$ for which $\rho_\beta < 1$, which in turn implies that

$$E\left[\max_{1 \leq i \leq N} X_i\right] = \frac{1}{\beta} E\left[\log\left(\bigvee_{i=1}^N e^{\beta X_i}\right)\right] \leq \frac{1}{\beta} \log E\left[\sum_{i=1}^N e^{\beta X_i}\right] < 0 \tag{1.1.18}$$

by Jensen's inequality. Hence, W is the maximum of a (perturbed) branching random walk with negative drift on each of its branches, which explains why the tail events $\{W > t\}$ are rare.

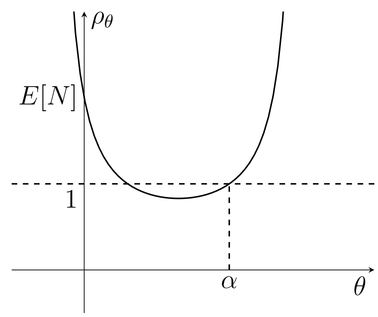


Figure 1.3: The relevant situation for the branching Laplace transform

1.2 Large Deviations For Interacting Particle Models

The models studied in Chapters 4 and 5 are refered to as *interacting particle systems*. Here we discuss some foundational asymptotic results about such systems and motivate the extensions of these basic models to those studied in subsequent chapters.

1.2.1 Large Deviation and Laplace Principles

In this section we provide some basic definitions and examples in the theory of large deviations. In Chapters 4 and 5, we prove large deviations principles for certain classes of interacting particle systems via weak convergence methods, and here we motivate the methods therein.

Suppose $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of random variables taking values in a Polish space S. Here and throughout, by a *Polish space*, we mean a complete, separable metric space. We say that a function $I : S \to [0, \infty]$ is a *rate function* if it has compact level sets, that is, $\{x \in S : I(x) \leq c\}$ is compact in S for all $c \in [0, \infty)$. Often in the literature, I is called a rate function if it is lower semi-continuous, i.e. the level sets are closed, and it is called a *good rate function* if furthermore the level sets are compact. We will not be working with rate functions that are not "good," and so we will stick with the former terminology. The following gives the definition of a large deviation principle (LDP).

Definition 1.2.1. *The sequence $\{X_n\}_{n\in\mathbb{N}}$ is said to satisfy a large deviation principle on S with rate function I and speed $a_n \to \infty$ if both of the following hold.*

(a) The large deviation upper bound: for each closed $F \subset S$,

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P(X_n \in F) \leq - \inf_{x\in F} I(x),$$

(b) The large devation lower bound: for each open $G \subset S$,

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P(X_n \in G) \geq -\inf_{x\in G} I(x).$$

Formally, such a result says that, for large n and Borel sets $A \subset S$,

$$P(X_n \in A) \approx e^{-n \inf_{x\in A} I(x)}.$$

That is, the rate function quantifies the exponential rate of decay of the probabilities $P(X_n \in A)$ as $n \to \infty$. A large deviation principle can be seen as a refinement of a law of large numbers (LLN) in the sense that it quantifies the rate of convergence to the law of large numbers limit.

In Chapters 4 and 5, we are interested in LDPs for empirical measure processes, i.e. sequences

$$\mu^n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

where $\{X_i\}$ take values in a space S and δ_x denotes the Dirac measure at x, i.e. $\delta_x(A) = 1(x \in A)$ for measurable sets A. For the simple case of an i.i.d. sequence $\{X_i\}$, the LLN behavior of such objects is governed by the well-known Glivenko-Cantelli lemma. For a Polish space S, let $\mathcal{P}(S)$ denote the space of probability measures on S with the topology of weak convergence.

Theorem 1.2.1 (Glivenko-Cantelli). *If $\{X_n : n \in \mathbb{N}\}$ are i.i.d. random variables taking values in a Polish space S, each with distribution μ, then, as $n \to \infty$,*

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \to \mu \quad a.s. \quad in\ \mathcal{P}(S).$$

Recalling the definition of convergence in $\mathcal{P}(S)$, the above says that

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \int_S f\, d\mu$$

for all continuous and bounded $f : S \to \mathbb{R}$. See [21, Lemma 3.2] for a proof.

In the theory of large deviations, particularly that for empirical measures, one of the basic objects is the relative entropy function. For $\mu \in \mathcal{P}(S)$, the *relative entropy* $R(\cdot\|\mu) : \mathcal{P}(S) \to [0, \infty]$ is given by

$$R(\gamma\|\mu) = \begin{cases} \int_S \log\left(\frac{d\gamma}{d\mu}\right) d\gamma & \text{if } \gamma \ll \mu, \\ \infty & \text{otherwise,} \end{cases}$$

where $\gamma \ll \mu$ denotes absolute continuity of γ with respect to μ, and $\frac{d\gamma}{d\mu}$ denotes the Radon-Nikodym derivative of γ with respect to μ. It is easy to check that, in fact, $R(\gamma\|\mu) \geq 0$ for $\gamma, \mu \in \mathcal{P}(S)$. The following result, known as Sanov's theorem, is one of the fundamental results in large deviations theory.

Theorem 1.2.2 (Sanov)**.** *If $\{X_n : n \in \mathbb{N}\}$ are i.i.d. random variables taking values in a Polish space S, each with distribution μ, then the sequence*

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}, \quad n \in \mathbb{N},$$

satisfies the large deviation principle on $\mathcal{P}(S)$ with rate function $I(\cdot) = R(\cdot\|\mu)$ and speed n.

Chapters 4 and 5 uses weak convergence methods to prove large deviation principles for certain interacting particle models. A reformulation of a large deviation principle, called a Laplace principle, is particularly convenient for the application of weak convergence methods. Let I be a rate function on S. A sequence $\{X_n\}$ taking values in S is said satisfy the *Laplace principle on S with rate function I* if, for all bounded continous functions $F : S \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{a_n} \log E\left[e^{-a_n F(X_n)}\right] = -\inf_{x\in S}\left(F(x) + I(x)\right). \tag{1.2.1}$$

It is well known [98] that for a given rate function I, a sequence $\{X_n\}$ satisfies the Laplace principle with rate function I if and only if it satisfies both the large deviation upper bound and the large deviation lower bound with rate function I, However for completeness we provide a proof in the Appendix.

Theorem 1.2.3. *A sequence $\{X_n\}$ satisfies a large deviation principle on a Polish space (S, d) with rate function I and speed a_n if and only if* (1.2.1) *holds for all $F \in \mathcal{C}_b(S)$.*

1.2.2 Weakly Interacting Diffusions

Here we motivate the models studied in Chapters 4 and 5 by describing some special, simple cases. Consider a particle whose state is subject to some driving force and a random disturbance and can be described by an SDE of the form

$$dX(t) = b(X(t))\,dt + \sigma(X(t))\,dW(t), \quad t \geq 0, \tag{1.2.2}$$

where $X(t)$ indicates the state of the particle at time t and W is a Brownian motion of appropriate dimension. A system of n particles could be described by a system of n equations of the type in (1.2.2), driven by mutually independent noise terms, however to describe real physical systems of multiple particles, the interaction between particles is natural to consider. As noted by Tanaka and Hitsuda in [97], certain second order partial differential equations (PDEs) that describe physical systems, such as Boltzmann's famous equation in the kinetic theory of monatomic gases [66], are related to certain nonlinear Markov processes that are described through the mean field limit of a collection of weakly interacting particles. By the *mean field* or *weak interaction* of n particles whose states are described by $X_1(t), X_2(t), \ldots, X_n(t)$, we mean an interaction that is given through the empirical measure process

$$\mu^n(t) \doteq \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(t)}, \quad t \geq 0, \tag{1.2.3}$$

where δ_x is the Dirac measure on the appropriate measure space. For a measurable set A, $\mu^n(t, A)$ is then the proportion of particles that are in a particular state described by A at time t. The contribution of each particle to the sum in (1.2.3) diminishes as n becomes large, and it is in this sense that the interaction is considered "weak."

When considering a weakly interacting particle system, one is usually interested in the asymptotics of $\mu^n(t)$ as $n \to \infty$. In the case of n particles each described by (1.2.2) with mutually independent Brownian motions $W_1, \ldots, W_n$, the states of the particles are independent, and the LLN asymptotics of $\mu^n(t)$ are governed by the Glivenko-Cantelli Lemma given in the previous section. The rate of convergence, as given by the corresponding LDP, is described by Sanov's theorem.

Extending (1.2.2) to include a weak interaction leads to the system of SDEs

$$dX_i(t) = b(X_i(t), \mu^n(t))\, dt + \sigma(X_i(t), \mu^n(t))\, dW_i(t), \quad t \geq 0, \quad 1 \leq i \leq n, \tag{1.2.4}$$

where $\{W_i\}$ are mutually independent Brownian motions. That is, the dynamics of each particle at time t depends on the state of that particle and the states of the other particles through $\mu^n(t)$. Models like these (as well as those studied in Chapters 4 and 5 are often referred to as *weakly interacting particle systems* and have been extensively studied, see [82, 18, 32, 33, 96, 85, 94, 51, 91, 83] and many others. Originally motivated by problems in statistical physics, in recent years such systems have arisen in many applied probability problems such as stochastic networks [5, 52], information theory [15, 16], mathematical neuroscience [10], population opinion dynamics [50], nonlinear filtering [71, 34], and mathematical finance [46, 47], among others.

As an example, we consider the model in [33] on $\mathbb{R}^d$:

$$dX_i(t) = b(X_i(t), \mu^n(t))\, dt + \sigma(X_i(t))\, dW_i(t), \quad t \geq 0, \quad 1 \leq i \leq n, \tag{1.2.5}$$

where $\mu^n(t)$ is the empirical measure of $X_1(t), \ldots, X_n(t)$ and $\{W_i : i \in \mathbb{N}\}$ are mutually independent Brownian motions on $\mathbb{R}^d$. Under the condition that $\mu^n(0) \to \mu_0$ in $\mathcal{P}(\mathbb{R}^d)$ for some μ_0, there is a deterministic $\mathcal{P}(\mathbb{R}^d)$-valued process $\mu(\cdot)$ such that

$$\mu^n(\cdot) \Rightarrow \mu(\cdot),$$

and this limit can be characterized as a weak solution to the *Mckean-Vlasov equation*

$$\frac{d}{dt}\mu(t) = \mathcal{L}^*(\mu(t))\mu(t), \quad \mu(0) = \mu_0. \tag{1.2.6}$$

where $\mathcal{L}^*$ denotes the formal adjoint of the operator

$$\mathcal{L}(\mu)f(x) = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}f(x) + \sum_{i=1}^{d} b_i(x,\mu)\frac{\partial}{\partial x_i}f(x), \quad a(x) = \sigma(x)\sigma(x)^{\mathsf{T}},$$

for smooth $f : \mathbb{R}^d \to \mathbb{R}$. The PDE (1.2.6) is interpreted in the weak sense. That is, for smooth test functions f with compact support,

$$\int_0^t \langle f, \mu(s)\rangle \, ds = \langle f, \mu_0\rangle + \int_0^t \langle \mathcal{L}(\mu(s))f, \mu(s)\rangle \, ds,$$

where $\langle f, \mu\rangle$ denotes $\int f \, d\mu$.

In Chapter 4, we consider the case where the system is driven by a common noise. Consider for simplicity the following elementary model where the limiting dynamics are described by *Freidlin-Wentzell asymptotics*: let $\{X^\varepsilon : \varepsilon > 0\}$ be the family of processes on a Polish space S (typically $\mathbb{R}^d$ for some d) described by

$$dX^\varepsilon(t) = b(X^\varepsilon(t))\, dt + \sqrt{\varepsilon}\, dW(t), \qquad t \geq 0. \tag{1.2.7}$$

We are interested in the asymptotics as $\varepsilon \to 0$, and for this reason we say the noise is "small." The rate at which the size of the noise decreases is governed by $\sqrt{\varepsilon}$, however we note that as long as the size goes to zero, the LLN limit is not changed by different rates. Indeed, suppose $X^\varepsilon(0) = x_0 \in S$ for each $\varepsilon > 0$. Then rewriting (1.2.7) as

$$X^\varepsilon(t) = x_0 + \int_0^t b(X^\varepsilon(s))\, ds + \sqrt{\varepsilon} W(t),$$

one can show that if b satisfies appropriate conditions (e.g. it is a Lipschitz map), then $\{X^\varepsilon : \varepsilon > 0\}$ is tight, and along every subsequence, $X^\varepsilon \Rightarrow x$, where x is the unique solution to the ordinary differential equation

$$\frac{d}{dt}x(t) = b(x(t)), \quad x(0) = x_0. \tag{1.2.8}$$

While the $\sqrt{\varepsilon}$ rate does not affect the law of large numbers limit, one expects that the rate at which the driving noise gets small changes the rate at which that limit is reached, which can be made precise by establishing a large deviation principle for $\{X^\varepsilon\}$.

To illustrate what a large deviation principle looks like when there is interaction in the system through a small noise, let $\mathcal{C}([0,T],\mathbb{R}^d)$ denote the space of continuous functions from $[0,T]\to\mathbb{R}^d$ for some $T,d<\infty$ equipped with the supremum norm. It is well known that the sequence $\{X^\varepsilon\}$ satisfies the large deviation principle (as $\varepsilon\to 0$) on $\mathcal{C}([0,T],\mathbb{R}^d)$ with speed ε^{-1} rate function

$$I(\varphi)=\frac{1}{2}\int_0^T\left\|\frac{d}{dt}\varphi(t)-b(\varphi(t))\right\|^2\,dt.$$

Consider now the setting where instead of a single SDE one has a system of n equations as in (1.2.5) where each equation has in addition a common small Brownian motion, i.e. a term of the form $\kappa(n)dB(t)$ where B is Brownain motion and $\kappa(n)\to 0$ as $n\to\infty$. Once again one expects that the law of large number behavior of the associated empirical measure does not depend on the manner in which $\kappa(n)\to 0$, but that the rate at which this convergence occurs, as quantified by a large deviation principle, does depend crucially on the properties of κ. Chapter 4 addresses such questions by establishing large deviation principles for a broad family of weakly interacting diffusions with small common noise under various types of conditions on κ.

1.2.3 Control Representations

The starting point for the work in Chapters 4 and 5 are certain variational representations for exponential functionals of Brownian motion established in [17]. The above models are driven by a finite-dimensional Brownian motion $\{W(t):0\le t\le T\}$, and so, with proving a Laplace principle in mind, we are interested in a representation of the quantity

$$-\log E\left[e^{-f(W)}\right],$$

where $f:\mathcal{C}([0,T],\mathbb{R}^d)\to\mathbb{R}$ is a bounded and measurable function. Suppose that $\{\mathcal{F}(t):0\le t\le T\}$ is a usual filtration on a space $(\Omega,\mathcal{F},P)$ (namely it is right continuous and has all null sets) relative to which W is s d-dimensional Brownian motion on $[0,T]$. A process $\{X(t):0\le t\le T\}$ is said to be *progressively measurable* with respect to $\mathcal{F}(t)$ if for each t, the map from $([0,t]\times\Omega,\mathcal{B}([0,t])\otimes\mathcal{F}(t))$ to $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$ given by

$$(s,\omega)\mapsto X(s,\omega)$$

is measurable. Denote by $\mathcal{A}_d$ the set of all $\mathcal{F}(t)$-progressively measurable $\mathbb{R}^d$-valued processes $\{v(t):0\le t\le T\}$ such that

$$E\left[\int_0^T\|v(s)\|^2\,ds\right]<\infty.$$

We refer to $\mathcal{A}_d$ as a space of *controls*. We then have the following (see [17, 20]).

Theorem 1.2.4. *If $f : \mathcal{C}([0,T], \mathbb{R}^d) \to \mathbb{R}$ is a bounded and Borel-measurable function, then*

$$-\log E\left[e^{-f(W)}\right] = \inf_{v \in \mathcal{A}_d} E\left[\frac{1}{2}\int_0^T \|v(s)\|^2\, ds + f\left(W + \int_0^\cdot v(s)\, ds\right)\right].$$

To illustrate the usefulness of the above theorem, consider the generalization of the model in (1.2.5) analyzed in [22], where for each $n \in \mathbb{N}$, the particles $(X_1^n, X_2^n, \ldots, X_n^n)$ are are driven by the independent Brownian motions $\{W_i\}$ on $\mathbb{R}^d$:

$$dX_i^n(t) = b(X_i^n(t), \mu^n(t))\, dt + \sigma(X_i^n(t), \mu^n(t))\, dW_i(t), \quad t \in [0,T], \quad 1 \le i \le n.$$

Suppose for each $n \in \mathbb{N}$, existence and uniqueness holds in the strong sense for the above model. Then there is some Borel-measurable map $h^n = (h_1^n, \ldots, h_n^n)$ such that for each $1 \le i \le n$, a solution is given by

$$X_i^n(\cdot, \omega) = h_i^n(W(\cdot, \omega)) \quad \text{for a.e. } \omega,$$

where $W = (W_1, \ldots, W_n)$. Hence, if $F : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ is bounded and continuous, and

$$f^n(\cdot) = F\left(\frac{1}{n}\sum_{i=1}^n \delta_{h_i^n(\cdot)}\right),$$

then $(nf^n) : \mathcal{C}([0,T], \mathbb{R}^d) \to \mathbb{R}$ is bounded and Borel-measurable, and

$$-\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] = -\frac{1}{n}\log E\left[e^{-nf^n(W)}\right].$$

The limit in n of the left hand side above is what we want to analyze to establish a Laplace principle for μ^n, and the right hand side is the subject of Theorem 1.2.4. Then under some change of measure, an application of Girsanov's theorem says further that if $u_i \in \mathcal{A}_d$, $1 \le i \le n$,

$$W_i + \int_0^\cdot u_i(s)\, ds, \quad 1 \le i \le n,$$

is a d-dimensional Brownian motion. Hence the *controlled* system

$$d\bar{X}_i^n(t) = b(\bar{X}_i^n(t), \bar{\mu}^n(t))\, dt + \sigma(\bar{X}_i^n(t), \bar{\mu}^n(t))\, dW_i(t) + \sigma(\bar{X}_i^n(t), \bar{\mu}^n(t)) u_i(t)\, dt, \quad t \in [0,T], \quad 1 \le i \le n,$$

where $\bar{\mu}^n(t)$ denotes the empirical measure of $(\bar{X}_1^n(t), \ldots, \bar{X}_n^n(t))$, has a unique strong solution that satisfies

$$\bar{X}_i^n(\cdot, \omega) = h_i^n\left(W(\cdot, \omega) + \int_0^{\cdot} u(s, \omega)\, ds\right) \quad \text{for a.e. } \omega,$$

where $u = (u_1, \ldots, u_n) \in \mathcal{A}_{nd}$. Furthermore, for F and f^n as before,

$$E\left[f^n\left(W + \int_0^{\cdot} u(s)\, ds\right)\right] = E\left[F(\bar{\mu}^n)\right],$$

and so, applying Theorem 1.2.4,

$$-\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] = \inf_{u\in\mathcal{A}_{nd}} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i(s)\|^2\, ds + F(\bar{\mu}^n)\right].$$

Now a Laplace principle can be obtained by analyzing the weak limit points of the processes on the right hand side of the above display. Such a representation is the basis of the proofs in Chapters 4 and 5.

Large deviation proof methods based on control representations naturally result in corresponding control representations for the large deviation rate functions. In order to motivate the representations of the rate functions that show up in Chapters 4 and 5, we begin with the following elementary setting. For fixed $x \in \mathbb{R}^m$ consider the empirical measure

$$\mu^n = \frac{1}{n}\sum_{i=1}^n \delta_{\{x+W_i\}}$$

for independent m-dimensional Brownian motions $\{W_i\}$ on the time interval $[0, T]$. By Sanov's theorem, $\{\mu^n\}$ satisfies a large deviation principle on $\mathcal{P}(\mathcal{C}([0,T], \mathbb{R}^m))$ with rate function I given as

$$I(\gamma) = R(\gamma\|\theta_x), \qquad \gamma \in \mathcal{P}(\mathcal{C}([0,T], \mathbb{R}^m)), \tag{1.2.9}$$

where $\mathcal{C}([0,T], \mathbb{R}^m)$ is the space of continuous functions from $[0, T]$ to $\mathbb{R}^m$ equipped with the uniform topology, θ_x denotes the Wiener measure with initial value x, and the quantity $R(\gamma\|\theta_x)$ denotes the relative entropy of γ with respect to θ_x, defined in Section 1.2.1

The corresponding control representation for the rate function, although notationally more demanding, is useful when studying more general settings. For a Polish space S, we will denote by $\mathcal{B}(S)$ the Borel σ-field on S. Let $\mathcal{R}$ denote the set of all finite measures r on $\mathcal{B}(\mathbb{R}^m \times [0, T])$ such that $r(\mathbb{R}^m \times [0, t]) = t$ for all $t \in [0, T]$. This space is equipped with the topology of weak convergence. Let $\mathcal{R}_1 \subset \mathcal{R}$ be defined as

$$\mathcal{R}_1 \doteq \left\{r \in \mathcal{R} : \int_{\mathbb{R}^m\times[0,T]} \|y\|\, r(dy, dt) < \infty\right\}.$$

Then $\mathcal{R}_1$ is a Polish space when equipped with the the Wasserstein-1 metric. Under this metric, $r_n \to r$ in $\mathcal{R}_1$ if and only if $r_n \to r$ as a sequence in $\mathcal{R}$ and $\int_{\mathbb{R}^m \times [0,T]} y\, r_n(dy, dt) \to \int_{\mathbb{R}^m \times [0,T]} y\, r(dy, dt)$. Let

$$\mathcal{Z}_1 \doteq \mathcal{X} \times \mathcal{R}_1 \times \mathcal{W}, \qquad \text{where } \mathcal{X} = \mathcal{W} = \mathcal{C}([0,T], \mathbb{R}^m),$$

and denote by (X, ρ, W) the three coordinate maps on this space, namely

$$X(\xi, r, w) = \xi, \quad \rho(\xi, r, w) = r, \quad W(\xi, r, w) = w, \quad (\xi, r, w) \in \mathcal{Z}_1.$$

Define

$$\mathcal{P}_2(\mathcal{Z}_1) \doteq \left\{ \Theta \in \mathcal{P}(\mathcal{Z}_1) : E_\Theta \left[\int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \rho(dy, dt) \right] < \infty \right\},$$

where E_Θ denotes expectation on $(\mathcal{Z}_1, \mathcal{B}(\mathcal{Z}_1), \Theta)$. Let $\mathcal{E}_1$ denote the subset of $\mathcal{P}_2(\mathcal{Z}_1)$ consisting of probability measures Θ such that, under Θ, $W(t)$ is a standard Brownian motion with respect to the canonical filtration $\mathcal{F}_t \doteq \sigma\{X(s), W(s), \rho(A \times [0,s]); A \in \mathcal{B}(\mathbb{R}^m), s \leq t\}$, and a.s.

$$X(t) = x + \int_{\mathbb{R}^m \times [0,t]} y\, \rho(dy, ds) + W(t), \qquad 0 \leq t \leq T. \tag{1.2.10}$$

Then the rate function I in (1.2.9) has the following alternative representation:

$$I(\gamma) = \inf_{\Theta \in \mathcal{E}_1 : [\Theta]_1 = \gamma} E_\Theta \left[\frac{1}{2} \int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \rho(dy, ds) \right], \tag{1.2.11}$$

where $[\Theta]_1$ is the marginal of Θ on the first coordinate. Viewing ρ as a (relaxed) control, the right side of the above display gives a representation for the rate function as the value function of a stochastic control problem in which the goal is to produce a state process X with a specified law γ using the state dynamics (1.2.10) and a (nonanticipative) control process ρ which has the least cost, where the cost is given by the expectation on the right side of (1.2.11).

The above interpretation is a useful point of view and analogous stochastic control representations can be given more generally. Consider for example the case where we are given an i.i.d. collection of d-dimensional diffusions $\{X_i : i \in \mathbb{N}\}$ described through the stochastic differential equations

$$X_i(t) = x + \int_0^t b(X_i(s))\, ds + \int_0^t \sigma(X_i(s))\, dW_i(s), \tag{1.2.12}$$

where $x \in \mathbb{R}^d$ is an initial condition, and where for simplicity we assume that the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are Lipschitz functions so that the equations have a unique pathwise solution. Letting

$$\mu^n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}, \tag{1.2.13}$$

the rate function associated with the LDP for μ^n on $\mathcal{P}(\mathcal{C}([0,T],\mathbb{R}^d))$ takes the same form as (1.2.11) except $\mathcal{X} = \mathcal{C}([0,T],\mathbb{R}^d)$ and the class $\mathcal{E}_1$ is now the collection of all probability measures in $\mathcal{P}_2(\mathcal{Z}_1)$ under which W is as before and (X,ρ,W) are related as

$$X(t) = x + \int_0^t b(X(s))\,ds + \int_0^t \sigma(X(s))\,dW(s) + \int_{\mathbb{R}^m\times[0,t]} \sigma(X(s))y\,\rho(dy,ds).$$

Note that the system of equations in (1.2.12) has no interaction between particles. The models considered in Chapter 4 concern particles that interact partly through a driving Brownian motion that is common to all particles. We can introduce this small amount of coupling between the equations to the model (1.2.12) as follows:

$$X_i^n(t) = x + \int_0^t b(X_i^n(s))\,ds + \int_0^t \sigma(X_i^n(s))\,dW_i(s) + \kappa(n)\int_0^t \alpha(X_i^n(s))\,dB(s), \tag{1.2.14}$$

where B is a k-dimensional standard Brownian motion independent of $\{W_i\}$, $\alpha : \mathbb{R}^d \to \mathbb{R}^{d\times k}$ is a Lipschitz map, and $\kappa(n) \to 0$ as $n \to \infty$. In this case, since $\{X_i^n : 1 \le i \le n\}$ are not independent, the large deviation behavior of

$$\mu^n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^n} \tag{1.2.15}$$

cannot be deduced from Sanov's theorem, and in fact this behavior crucially depends on the manner in which $\kappa(n) \to 0$. The measures μ^n in (1.2.13) and in (1.2.15) converge to the same LLN limit but the rates of convergence as measured by the large deviation rate function are different. Indeed, as an elementary corollary of Theorems 4.1.1 and 4.1.3 in Chapter 4, the convergence rate for different choices of the small noise coefficient $\kappa(n)$ can be characterized (see Remark 4.1.3). Specifically, when $\kappa(n) = n^{-1/2}$ the rate function is governed by a different type of stochastic control problem than (1.2.11) that can be described as follows. For $\varphi \in L^2([0,T],\mathbb{R}^k)$, the space of square-integrable functions from $[0,T]$ into $\mathbb{R}^k$, let $\mathcal{E}_1[\varphi]$ denote the subset of $\mathcal{P}_2(\mathcal{Z}_1)$ consisting of all probability measures under which W is, as before, a m-dimensional Brownian motion with respect to the canonical filtration $\{\mathcal{F}_t\}$, and the coordinate processes X, ρ, and W are related to φ through the equation

$$X(t) = x + \int_0^t b(X(s))\,ds + \int_0^t \sigma(X(s))\,dW(s) + \int_{\mathbb{R}^m\times[0,t]} \sigma(X(s))y\,\rho(dy,ds)$$

$$+\int_0^t \alpha(X(s))\varphi(s)\,ds.$$

Then the rate function $I(\cdot)$ associated with the empirical measures μ^n in (1.2.15), with $\kappa(n) = n^{-1/2}$, is given as

$$I(\gamma) = \inf_{\varphi\in L^2([0,T],\mathbb{R}^k)} \left\{ \inf_{\Theta\in\mathcal{E}_1[\varphi]:[\Theta]_1=\gamma} E_\Theta \left[\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\,\rho(dy,dt)\right] + \frac{1}{2}\int_0^T \|\varphi(t)\|^2\,dt \right\}, \tag{1.2.16}$$

where $L^2([0,T],\mathbb{R}^k)$ denotes the space of square integrable functions $\varphi : [0,T] \to \mathbb{R}^k$. The right side of (1.2.16) is once more the value function of a stochastic control problem, however this time there are two types of controls. One of the controls, represented by ρ, is random and nonanticipative and arises from the aggregated contributions of the individual Brownian noises, whereas the second control, represented by φ, is nonrandom and corresponds to the small common Brownian noise that impacts all particles.

1.3 Contributions

The results of Chapter 2 can be viewed as a generalization of the Cramér-Lundberg asymptotic and Siegmund's algorithm in two directions, the main one being to the case of branching random walks. Historically, the asymptotic derived in terms of ladder heights for the random walk predates the use of Goldie's implicit renewal theory, which is a more robust method in the sense that it allows for a perturbation and is applicable to branching recursions other than Lindley's equation. For the branching case, the asymptotic was established (along with that for other recursions) by extending implicit renewal theory as discussed in the previous section. The main goal of Chapter 2 then is not to prove the asymptotic result, but rather to approach the problem by deriving a representation (2.1.5) that generalizes (1.1.7), which can be analyzed via direct renewal theoretic arguments, obtaining an alternate expression for the asymptotic constant. The other, potentially more important consequence of such a representation is that out of it falls an importance sampling estimator for the branching recursion in the spirit of Siegmund's algorithm. While there exist methods for approximating branching recursions with non-branching ones to arrive at simulation schemes (see e.g. [29]), previous analysis, including the asymptotic result itself, does not allow for an unbiased, strongly efficient (in the sense of bounded relative error) sampler for the tail probabilities $P(W > t)$.

The use of a spine change of measure on the underlying weighted Galton-Watson process is a natural way to arrive at a representation that mimics the non-branching scenario, since such a change of measure results in "tilting" a single path down the tree. While the formulation of the product martingale used to construct the change of measure in Chapter 2 is non-standard, the work of Lyons [76] suggests that there is a connection with the standard Biggins-Kyprianou martingale [11, 72] used frequently in the branching process literature. Change of measures that tilt a single branch of a Galton-Watson process have been used also in other contexts [77], however they have not been exploited to obtain efficient rare event samplers as is done here.

The second generalization of the classical Cramér-Lundberg theory that appears in Chapter 2 is to the case of the perturbed random walk, both non-branching and branching. The inclusion of the perturbation precludes the use of ladder height analysis in the renewal-theoretic results, and we opt instead to use the Markovian nature of the branching process to apply a suitable Markov renewal theorem. While the asymptotic behavior of tail events of the perturbed random walk on $\mathbb{R}$ have been studied and correct asymptotic decay rates have been established [6], as noted in Remark 2.1.2(b), the change of measure has been applied incorrectly in the literature. Hence, a more minor contribution of Chapter 2 is that it provides the equivalent representation to (1.1.7) for the perturbed random walk as well as the corresponding importance sampling estimator.

The representation (2.1.5) and the structure of the tree under the change of measure (namely that it tilts one path while leaving the distribution elsewhere invariant) opens up the possibility for even more efficient rare event sampling schemes to be developed. In fact, this is what is done in Chapter 3. In particular, a sampling scheme is developed therein with computational complexity that does not depend on the offspring distribution of the Galton-Watson process, where complexity is measured in terms of the number of input random vectors to the high order Lindley equation that need to be generated. The unbiased algorithm in Chapter 2 requires the simulation of a tree up to a generation on the order of t for a single estimate of $P(W > t)$, resulting in a computational complexity that is exponential in t. On the other hand, the biased algorithm in Chapter 3 has complexity that is linear in t, enabling efficient sampling even when t or the mean of the offspring distribution is very large.

The large deviation analysis of Chapter 4 is done for an interacting particle system that is a combination of models that have been previously studied. In particular, the model concerns weakly-interacting Brownian particles that are driven by both independent and small, common sources of noise. Such systems with a "common factor" arise in many fields, including in neuroscience, where they model systematic noise in the external input to neuronal ensembles [43], and in mathematical finance, where they model global sources of risk [30]. Additionally, an interacting particle model approximation of Feynman-Kac functionals is considered. The main contribution of Chapter 4 is to specify the large deviation rate function in terms of those for each of the two contributing models, and describe how it changes with the rate at which the common noise vanishes.

The large deviation principles are proved for the empirical measures of the particles as n, the number of particles, increases, and the rate function is shown to take qualitatively different forms when the small noise scales as n^{-r} for $r < 1/2$, $r = 1/2$, and $r > 1/2$. In particular, in the $r < 1/2$ regime, the rate function corresponds to that for the model dominated by the independent sources of noise, and when $r > 1/2$, the rate function looks like the rate function that appears in Friedlin-Wentzell small-noise asymptotic results. The $r = 1/2$ case is balanced and yields a rate function that is the sum of each contributing part.

The model in Chapter 5 is simpler, consisting of a weakly-interacting particle model driven by independent, small sources of noise. In addition to considering the large deviation behavior of the empirical measures of the particles, a large deviation principle is also established for certain functionals of the particle states called *stochastic currents*, which a.s. take values in certain negative fractional-order Sobolev spaces. A joint large deviation principle for the empirical measures and stochastic currents for a simpler version of this model was already established in [88], but the contributions of Chapter 5 are improvements upon this result in several ways.

The first is that weak convergence proof techniques are employed to significantly reduce the length of the proof. These techniques also easily allow us to consider a more general version of the model that includes a non-constant diffusion coefficient that includes interaction among the particles. In particular, this also allows for degeneracy of the diffusion coefficient, unlike in [88]. Additionally, the LDP in [88] for the stochastic currents is proved in the weak-$*$ topology on the appropriate Sobolev space, and we use Sobolev embedding techniques to establish the LDP in the much stronger norm topology. The main technical challenge to this last improvement lies in using weak convergence arguments when Sobolev spaces with the norm topology lack certain compactness properties that usually hold in this context. This is overcome with the standard technique of slightly widening the time domain and proving auxiliary compact embedding lemmas.

CHAPTER 2

Importance Sampling For Maxima on Trees

This chapter concerns the SFPE

$$W \stackrel{\mathcal{D}}{=} \max\left\{Y, \max_{1\leq i\leq N}(X_i + W_i)\right\}, \tag{2.0.1}$$

where the $\{W_i\}$ are i.i.d. copies of W, independent of the vector $(Y, N, \{X_i\})$, with $N \in \mathbb{N}$, which is known in the literature as the high-order Lindley equation [12, 64, 68, 87]. The special case of $N \equiv 1$ and $Y \equiv 0$, known as the Lindley equation,

$$W \stackrel{\mathcal{D}}{=} \max\{0, X + W\}, \tag{2.0.2}$$

is perhaps one of the best studied recursions in applied probability, since it describes the stationary distribution of the waiting time in a single-server queue fed by a renewal process and having i.i.d. service times; see the discussion in the Introduction, and see Asmussen [7] and Cohen [25] for a more comprehensive overview. If we replace the zero in (2.0.2) with a random Y we obtain a recursion satisfied by the all-time supremum of a "perturbed" random walk, where the Y denotes the perturbation. This type of distributional recursion was analyzed, for example, in [6, 53, 61]. The branching form (2.0.1) appears in the study of queueing networks with synchronization requirements [68, 87] and in the analysis of the maximum displacement of a branching random walk [12].

Given both the theoretical and practical importance of the special endogenous solution to (2.0.1), the focus of this paper is the study of its asymptotic tail behavior, i.e., $P(W > t)$ for large t. Recall from Theorem 1.1.6 in Section 1.1.4 that tail events for the endogenous solution to (2.0.1) decay exponentially, i.e., provided there exists $\alpha > 0$ such that

$$E\left[\sum_{i=1}^{N} e^{\alpha X_i}\right] = 1 \quad \text{and} \quad 0 < E\left[\sum_{i=1}^{N} e^{\alpha X_i} X_i\right] < \infty,$$

then

$$P(W > t) \sim H e^{-\alpha t}, \qquad t \to \infty,$$

as established in [64] using implicit renewal theory [49, 62, 63]. However, the constant H provided by the theorem is implicitly defined in terms of W itself, making its interpretation even less obvious than in the non-branching case.

Hence, the goal of this chapter is to provide an alternative representation for $P(W > t)$ yielding: 1) an unbiased and easy to simulate algorithm for $P(W > t)$ for all values of t, and 2) an alternative expression for H that better reflects the behavior of the underlying weighted branching random walk leading to the event $\{W > t\}$. The main tool enabling our first goal is a new interpretation of the measure $E\left[\sum_{i=1}^{N} e^{\alpha X_i} 1(X_i \in dx)\right]$ appearing in the renewal theoretic approaches for establishing the existence of H [31, 60, 74, 75, 12, 62, 63] in terms of a distinguished path of the branching process, with which we form a change of measure. The second goal, that of obtaining an alternative representation for H, is attained by applying the Markov renewal theorem from [4] to our representation. The proposed simulation algorithm yields an unbiased and strongly efficient estimator for the probability $P(W > t)$, much in the spirit of the importance sampling approach provided by Siegmund's algorithm for the Lindley equation (see Section 1.1.3). Importance samplers were also constructed in [14, 27] for the tail distribution of the solution of an affine version of the equation (when $N \equiv 1$), in [92] in the context of sequential analysis, in [48] for Markov chains and semi-Markov processes, and in [26] for Markov-modulated walks. For general review on rare-event simulation we refer the reader to [13, 19].

The change of measure we propose is of independent interest, since it formally differs from the typical one encountered in the weighted branching processes literature. It is constructed along a random path $\{\boldsymbol{J}_r\}_{r\geq 0}$ of the underlying weighted branching process, which we refer to as the *spine*, and changes its drift while leaving all other paths unchanged. The likelihood martingale $L_n = \prod_{r=0}^{n-1} D_{\boldsymbol{J}_r}$ used in our approach is constructed as a product of certain random variables $D_{\boldsymbol{i}}$ along the spine, and it is different from the Biggins-Kyprianou martingale $W_n(\alpha)$ defined in (1.1.17), which is constructed along the 'width' of the tree. Finally, our change of measure approach also provides important insights into the exponential asymptotics described by the implicit renewal theorem [49, 64].

Most of the analysis in this chapter is done in terms of the multiplicative version

$$R \stackrel{\mathcal{D}}{=} Q \vee \left(\bigvee_{i=1}^{N} C_i R_i \right) \tag{2.0.3}$$

of (2.0.1), where $Y = \log Q$, $X_i = \log C_i$, and $W = \log R$, and in terms of the corresponding multiplicative random walk. Recall from Section 1.1.4 the construction of the weighted branching process and the tree $\mathcal{T}$ constructed from i.i.d. copies

$$\{\boldsymbol{\psi}_{\boldsymbol{i}} : \boldsymbol{i} \in U\} = \left\{ (Q_{\boldsymbol{i}}, N_{\boldsymbol{i}}, \{C_{(\boldsymbol{i},j)}\}_{j\geq 1}) : \boldsymbol{i} \in U \right\},$$

where U is the set of all finite strings of positive integers. Also recall that the endogenous solution to (2.0.3) is given by

$$R \doteq \bigvee_{\boldsymbol{i}\in\mathcal{T}} \Pi_{\boldsymbol{i}} Q_{\boldsymbol{i}}, \tag{2.0.4}$$

where $\Pi_{\boldsymbol{i}} = \prod_{j=1}^{k} C_{\boldsymbol{i}|j}$ when $|\boldsymbol{i}| = k$.

Our analysis of R will rely on a set of assumptions satisfied by the generic branching vector $\boldsymbol{\psi} = (N, Q, C_1, C_2, \dots)$.

Condition 2.0.1. *$(N, Q, C_1, C_2, \dots)$ is nonnegative a.s. with $N \in \mathbb{N}_+ \cup \{\infty\}$, and $P(Q > 0) > 0$. Furthermore, for some $\alpha > 0$,*

(a) $E\left[\sum_{i=1}^{N} C_i^{\alpha}\right] = 1$ and $E\left[\sum_{i=1}^{N} C_i^{\alpha} \log C_i\right] \in (0, \infty)$,

(b) $E\left[\sum_{i=1}^{N} C_i^{\beta}\right] < 1$ for some $0 < \beta < \alpha$ and $E[Q^{\alpha}] < \infty$,

(c) $P\left(\sum_{i=1}^{N} C_i^{\alpha} = 0\right) = 0$,

(d) The probability measure $\eta(dx) = E\left[\sum_{i=1}^{N} C_i^{\alpha} 1(\log C_i \in dx)\right]$ is nonlattice,

(e) $E\left[\left(\sum_{i=1}^{N} C_i^{\alpha}\right) \log^{+}\left(Q^{\alpha} \vee \sum_{i=1}^{N} C_i^{\alpha}\right)\right] < \infty$.

Since the approach followed here is different from the one used in the implicit renewal theorem found in [64], our assumptions for establishing the representation of the constant in Theorem 2.3.1 are slightly different. In particular, conditions (c) and (e) are new. Condition (c) will be needed to ensure that our change of measure is well-defined, and condition (e) will guarantee that the positive part of the perturbed branching random walk has finite mean under said change of measure. On the other hand, the implicit renewal theorem requires the following assumption, which we use only for the positivity of the constant in Theorem 2.3.1.

Condition 2.0.2. *$E\left[\left(\sum_{i=1}^{N} C_i\right)^{\alpha}\right] < \infty$ if $\alpha > 1$ and $E\left[\left(\sum_{i=1}^{N} C_i^{\alpha/(1+\epsilon)}\right)^{1+\epsilon}\right] < \infty$ for some $0 < \epsilon < 1$ if $0 < \alpha \leq 1$.*

Observe that apart from these assumptions, the dependence structure in the vector $\boldsymbol{\psi}$ is arbitrary.

2.1 Change of Measure Along a Path

Note that for any path $\boldsymbol{i}$ originating at the root of $\mathcal{T}$, we can define a random walk by setting $S_{\boldsymbol{i}} = \log \Pi_{\boldsymbol{i}}$. Moreover, by letting $Y_{\boldsymbol{i}} = \log Q_{\boldsymbol{i}}$, we obtain that

$$W \doteq \log R = \bigvee_{\boldsymbol{i} \in \mathcal{T}} (S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}) \tag{2.1.1}$$

represents the maximum of a perturbed branching random walk.

Since our goal is to analyze the tail distribution $P(W > t)$ (equivalently, of $P(R > t)$) for all values of t, the key idea of our analysis is to apply a change of measure to the perturbed branching random walk under which the event $\{W > t\}$ for large t is no longer rare. This is exactly the usual approach for studying the maximum of the standard random walk under Cramér conditions (i.e., the existence of $\alpha > 0$ such that $E\left[e^{\alpha X}\right] = 1$ and $0 < E\left[X e^{\alpha X}\right] < \infty$).

However, in the branching case the change of measure is not as straightforward as in the non-branching case, where we use the exponential martingale to define it (see Section 1.1.2).

Note that under the condition $E\left[\sum_{i=1}^{N} C_i^{\beta}\right] < 1$ for some $\beta > 0$, the paths in the tree $\mathcal{T}$ have negative drift (see the calculation (1.1.18)). The change of measure we seek is obtained by making the drift of one path positive. Starting at the root, we will pick this *chosen path* by selecting one of its offspring at random, with a probability proportional to its weight raised to the α power. This procedure will allow us to define a suitable mean one nonnegative martingale to induce a change of measure on the entire tree. As we will show later, the change in the drift will not affect any subtrees whose roots are not part of the chosen path, allowing us to isolate the (small) set of paths responsible for the rare event $\{W > t\}$.

More precisely, let $\boldsymbol{J}_0 = \emptyset$ denote the root of $\mathcal{T}$. We now recursively define the random indices along the chosen path, $\{\boldsymbol{J}_k : k \geq 1\}$, as follows:

$$P\left(\boldsymbol{J}_k = (\boldsymbol{J}_{k-1}, i) \,\middle|\, \boldsymbol{\psi}_{\boldsymbol{J}_{k-1}}\right) = \frac{C^{\alpha}_{(\boldsymbol{J}_{k-1},i)}}{D_{\boldsymbol{J}_{k-1}}}, \qquad 1 \leq i \leq N_{\boldsymbol{J}_{k-1}}, \quad k \geq 1,$$

where

$$D_{\boldsymbol{i}} = \sum_{r=1}^{N_{\boldsymbol{i}}} C^{\alpha}_{(\boldsymbol{i},r)}$$

for any $\boldsymbol{i} \in U$, with generic copy $D = \sum_{i=1}^{N} C_i^{\alpha}$. From now on, we will refer to this chosen path along with its offspring and sibling nodes as the *spine*. Note that the sequence of indexes $\{\boldsymbol{J}_k : k \geq 0\}$ identifies all the nodes in the spine, with node $\boldsymbol{J}_k$ denoting the one in the kth generation of $\mathcal{T}$.

We now use the spine to define a mean one nonnegative martingale for our change of measure. To this end, define

$$L_0 = 1, \qquad\qquad L_k = \prod_{r=0}^{k-1} D_{\boldsymbol{J}_r}\,, \qquad k \geq 1\,,$$

and note that if we let $\mathcal{F}_k = \sigma(\boldsymbol{\psi}_{\boldsymbol{i}} : \boldsymbol{i} \in A_s,\ s < k)$ and $\mathcal{G}_k = \sigma(\mathcal{F}_k \cup \sigma(\boldsymbol{J}_s : s \leq k))$ for $k \geq 1$ and $\mathcal{F}_0 = \mathcal{G}_0 = \sigma(\varnothing)$, then

$$E\left[\left. L_k \right| \mathcal{G}_{k-1}\right] = L_{k-1} E\left[\left. D_{\boldsymbol{J}_{k-1}} \right| \mathcal{G}_{k-1}\right] = L_{k-1}\,.$$

Therefore $\{L_k : k \geq 0\}$ is a nonnegative martingale with mean one, measurable with respect to the filtration $\{\mathcal{G}_k : k \geq 0\}$. Setting

$$\widetilde{P}(A) = E[1(A) L_k]\,, \qquad \text{for } A \in \mathcal{G}_k \quad \text{and all } k \geq 0\,, \tag{2.1.2}$$

we obtain a new probability measure on $\mathcal{G} = \sigma\left(\bigcup_{k\geq 1} \mathcal{G}_k\right)$ (see Theorem 1.1.4(i)). Note in particular that $\boldsymbol{\psi} = \boldsymbol{\psi}_\emptyset$ satisfies

$$\widetilde{P}(\boldsymbol{\psi} \in B) = E\left[1(\boldsymbol{\psi} \in B)L_1\right] = E\left[1(\boldsymbol{\psi} \in B)D\right]$$

for Borel sets B.

Remark 2.1.1. (a) By Theorem 1.1.4(ii), P is absolutely continuous with respect to the measure induced by (2.1.2) up to each finite time step k. Namely,

$$P(A) = \widetilde{E}\left[1(A)L_k^{-1}\right], \qquad A \in \mathcal{G}_k,$$

and so the support of $\boldsymbol{\psi}$ does not change under the new measure. In particular, we can write

$$P(L_1 = 0) = \widetilde{E}\left[1(L_1 = 0)L_1^{-1}\right].$$

Since $1(L_1 = 0)L_1^{-1} \in \{0, \infty\}$, it must be the case that $\widetilde{P}(L_1 = 0) = 0$, which in turn must mean $P(L_1 = 0) = 0$. The corresponding assumption is given by Condition 2.0.1(c). In particular, we assume that $P(N \geq 1) = 1$. It is common for conditioning that the tree does not die to appear in the literature in the context of spine changes of measure. See, for instance, [89].

(b) It is worth mentioning that both Goldie's implicit renewal theorem [49] and the implicit renewal theorem on trees [62, 64] allow $P(L_1 = 0) > 0$, which is precluded by Condition 2.0.1. Our current setting is less general because it clearly identifies the most likely path to the rare event $\{W > t\}$ in cases where it is solely determined by the behavior of the spine. However, the implicit renewal theorems cover cases where the most likely path to the rare event is somewhat different than the one we will describe, which translates into the same exponential decay but with a different constant.

As mentioned earlier, the change of measure defined above only affects the drift of the random walk and the perturbation along the spine. Moreover, it preserves the branching property, i.e., the independence between the vectors $\{\boldsymbol{\psi}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{T}\}$. The following result formalizes this statement; its proof is given in Section 2.5. Throughout, we use the convention $\sum_{i=a}^{b} x_i \equiv 0$ whenever $a > b$.

Lemma 2.1.1. *Suppose Condition 2.0.1(a) holds. For any measurable set $B \in \mathbb{N} \times \mathbb{R}^\infty$, and any $\boldsymbol{i} \in A_k$,*

$$\widetilde{P}(\boldsymbol{J}_k = \boldsymbol{i}) = \prod_{r=1}^{k} E[C_{i_r}^{\alpha} 1(N \geq i_r)],$$

$$\widetilde{P}(\boldsymbol{\psi}_{\boldsymbol{i}} \in B | \boldsymbol{i} \neq \boldsymbol{J}_k) = P\left((N, Q, C_1, C_2, \dots) \in B\right),$$

$$\widetilde{P}(\boldsymbol{\psi}_{\boldsymbol{i}} \in B | \boldsymbol{i} = \boldsymbol{J}_k) = E\left[1\left((N, Q, C_1, C_2, \dots) \in B\right) \sum_{j=1}^{N} C_j^\alpha\right].$$

Moreover, under $\widetilde{P}$, the vectors $\{\boldsymbol{\psi}_{\boldsymbol{i}} : \boldsymbol{i} \in A_k\}$ are conditionally independent given $\mathcal{G}_{k-1}$ for any $k \geq 1$.

Recall that by taking the logarithm of the weights we can define a perturbed random walk along every path $\mathbf{i} \in \mathcal{T}$. The one along the spine will be special, since it is the one being affected by the change of measure, and will be the only one guaranteed to eventually exceed any level t. To make this precise, let us define $X_{\boldsymbol{i}} = \log C_{\boldsymbol{i}}$ and note that for any $\boldsymbol{i} \in A_k$,

$$S_{\boldsymbol{i}} = \log \Pi_{\boldsymbol{i}} = X_{\boldsymbol{i}|1} + \cdots + X_{\boldsymbol{i}|k-1} + X_{\boldsymbol{i}},$$

where the $\{X_{\boldsymbol{i}|r}\}_{1 \leq r \leq k}$ are independent of each other, although not necessarily identically distributed. To identify the spine we use the notation $\hat{X}_k = X_{\boldsymbol{J}_k} = \log C_{\boldsymbol{J}_k}$, identify the random walk along the chosen path by

$$V_k = \hat{X}_1 + \cdots + \hat{X}_k, \qquad V_0 = 0, \tag{2.1.3}$$

and use

$$\xi_k = Y_{\boldsymbol{J}_k} = \log Q_{\boldsymbol{J}_k}, \qquad \xi_0 = Y_\emptyset = \log Q_\emptyset,$$

for its perturbation. The following result establishes that $\{V_k : k \geq 0\}$ defines a random walk with i.i.d. increments and positive drift.

Lemma 2.1.2. *Suppose Condition 2.0.1(a) holds. For all $k \geq 1$ and $x_1, \dots x_k, y \in \mathbb{R} \cup \{\infty\}$, we have*

$$\widetilde{P}\left(\hat{X}_1 \leq x_1, \dots, \hat{X}_k \leq x_k,\, \xi_k \leq y\right) = E\left[1(Q \leq e^y) \sum_{i=1}^{N} C_i^\alpha\right] \prod_{r=1}^{k} G(x_r),$$

where

$$G(x) = \sum_{i=1}^{\infty} E\left[1(C_i \leq e^x,\, N \geq i) C_i^\alpha\right] = E\left[\sum_{i=1}^{N} 1(\log C_i \leq x) C_i^\alpha\right].$$

In particular, the $\{\hat{X}_i : i \geq 1\}$ are i.i.d. with common distribution G under $\widetilde{P}$, $\widetilde{E}\left[|\hat{X}_1|\right] < \infty$, and have mean

$$\mu \doteq \widetilde{E}\left[\hat{X}_1\right] = E\left[\sum_{i=1}^{N} C_i^\alpha \log C_i\right] \in (0, \infty).$$

We now explain how to compute the probability $P(W > t)$ using the change of measure described above. We start by defining the hitting time of level t for the perturbed branching random walk defined by $\{(S_{\boldsymbol{i}}, Y_{\boldsymbol{i}}) : \boldsymbol{i} \in \mathcal{T}\}$, which we denote $\gamma(t) = \inf\{\boldsymbol{i} \in \mathcal{T} : S_{\boldsymbol{i}} + Y_{\boldsymbol{i}} > t\}$, where the infimum is taken according to the length-lexicographic

$\prec$ order defined in Section 1.1.4. We use $\nu(t) = |\gamma(t)|$ to denote the generation in the weighted branching process where the perturbed random walk along a path exceeds level t. Next, define the hitting time of level t along the spine, $\tau(t) = \inf\{k \geq 0 : V_k + \xi_k > t\}$.

Note that $\nu(t) + 1$ and $\tau(t) + 1$ are stopping times for the weighted branching process with respect to the filtration $\{\mathcal{G}_k : k \geq 0\}$, and since it is possible for a path different from the spine to hit level t before the spine does, then

$$\nu(t) \leq \tau(t),$$

with equality possible, e.g. if $\boldsymbol{J}_{\tau(t)} = \gamma(t)$. Moreover, since $W = \bigvee_{\boldsymbol{i} \in \mathcal{T}} (S_{\boldsymbol{i}} + Y_{\boldsymbol{i}})$, it follows that

$$P(W > t) = P(\nu(t) < \infty).$$

The next step is to apply the change of measure and derive an alternative representation for $P(\nu(t) < \infty)$. To this end, observe that on the set $\{\gamma(t) = \boldsymbol{i}\}$, we have $N_{\boldsymbol{i}|r-1} \geq i_r$ for all $r = 1, \ldots, k$ and that $\{\gamma(t) = \boldsymbol{i},\, \boldsymbol{J}_k = \boldsymbol{i}\} = \{\tau(t) = k,\, \boldsymbol{J}_{\tau(t)} = \gamma(t) = \boldsymbol{i}\}$. Also note that for $\boldsymbol{i} \in A_k$,

$$P\left(\boldsymbol{J}_k = \boldsymbol{i} \middle| \mathcal{F}_k \right) = \frac{\prod_{r=1}^{k} C_{\boldsymbol{i}|r}^{\alpha} 1(N_{\boldsymbol{i}|r-1} \geq i_r)}{\prod_{r=0}^{k-1} D_{\boldsymbol{i}|r}}.$$

Therefore, since $P\left(\boldsymbol{J}_k = \boldsymbol{i} \middle| \mathcal{F}_k \right) = P\left(\boldsymbol{J}_k = \boldsymbol{i} \middle| \mathcal{F}_{k+1} \right)$, and since $\prod_{r=0}^{k-1} D_{\boldsymbol{i}|r} > 0$ a.s. for all $\boldsymbol{i} \in A_k$ and all k, we have

$$\begin{aligned}
e^{\alpha t} P(W > t) &= E\left[\sum_{k=0}^{\infty} \sum_{\boldsymbol{i} \in A_k} e^{\alpha t} 1(\gamma(t) = \boldsymbol{i}) \right] \\
&= \sum_{k=0}^{\infty} E\left[\sum_{\boldsymbol{i} \in A_k} 1(\gamma(t) = \boldsymbol{i}) e^{\alpha t} \frac{\Pi_{\boldsymbol{i}}^{\alpha}}{\Pi_{\boldsymbol{i}}^{\alpha}} \cdot \frac{\prod_{r=0}^{k-1} D_{\boldsymbol{i}|r}}{\prod_{r=0}^{k-1} D_{\boldsymbol{i}|r}} \right] \\
&= \sum_{k=0}^{\infty} E\left[\sum_{\boldsymbol{i} \in A_k} 1(\gamma(t) = \boldsymbol{i}) e^{-\alpha(S_{\boldsymbol{i}} - t)} \cdot \prod_{r=0}^{k-1} D_{\boldsymbol{i}|r} \cdot P\left(\boldsymbol{J}_k = \boldsymbol{i} \middle| \mathcal{F}_{k+1} \right) \right] \\
&= \sum_{k=0}^{\infty} E\left[E\left[\sum_{\boldsymbol{i} \in A_k} 1(\gamma(t) = \boldsymbol{i}) e^{-\alpha(S_{\boldsymbol{i}} - t)} L_k 1(\boldsymbol{J}_k = \boldsymbol{i}) \middle| \mathcal{F}_{k+1} \right] \right] \\
&= \sum_{k=0}^{\infty} E\left[1(\gamma(t) = \boldsymbol{J}_{\tau(t)},\, \tau(t) = k) e^{-\alpha(V_{\tau(t)} - t)} L_k \right]. \qquad (2.1.4)
\end{aligned}$$

Now recall that although $\tau(t)$ and $|\gamma(t)|$ are not stopping times with respect to $\{\mathcal{G}_k : k \geq 0\}$, $\tau(t) + 1$ and $|\gamma(t)| + 1$ are. Hence, multiplying and dividing by $D_{\boldsymbol{J}_k}$ we obtain

$$e^{\alpha t} P(W > t) = \sum_{k=0}^{\infty} E\left[1(\gamma(t) = \boldsymbol{J}_{\tau(t)},\, \tau(t) + 1 = k + 1) e^{-\alpha(V_{\tau(t)} - t)} D_{\boldsymbol{J}_k}^{-1} L_{k+1} \right]$$

$$= \widetilde{E}\left[1(\gamma(t) = \boldsymbol{J}_{\tau(t)}, \tau(t) < \infty) e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right].$$

Since $\{V_k + \xi_k : k \geq 0\}$ is a perturbed random walk with positive drift under $\widetilde{P}$, we have the following.

Lemma 2.1.3. *Under Condition 2.0.1(a)-(c), $\widetilde{P}(\tau(y) < \infty) = 1$ for any $y \in \mathbb{R}$.*

Hence, we obtain

$$e^{\alpha t} P(W > t) = \widetilde{E}\left[1(\boldsymbol{J}_{\tau(t)} = \gamma(t)) e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right]. \tag{2.1.5}$$

Note that the right-hand-side of (2.1.5) is an explicit function of the first $\tau(t)$ generations of a weighted branching process with a distinguished spine, which can be directly estimated using standard Monte Carlo methods, as discussed in Section 2.4.

Remark 2.1.2. (a) Note that if Q is independent of $(N, C_1, C_2, \dots)$, then we can use the filtration $\mathcal{F}'_0 = \sigma(Q_\emptyset)$, $\mathcal{F}'_k = \sigma\left(\psi_{\boldsymbol{i}} : \boldsymbol{i} \in A_s, s < k; Q_{\boldsymbol{j}} : \boldsymbol{j} \in A_k\right)$ and its corresponding $\mathcal{G}'_0 = \mathcal{F}'_0$, $\mathcal{G}'_k = \sigma\left(\mathcal{F}'_k \cup \sigma(\boldsymbol{J}_s : s \leq k)\right)$, with respect to which both $\tau(t)$ and $\nu(t)$ are stopping times, and obtain the simpler expression

$$e^{\alpha t} P(W > t) = \widetilde{E}\left[1(\boldsymbol{J}_{\tau(t)} = \gamma(t)) e^{-\alpha(V_{\tau(t)}-t)}\right].$$

(b) In the non-branching case ($N \equiv 1$), equation (2.1.5) reduces to

$$P(W > t) = \widetilde{E}\left[e^{-\alpha V_{\tau(t)+1}}\right], \tag{2.1.6}$$

which we point out is different from equation (3.4) in [6], since their expression has $V_{\tau(t)}$ instead of $V_{\tau(t)+1}$. As explained earlier, $\tau(t)$ is not a stopping time with respect to the natural filtration $\sigma(\hat{X}_i : 1 \leq i \leq n)$ of the martingale L_n, so the change of measure argument in [6] needs to be modified (see Theorem 3.2 in Chapter XIII of [7]). Once we consider the augmented filtration $\mathcal{F}_k$ (which is equal to $\mathcal{G}_k$ in the non-branching case) and apply the change of measure up to the stopping time $\tau(t) + 1$, we obtain the expression given by (2.1.6).

(c) The case where the Q is bounded is also special in the sense of the theory needed for its analysis. In particular, the exponential asymptotics of $P(W > t)$ can be easily obtained without using the augmented filtration nor any implicit renewal theory. To illustrate this we include in Section 2.5 (see Theorem 2.5.1) a very short proof of Theorem 1 in [6], for the non-branching case. Since the focus of here is to obtain a more explicit representation for the constant H obtained through the implicit renewal theorem on trees (Theorem 3.4 in [64]), we do not pursue the bounded Q case separately in the branching setting.

(d) Moreover in the case of a.s. bounded Q, say $P(Q \leq q) = 1$, we can obtain a Cramér-Lundberg type of inequality for $P(W > t)$ by defining $\gamma^*(t) = \inf\{\boldsymbol{i} \in \mathcal{T} : S_{\boldsymbol{i}} > t\}$, $\nu^*(t) = |\gamma^*(t)|$, and $\tau^*(t) = \inf\{n \geq 1 : V_n > t\}$, and noting that both $\nu^*(t)$ and $\tau^*(t)$ are stopping times with respect to the filtration $\sigma(\{(N_{\boldsymbol{i}}, C_{(\boldsymbol{i},1)}, C_{(\boldsymbol{i},2)}, \dots) : \boldsymbol{i} \in A_s,\ s < k\}, \{\boldsymbol{J}_s : s \leq k\})$. The same change of measure arguments used above yield for any $t > c \doteq \log q$:

$$
\begin{aligned}
P(W > t) &= P(\nu^*(t-c) \leq \nu(t) < \infty) \\
&\leq P(\nu^*(t-c) < \infty) \\
&= \widetilde{E}\left[1(\boldsymbol{J}_{\tau^*(t-c)} = \gamma^*(t-c)) e^{-\alpha V_{\tau^*(t-c)}}\right] \\
&\leq \widetilde{E}\left[e^{-\alpha(V_{\tau^*(t-c)} - t + c)}\right] e^{-\alpha(t-c)} \\
&\leq q^\alpha e^{-\alpha t}.
\end{aligned}
$$

This inequality for all $t \geq c$ holds under Assumption 2.0.1(a), and cannot be obtained using only the implicit renewal theorem for trees in [64].

2.2 The Markov Renewal Theorem

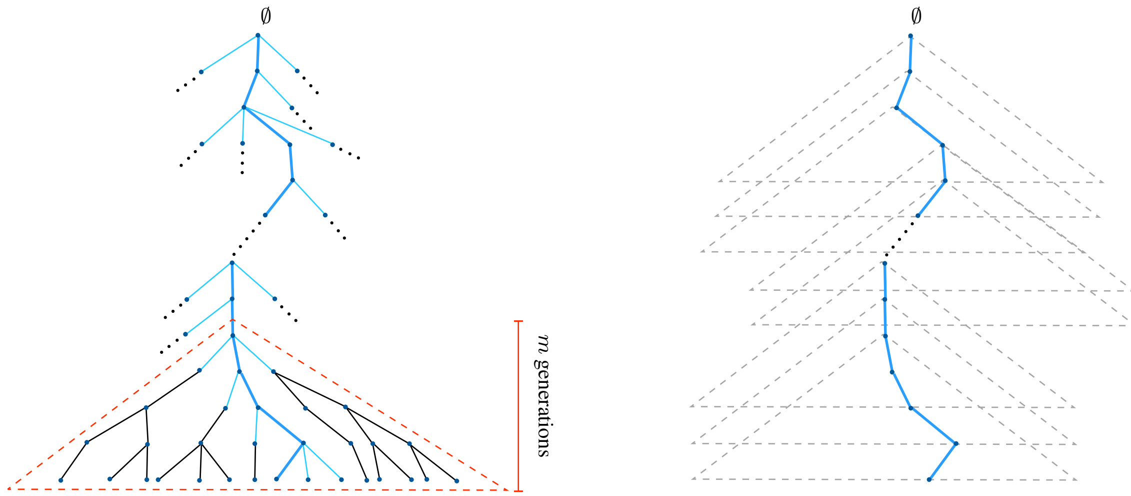
As pointed out, the new expression provided by (2.1.5) can easily be estimated via simulation. however, it can also be directly analyzed to obtain an alternative representation for the constant H in $P(W > t) \sim He^{-\alpha t}$, $t \to \infty$. The idea behind this analysis is the use of renewal theory on the expectation

$$
\widetilde{E}\left[1(\boldsymbol{J}_{\tau(t)} = \gamma(t)) e^{-\alpha(V_{\tau(t)} - t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right].
$$

Note that although the exponential term inside the expectation depends only on the random walk $\{V_k : k \geq 0\}$ and its hitting time of level t, the event $\{\boldsymbol{J}_{\tau(t)} = \gamma(t)\}$ depends on the history of the tree $\mathcal{T}$ up to generation $\nu(t)$. Hence, any renewal argument would need to include the latter, which complicates matters since its exponential growth (whenever $E[N] > 1$) implies it does not naturally renew at any point. However, intuitively, only the paths that branch out from the spine close to the time when the spine is likely to reach level t are likely to reach level t at all. This means that it should suffice to focus only on these paths, say a subtree of height m rooted at the spine that moves along the random walk $\{V_k : k \geq 0\}$; see Figure 2.1a. Since the sequence of such height-m subtrees forms a Harris chain, the key to our main theorem is the use of the Markov renewal theorem in [4].

To formalize this idea, we define the subtrees of height m rooted at node $\boldsymbol{J}_k$ (the kth node along the spine) according to:

$$
\mathcal{T}_k^{(m)} = \bigcup_{n=0}^{m-1} A_{n,\boldsymbol{J}_k}, \qquad k \geq 0, \tag{2.2.1}
$$



(a) The spine of $\mathcal{T}$.

(b) The subtrees of height m rooted in the spine, $\mathcal{T}_k^{(m)}$, $k \geq 0$.

Figure 2.1: The spine of $\mathcal{T}$ and the Markov chain consisting of subtrees.

where $A_{n,\boldsymbol{i}} = \{(\boldsymbol{i}, \boldsymbol{j}) \in \mathcal{T} : |\boldsymbol{j}| = n\}$ is the nth generation of the subtree rooted at node $\boldsymbol{i}$ (See Figure 2.1b). Focusing on these subtrees rooted at the spine allows us to analyze the expectation $\widetilde{E}\left[1(\boldsymbol{J}_{\tau(t)} = \gamma(t))e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right]$ using the Markov renewal theorem in [4]. Note that even in the non-branching case $(N \equiv 1)$, the perturbations that the Q's represent make it difficult to identify clear regeneration epochs for the process $\{V_k + \xi_k : k \geq 0\}$, which is a problem that is solved by looking not only at the current value of $V_n + \xi_n$, but also at its m-step history. We give more details on this idea and the intuition behind it in Section 2.5.

2.3 Main Result

Here we present the main theoretical result of this chapter. Recall $\mu = \widetilde{E}[\hat{X}_1] = E\left[\sum_{i=1}^{N} C_i^\alpha \log C_i\right]$.

Theorem 2.3.1. *If* $(N, Q, C_1, C_2, \ldots)$ *satisfies Condition 2.0.1 for some* $\alpha > 0$ *and* $W = \log R$*, where* R *is the endogenous solution to* (2.0.3) *given by* (2.0.4)*, then*

$$P(W > t) \sim He^{-\alpha t} \quad \textit{as} \quad t \to \infty,$$

where

$$H = \lim_{m\to\infty} \frac{\widetilde{E}\left[\left(e^{\alpha\xi_m} - e^{\alpha\left(\bigvee_{\boldsymbol{i}\prec \boldsymbol{J}_m}(S_{\boldsymbol{i}}+Y_{\boldsymbol{i}})-V_m\right)}\right)^+ D^{-1}_{\boldsymbol{J}_m}\right]}{\alpha\mu}.$$

If furthermore Condition 2.0.2 holds, then $H > 0$.

Remark 2.3.1. It it interesting to compare the expression for H in the theorem with its counterpart obtained through the use of the implicit renewal theorem on trees (Theorem 1.1.6 in the Introduction and Theorem 3.4 in [64]), which written in terms of W under our current assumptions* becomes

$$\frac{E\left[e^{\alpha Y} \vee \bigvee_{i=1}^{N} e^{\alpha(X_i+W_i)} - \sum_{i=1}^{N} e^{\alpha(X_i+W_i)}\right]}{\alpha\mu}, \tag{2.3.1}$$

where the $\{W_i\}$ are i.i.d. copies of W independent of the vector $(N, Y, X_1, X_2, \dots)$. As we can see, the two representations are significantly different, despite the fact that they are necessarily equal to each other. However, the representation given by Theorem 2.3.1 applies only to our setting where the rare event is determined by the spine, which under $\widetilde{P}$ behaves very differently than all other paths in the tree, while the constant obtained through the implicit renewal theorem on trees also works for the case where $P(L_1 = 0) > 0$.

Let $H_m = (\alpha\mu)^{-1}\widetilde{E}\left[\left(e^{\alpha\xi_m} - e^{\alpha\left(\bigvee_{i \prec J_m}(S_i+Y_i)-V_m\right)}\right)^{+} D_{\boldsymbol{J}_m}^{-1}\right]$. The rate of convergence of $H_m \to H$ can be shown to be geometrically fast when we add a slightly stronger assumption on the distribution of N than in Condition 2.0.1. More precisely, as a consequence of the proof of Theorem 2.3 we have the following.

Corollary 2.3.1. *Suppose Conditions 2.0.1 and 2.0.2 hold, and in addition suppose that*

$$\widetilde{E}[N] = E\left[N \sum_{i=1}^{N} C_i^{\alpha}\right] < \infty.$$

Then, there is a constant $0 < c < 1$ *such that*

$$|H - H_m| = O\left(c^m\right).$$

2.4 An Importance Sampling Algorithm

As discussed in Section 1.1.3, the same exponential change of measure used to establish the Cramér-Lundberg asymptotic in the non-branching case is well-known to provide an unbiased and strongly efficient estimator for the rare event probability $P(W > t)$ when t is large. Throughout this section we assume that $N < \infty$ a.s.

To relate our estimator to the one used in the non-branching case, suppose first that the goal is to estimate the tail distribution of the all-time maximum of the random walk $S_n = X_1 + \cdots + X_n$ when $E[X_1] < 0$. For large values of t, estimating $P(W > t) = P(\sup_n S_n > t)$ using the naive estimator $1(W > t)$ would require prohibitively large sample sizes, since its relative error grows unboundedly, i.e.,

$$\frac{\mathrm{Var}(1(W > t))}{P(W > t)^2} = \frac{P(W > t)P(W \le t)}{P(W > t)^2} \to \infty \quad \text{as} \quad t \to \infty,$$

*Theorem 3.4 in [64] allows Q, and therefore R, to take negative values.

However, whenever there exists $\alpha > 0$ such that $E\left[e^{\alpha X_1}\right] = 1$ and $E\left[X_1 e^{\alpha X_1}\right] \in (0, \infty)$, Siegmund's algorithm [92] takes advantage of the representation

$$P(W > t) = \widetilde{E}\left[e^{-\alpha S_{\tau(t)}} 1(\tau(t) < \infty)\right],$$

where the expectation is computed under the change of measure $\widetilde{P}(A) = E\left[1(A) e^{\alpha S_n}\right]$ for any set A measurable with respect to $\sigma(X_1, \ldots, X_n)$. Since under $\widetilde{P}$ the random walk has positive drift, $\widetilde{P}(\tau(t) < \infty) = 1$ and we obtain the estimator:

$$Z(t) = e^{-\alpha S_{\tau(t)}}.$$

This estimator is known to be strongly efficient, in the sense that it has bounded relative error, i.e,

$$\limsup_{t \to \infty} \frac{\widetilde{\mathrm{Var}}(Z(t))}{P(W > t)^2} < \infty,$$

where $\widetilde{\mathrm{Var}}(\cdot)$ denotes the variance under $\widetilde{P}$. Furthermore, since it can be shown that $t/\tau(t) \to \mu = E\left[X_1 e^{\alpha X_1}\right]$ $\widetilde{P}$-a.s., then computing $Z(t)$ requires that we simulate around t/μ steps of the random walk. For further details we refer to Section 1.1.3 herein and Chapter VI in [8].

Our proposed simulation approach for the branching case follows the same ideas described above. However, the issues we encounter while using a naive Monte Carlo approach are considerably worse, since simulating k generations of a tree requires, in general, an exponential in k number of random variables. Observe that in similar situations, the population dynamics algorithm [3, 84, 86] has been used to construct dependent samples which still yield strongly consistent estimators. However, our problem here is that we are interested in estimating the probability $P(W > t)$ for both moderate and large values of t, and in the latter case the size of such samples would again have to be prohibitively large in order to obtain enough observations larger than t.

Alternatively, we could try to estimate the expectation in the asymptotic expression

$$P(W > t) \sim \frac{E\left[e^{\alpha Y} \vee \bigvee_{i=1}^{N} e^{\alpha(X_i + W_i)} - \sum_{i=1}^{N} e^{\alpha(X_i + W_i)}\right]}{\alpha\mu} \cdot e^{-\alpha t}, \qquad t \to \infty,$$

provided by Theorem 3.4 in [64], since the population dynamics algorithm could be used to efficiently and accurately estimate the expectation involved. However, we would still have a bias due to the limit in t that cannot be explicitly computed, despite the availability of convergence rates in the implicit renewal theorem [65]. Instead, our proposed estimator follows the idea behind Siegmund's algorithm and is based on the representation

$$P(W > t) = \widetilde{E}\left[1(\boldsymbol{J}_{\tau(t)} = \gamma(t)) e^{-\alpha V_{\tau(t)}} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right],$$

derived in Section 2.1. Note that by Lemma 2.1.3, under Condition 2.0.1(a)-(c) we have $\widetilde{P}(\nu(t) \leq \tau(t) < \infty) = 1$, which suggests the estimator

$$Z(t) = 1(\boldsymbol{J}_{\tau(t)} = \gamma(t))e^{-\alpha V_{\tau(t)}} D^{-1}_{\boldsymbol{J}_{\tau(t)}}, \tag{2.4.1}$$

where the underlying tree $\mathcal{T}$ is simulated under the measure $\widetilde{P}$ up to the stopping time $\tau(t) + 1$.

Remark 2.4.1. By the discussion in Remark 2.1.2(a), when Q is independent of $(N, C_1, C_2, \ldots)$, the estimator

$$Z(t) = 1(\boldsymbol{J}_{\tau(t)} = \gamma(t))e^{-\alpha V_{\tau(t)}} \tag{2.4.2}$$

is also unbiased for $P(W > t)$. This is the prefered estimator in this case since the $D^{-1}_{\boldsymbol{J}_{\tau(t)}}$ in (2.4.1) is an unnecessary, independent source of variability. In the case that both $N \equiv 1$ and Q is independent of C_1, (2.4.2) reduces to the estimator in Siegmund's algorithm.

As with the non-branching case, we expect the spine to reach level t in about t/μ steps, or equivalently, t/μ generations of $\mathcal{T}$. The precise result is stated below; note that its proof is not a straightforward consequence of the strong law of large numbers due to the presence of the perturbations.

Lemma 2.4.1. *Under Condition 2.0.1, $\tau(t) \to \infty$ $\widetilde{P}$-a.s. as $t \to \infty$. In particular, $\tau(t) \sim t/\mu$ as $t \to \infty$ $\widetilde{P}$-a.s.*

Just as the estimator in Siegmund's algorithm, our proposed estimator is strongly efficient, although under a strengthened moment condition due to the perturbations.

Lemma 2.4.2. *Suppose Conditions 2.0.1 and 2.0.2 hold, so that $H > 0$. If $E\left[Q^{2\alpha} D^{-1}\right] < \infty$, then $Z(t)$ as defined by* (2.4.1) *has bounded relative error. If Q is independent of $(N, C_1, C_2, \ldots)$ and $E\left[Q^{2\alpha}\right] < \infty$, then $Z(t)$ in* (2.4.2) *has bounded relative error.*

In Table 2.1 we present an algorithm for simulating one copy of $Z(t)$ for fixed $t > 0$. At the start, we assume we have computed the value of α such that $E\left[\sum_{i=1}^{N} C_i^{\alpha}\right] = 1$ as well as the corresponding tilted distribution for the nodes along the spine under $\widetilde{P}$, and that we are capable of simulating $(N, Q, C_1, C_2, \ldots)$ both under P and under the tilted measure. To distinguish the two distributions, let $\tilde{\boldsymbol{\psi}} = (\tilde{N}, \tilde{Q}, \tilde{C}_1, \tilde{C}_2, \ldots)$ denote a vector having the tilted distribution:

$$\widetilde{P}(\tilde{\boldsymbol{\psi}} \in B) = \widetilde{P}\left(\boldsymbol{\psi}_{\boldsymbol{i}} \in B | \boldsymbol{i} = \boldsymbol{J}_k\right) = E\left[1(\boldsymbol{\psi} \in B) \sum_{i=1}^{N} C_j^{\alpha}\right], \qquad B \subseteq \mathbb{N} \times \mathbb{R}^{\infty},$$

and let $\boldsymbol{\psi} = (N, Q, C_1, C_2, \ldots)$ denote a vector having the original distribution under P; the simulation of the tree $\mathcal{T}$ will always be done under $\widetilde{P}$.

Table 2.1: Importance Sampling Algorithm

Input: $t > 0$

Output: A single copy of $Z = 1(\boldsymbol{J}_{\tau(t)} = \gamma(t)) e^{-\alpha V_{\tau(t)}} D^{-1}_{\boldsymbol{J}_{\tau(t)}}$

Generate $(N, Q, C_1, \ldots, C_N) \stackrel{\mathcal{D}}{=} \tilde{\boldsymbol{\psi}}$

Choose $j \in \{1, \ldots, N\}$ w.p. $C_j^\alpha / \sum_{i=1}^N C_i^\alpha$ and set $\boldsymbol{J}_1 \leftarrow j$

Set $Y = \log Q$ and $S_j \leftarrow \log C_j$ for $j = 1, \ldots, N$

Initialize $S_\emptyset \leftarrow 0$, $Y_\emptyset \leftarrow Y$, $\boldsymbol{i} \leftarrow \emptyset$, $\boldsymbol{J}_0 \leftarrow \emptyset$

while $S_{\boldsymbol{i}} + Y_{\boldsymbol{i}} \leq t$ **do**

    Update $\boldsymbol{i} \leftarrow \min\{\boldsymbol{j} : \boldsymbol{i} \prec \boldsymbol{j}\}$

    if $\boldsymbol{i} = \boldsymbol{J}_{|\boldsymbol{i}|}$ **then**

        Generate $\left(N_{\boldsymbol{i}}, Q_{\boldsymbol{i}}, C_{(\boldsymbol{i},1)}, \ldots, C_{(\boldsymbol{i},N_{\boldsymbol{i}})}\right) \stackrel{\mathcal{D}}{=} \tilde{\boldsymbol{\psi}}$

        Choose $j \in \{1, \ldots, N_{\boldsymbol{i}}\}$ w.p. $C_{(\boldsymbol{i},j)}^\alpha / \sum_{i=1}^{N_{\boldsymbol{i}}} C_{(\boldsymbol{i},i)}^\alpha$, and set $\boldsymbol{J}_{|\boldsymbol{i}|+1} \leftarrow (\boldsymbol{J}_{|\boldsymbol{i}|}, j)$

    else

        Generate $\left(N_{\boldsymbol{i}}, Q_{\boldsymbol{i}}, C_{(\boldsymbol{i},1)}, \ldots, C_{(\boldsymbol{i},N_{\boldsymbol{i}})}\right) \stackrel{\mathcal{D}}{=} \boldsymbol{\psi}$

    end if

    Set $Y_{\boldsymbol{i}} \leftarrow \log Q_{\boldsymbol{i}}$ and $S_{(\boldsymbol{i},j)} \leftarrow S_{\boldsymbol{i}} + \log C_{(\boldsymbol{i},j)}$ for $j = 1, \ldots, N_{\boldsymbol{i}}$

end while

if $\boldsymbol{i} = \boldsymbol{J}_{|\boldsymbol{i}|}$ **then**

    Set $Z \leftarrow e^{-\alpha S_{\boldsymbol{i}}} / \sum_{i=1}^{N_{\boldsymbol{i}}} C_{(\boldsymbol{i},i)}^\alpha$

else

    Set $Z \leftarrow 0$

end if

Output Z

2.4.1 Examples

We now illustrate the use of our proposed simulation algorithm by providing some examples for which both the random vectors $\boldsymbol{\psi}$ and $\tilde{\boldsymbol{\psi}}$ can be easily simulated. The particular form of the change of measure poses a simulation challenge since the tilt introduces dependence between N and the $\{C_i\}$ even if none exists under P. We start with three generic approaches for simulating $\tilde{\boldsymbol{\psi}}$ and then provide more concrete examples.

Example 2.4.1 (Acceptance-rejection for bounded C's, part I)**.** When the C_i are a.s. bounded, an acceptance-rejection algorithm based on the original distribution of $\boldsymbol{\psi}$ under P can be employed to generate a sample of $\tilde{\boldsymbol{\psi}}$. Suppose that $C_i \leq b_i$ a.s. for each i and note that

$$\widetilde{P}(\tilde{N} = n) = E\left[1(N = n) \sum_{i=1}^N C_i^\alpha\right] = P(N = n) \sum_{i=1}^n E\left[C_i^\alpha \middle| N = n\right],$$

so that

$$\widetilde{P}(\tilde{Q} \in dy, \tilde{N} = n, \tilde{C}_1 \in dx_1, \ldots, \tilde{C}_n \in dx_n)$$

$$
\begin{aligned}
&= P(N=n) E\left[1(Q \in dy, C_1 \in dx_1, \ldots, C_n \in dx_n) \sum_{i=1}^{n} C_i^{\alpha} \,\middle|\, N=n \right] \\
&= P(N=n) \left(\sum_{i=1}^{n} E[C_i^{\alpha} | N=n] \right) \cdot \frac{E\left[1(Q \in dy, C_1 \in dx_1, \ldots, C_n \in dx_n) \sum_{i=1}^{n} C_i^{\alpha} \middle| \, N=n \right]}{\sum_{i=1}^{n} E[C_i^{\alpha} | N=n]} \\
&= \widetilde{P}(\tilde{N}=n) \cdot \frac{E\left[1(Q \in dy, C_1 \in dx_1, \ldots, C_n \in dx_n) \sum_{i=1}^{n} C_i^{\alpha} \middle| \, N=n \right]}{\sum_{i=1}^{n} E[C_i^{\alpha} | N=n]}.
\end{aligned}
$$

Thus, the conditional density of $(\tilde{Q}, \tilde{C}_1, \ldots, \tilde{C}_n)$ given $\tilde{N}=n$ can be dominated as follows:

$$
\begin{aligned}
f_{\tilde{Q}, \tilde{C}_1, \ldots, \tilde{C}_n | \tilde{N}=n}(y, x_1, \ldots, x_n) &= \frac{\sum_{i=1}^{n} x_i^{\alpha}}{\sum_{i=1}^{n} E[C_i^{\alpha} | N=n]} f_{Q, C_1, \ldots, C_n | N=n}(y, x_1, \ldots, x_n) \\
&\leq \frac{\sum_{i=1}^{n} b_i^{\alpha}}{\sum_{i=1}^{n} E[C_i^{\alpha} | N=n]} f_{Q, C_1, \ldots, C_n | N=n}(y, x_1, \ldots, x_n),
\end{aligned}
$$

where $f_{Q, C_1, \ldots, C_n | N=n}$ denotes the conditional density of $(Q, C_1, \ldots, C_n)$ given $N=n$ with respect to P. Hence, after obtaining $\tilde{N}=n$ by simulation, an observation of $(\tilde{Q}, \tilde{C}_1, \ldots, \tilde{C}_n)$ can be obtained by using an acceptance-rejection procedure where we simulate U from a Uniform$[0,1]$ distribution and $(Q, C_1, \ldots, C_n)$ according to $f_{Q, C_1, \ldots, C_n | N=n}$, independent of each other, and then set $(\tilde{Q}, \tilde{C}_1, \ldots, \tilde{C}_n) = (Q, C_1, \ldots, C_n)$ if

$$
U \leq \frac{\sum_{i=1}^{n} C_i^{\alpha}}{\sum_{i=1}^{n} b_i^{\alpha}}.
$$

The acceptance probability given $\tilde{N}=n$ is $\left(\sum_{i=1}^{n} b_i^{\alpha}\right)^{-1}$.

Example 2.4.2 (Acceptance-rejection for bounded C's, part II)**.** Suppose that rather than having each of the C_i be bounded individually, we have that $D = \sum_{i=1}^{N} C_i^{\alpha} \leq b$ a.s. Now let $Z \sim$ Pareto$(a, 1)$ be independent of $(N, Q, C_1, \ldots, C_N)$, and let $f_{Q, C_1, \ldots, C_n | N=n}$ be the conditional density of $(Q, C_1, \ldots, C_n)$ given $N=n$ with respect to P, as in Example 2.4.1. Now note that the conditional density of $(\tilde{Q}, \tilde{C}_1, \ldots, \tilde{C}_n)$ given $\tilde{N}=n$ satisfies

$$
\begin{aligned}
&f_{\tilde{Q}, \tilde{C}_1, \ldots, \tilde{C}_n | \tilde{N}=n}(y, x_1, \ldots, x_n) \, dy \, dx_1 \cdots dx_n \\
&= \frac{E\left[1(Q \in dy, C_1 \in dx_1, \ldots, C_n \in dx_n) D \middle| \, N=n \right]}{E[D | N=n]} \\
&= \frac{E\left[1(Q \in dy, C_1 \in dx_1, \ldots, C_n \in dx_n) b^{-1} D \middle| \, N=n \right]}{E\left[b^{-1} D \middle| \, N=n \right]} \\
&= \frac{E\left[1(Q \in dy, C_1 \in dx_1, \ldots, C_n \in dx_n) 1(Z^a > b/D) \middle| \, N=n \right]}{P(Z^a > b/D | N=n)},
\end{aligned}
$$

where we have used the observation that $P(Z^a > b/D | D) = D/b$ and $P(Z^a > b/D | N=n) = E[D/b | N=n]$. Therefore, after simulating $\tilde{N}=n$, we can obtain $(\tilde{Q}, \tilde{C}_1, \ldots, \tilde{C}_n)$ by generating $Z \sim$ Pareto$(a, 1)$ and $(Q, C_1, \ldots, C_n)$

according to $f_{Q,C_1,\ldots,C_n|N=n}$, independent of each other, and then setting $(\tilde{Q}, \tilde{C}_1, \ldots, \tilde{C}_n) = (Q, C_1, \ldots, C_n)$ if $Z > (b/D)^{1/a}$. The acceptance probability given $\tilde{N} = n$ is $P(Z > (b/D)^{1/a}) = b^{-1}$.

Example 2.4.3. *[A mixture representation]* The change of measure induces a mixture density in the following way. If $\alpha > 0$ is such that $E\left[\sum_{i=1}^{N} C_i^\alpha\right] = 1$, then define the values $\{p_{i,n}, i \leq n, n \in \mathbb{N}\}$ by

$$p_{i,n} = \frac{E[C_i^\alpha | N = n]}{\sum_{j=1}^{n} E[C_j^\alpha | N = n]} \in [0, 1].$$

Then,

$$\begin{aligned}
&\widetilde{P}(\tilde{N} = n, \tilde{Q} \in dy, \tilde{C}_1 \in dx_1, \ldots, \tilde{C}_n \in dx_n) \\
&= E\left[1(N = n, Q \in dy, C_1 \in dx_1, \ldots, C_n \in dx_n) \sum_{i=1}^{N} C_i^\alpha\right] \\
&= P(N = n) \sum_{i=1}^{n} E[1(C_i \in dx_i) C_i^\alpha | N = n] P(Q \in dy, C_j \in dx_j, j \neq i | C_i = x_i, N = n) \\
&= \widetilde{P}(\tilde{N} = n) \sum_{i=1}^{n} p_{i,n} \tilde{f}_{i,n}(x_i)\, dx_i\, P(Q \in dy, C_j \in dx_j, j \neq i | C_i = x_i, N = n),
\end{aligned}$$

where

$$\tilde{f}_{i,n}(x)\, dx = \frac{E[1(C_i \in dx) C_i^\alpha | N = n]}{E[C_i^\alpha | N = n]} = \frac{x^\alpha f_{i,n}(x)\, dx}{E[C_i^\alpha | N = n]}$$

is the tilted marginal density of $\tilde{C}_i$ conditional on $\tilde{N} = n$, while $f_{i,n}$ is the marginal density of C_i conditional of $N = n$ under P.

Suppose now that $\tilde{f}_{i,n}$ specifies a distribution that can be efficiently simulated, and that it is possible to simulate the vector $(Q, C_1, \ldots, C_{i-1}, C_{i+1}, \ldots C_n)$ given $\{C_i = x, N = n\}$ under P. Then, conditional on $\tilde{N} = n$, the tilted vector $(\tilde{Q}, \tilde{C}_1, \ldots, \tilde{C}_n)$ can be simulated by picking $i \in \{1, \ldots, n\}$ according to the distribution $\{p_{i,n} : 1 \leq i \leq n\}$, generating $\tilde{C}_i$ according to $\tilde{f}_{i,n}$, and then generating $\{\tilde{Q}, \tilde{C}_j, j \neq i\}$ according to the conditional distribution of $\{Q, C_j, j \neq i\}$ given $\{C_i, N\}$ under P.

Consider the special case when the $\{C_i\}$ are i.i.d. and N, Q, and $\{C_i\}$ are mutually independent. Then,

$$\widetilde{P}(\tilde{N} = n) = E\left[1(N = n) \sum_{j=1}^{N} C_j^\alpha\right] = \frac{nP(N = n)}{E[N]}, \quad n \geq 1, \tag{2.4.3}$$

since α is such that $E[N]E[C_1^\alpha] = 1$. Hence, under the tilt, $\tilde{N}$ is the sized-biased version of N. Furthermore, $p_{i,n} = 1/n$, $\tilde{f}_{i,n} = \tilde{f}$ and $f_{i,n} = f$ for all i and n and some densities $\tilde{f}$, f. So upon simulating $\tilde{N}$ according to the size-biased distribution, the $\{\tilde{C}_1, \ldots, \tilde{C}_n\}$ can be simulated by picking $i \in \{1, \ldots, n\}$ uniformly at random, simulating

$\tilde{C}_i$ according to

$$\tilde{f}(x) = \frac{x^\alpha f(x)}{E[C_1^\alpha]},$$

and simulating the rest of the $\{\tilde{C}_j : j \neq i\}$ according to f. In this case the distribution of Q is invariant under the tilt.

Having now described three general methods for simulating the generic branching vector $\tilde{\psi}$ under the tilt induced by measure $\widetilde{P}$ for nodes along the spine, we now give some more concrete examples that lead to explicit distributions for both ψ and $\tilde{\psi}$.

Example 2.4.4. *[The branching version of the $M/M/1$ queue]* As mentioned earlier, the special case of (2.0.1) when $N \equiv 1$ corresponds to the Lindley equation satisfied by the single-server queue. In particular, if we choose $X = \chi - \tau$ where χ and τ are exponentially distributed and independent of each other, we obtain the $M/M/1$ queue. This choice of X is known to be closed under the change of measure induced by $\widetilde{P}$, in the sense that it remains a difference of two exponentials (but with different rates). As one would expect, this canonical example for the non-branching case is also valid in the branching one. Specifically, suppose that the $\{C_i : i \geq 1\}$ are i.i.d. and independent of N, with each of the $C_i = e^{\chi_i - \tau_i}$, where the $\{(\chi_i, \tau_i) : i \geq 1\}$ are i.i.d. copies of (χ, τ), with χ and τ exponentially distributed and independent of each other.

Suppose τ has rate λ and χ has rate θ, for which we have

$$f(x) = \frac{\theta\lambda}{\lambda + \theta}\left(x^{\lambda-1}1(x<1) + x^{-\theta-1}1(x \geq 1)\right), \qquad x \in (0,\infty),$$

in Example 2.4.3. Then, we can simulate $(\tilde{N}, \tilde{C}_1, \ldots, \tilde{C}_N)$ under $\widetilde{P}$ by first simulating $\tilde{N}$ according to the size-biased distribution of N (2.4.3), then pick an index $i \in \{1, \ldots, \tilde{N}\}$ uniformly at random, simulate each of the $\{\tilde{C}_j : j \neq i\} \stackrel{\mathcal{D}}{=} \{C_j : j \neq i\}$ through an inversion transform for each of the τ_j and χ_j, and then simulate $\tilde{C}_i = e^{\chi_i - \tau_i}$ according to the tilted density given by:

$$\tilde{f}(x) = \frac{x^\alpha f(x)}{E[C_1^\alpha]} = \frac{(\theta-\alpha)(\lambda+\alpha)}{\theta+\lambda}\left(x^{\alpha+\lambda-1}1(x<1) + x^{-(\theta-\alpha)-1}1(x \geq 1)\right), \qquad x \in (0,\infty),$$

which corresponds to simulating $\chi_i \sim \text{Exponential}(\theta - \alpha)$ and $\tau_i \sim \text{Exponential}(\lambda + \alpha)$, independent of each other.

Example 2.4.5 (Identical C's)**.** We now give three examples for which $C_i \equiv C$ for all $i \geq 1$.

(a) Suppose Q, N, C are mutually independent, $C \sim \text{Pareto}(a, b)$ with shape a and scale b, and $E[N] < b^{-\alpha}$, where the Cramér-Lundberg root α solves $\alpha = (1 - E[N]b^\alpha)a$. Then under $\widetilde{P}$, $\tilde{Q} \stackrel{\mathcal{D}}{=} Q$ is invariant and remains independent of $(\tilde{N}, \tilde{C})$, the law of $\tilde{N}$ is the sized-biased distribution given by (2.4.3), and $\tilde{C}$ is again Pareto independent of $\tilde{N}$ and $\tilde{Q}$, but with shape $a - \alpha$ and scale b.

(b) Suppose Q is independent of (N, C), $C \sim$ Exponential(λ), and conditional on C, $N \sim$ Poisson$(C) + 1$. Then after tilting, the law of $\tilde{Q}$ remains invariant, $\tilde{Q}$ remains independent of $\tilde{N}$ and $\tilde{C}$, $\tilde{N}$ has mass function

$$\widetilde{P}(\tilde{N} = n) = \frac{n\lambda\Gamma(n+\alpha)}{(n-1)!(\lambda+1)^{n+\alpha}}, \quad n \geq 1,$$

and conditional on $\tilde{N}$, we have $\tilde{C} \sim$ Gamma$(\tilde{N} + \alpha, \lambda + 1)$.

(c) Suppose that $Q \sim$ Gamma$(2, \beta)$, with shape 2 and rate β, and $N \sim$ Geometric$(1/2)$ with support on $\mathbb{N}_+$, are independent, and conditional on (N, Q), $C \sim$ Gamma$(N + 1, 2Q)$. Under the tilt, $\tilde{Q} \sim$ Gamma$(2 - \alpha, \beta)$, conditional on $\tilde{Q}$, $\tilde{C} \sim$ Gamma$(\alpha + 2, \tilde{Q})$, and conditional on $(\tilde{Q}, \tilde{C})$, $\tilde{N} \sim$ Poisson$(\tilde{Q}\tilde{C}) + 1$.

Example 2.4.6. *[C's on the N-simplex]* Let $B \sim$ Gamma(a, b), with shape a and rate b, let N have an arbitrary distribution that is independent of B, and let N and $(\beta_1, \ldots, \beta_N)$ be such that

$$\sum_{i=1}^{N} \beta_i = 1.$$

For example, conditional on N, $(\beta_1, \ldots, \beta_N) \sim$ Dirichlet$(\boldsymbol{\theta})$ for some concentration parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$, i.e., each β_i has a marginal Beta $\left(\theta_i, \sum_{k=1}^{N} \theta_k - \theta_i\right)$ distribution. Then let α be such that $E[B^\alpha] = 1$, let $C_i = B\beta_i^{1/\alpha}$ for $1 \leq i \leq N$, and let Q be arbitrarily distributed independent of everything else. Then α is the Cramér-Lundberg root since

$$E\left[\sum_{i=1}^{N} C_i^\alpha\right] = E\left[\sum_{i=1}^{N} B^\alpha \beta_i\right] = E[B^\alpha] = 1.$$

Under $\widetilde{P}$ the vector $\tilde{\boldsymbol{\psi}} = (\tilde{N}, \tilde{Q}, \tilde{C}_1, \tilde{C}_2, \ldots)$ remains in the same family of distributions, i.e., the marginal laws of $\tilde{N}$ and $\tilde{Q}$ are invariant, B is tilted to $\tilde{B} \sim$ Gamma$(a + \alpha, b)$, and the $\tilde{C}_i$ are constructed in the same way using the same β_i. In this case, it is the particular dependence of N on $\{C_i\}$ that ensures the invariance of $\tilde{N}$, since

$$\widetilde{P}(\tilde{N} = n) = E\left[1(N = n) \sum_{i=1}^{N} C_i^\alpha\right] = E[1(N = n)B^\alpha] = P(N = n).$$

Figure 2.2 shows how $\{S_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{T}\}$ behaves under $\widetilde{P}$. In this simulated walk, $N \sim$ Bernoulli$(1/4) + 1$ independent of $\{C_i\}$, which are exponentiated i.i.d. differences of exponentials as in Example 2.4.4 with $\theta = 5$ and $\lambda = 1/4$, and $Q \equiv 1$. On the tilted branch under $\widetilde{P}$, N has the size-biased distribution Bernoulli$(6/7) + 1$.

2.4.2 Numerical Experiments

Here we implement two examples of the importance sampling algorithm. The first is the branching M/M/1 queue of Example 2.4.4, in which we let χ have rate 5 and τ have rate $1/4$, and we let N be a truncated Poisson random variable

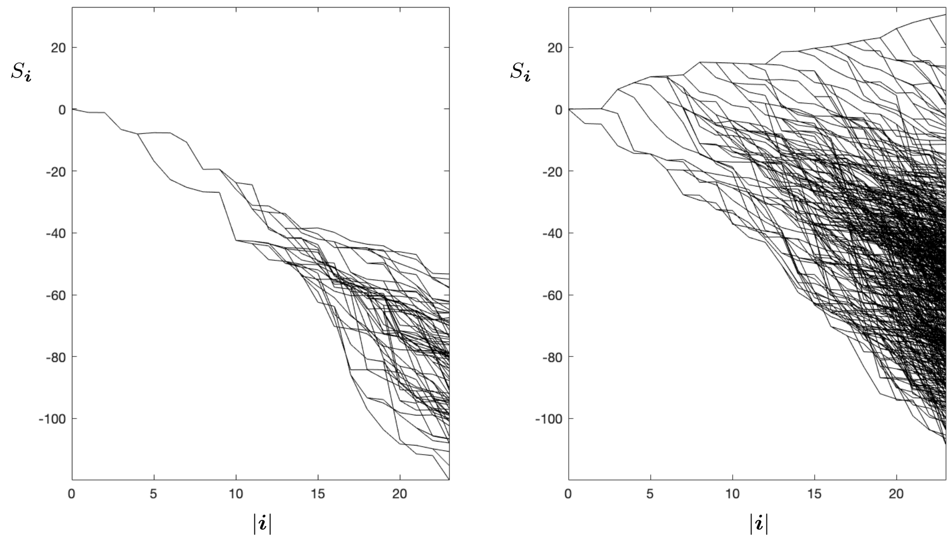


Figure 2.2: A branching random walk (without a perturbation) simulated under both P (left) and $\widetilde{P}$ (right).

with mean 2, i.e. $N \stackrel{\mathcal{D}}{=} K|K > 0$, where $K \sim$ Poisson(2). In this case $\alpha = 4.374$, and we include a perturbation $Y \sim$ Exponential(9) independent of $(N, \{C_i\})$, so that Q is a Pareto random variable with enough moments to ensure $E\left[Q^{2\alpha}\right] < \infty$ and provide our estimator with bounded relative error (see Lemma 2.4.2). Under the tilt, N has its size-biased distribution Poisson$(2) + 1$, and the $\{C_i\}$ are simulated as described in Example 2.4.3, by picking one uniformly at random and applying an exponential tilt. Since Q is independent of $(N, \{C_i\})$, we use the estimator in (2.4.2) (See Remarks 2.1.2(a) and 2.4.1). In Table 2.2 we show the numerical results, which include for a range of t values the sample average $\bar{Z}(t)$ based on 10,000 copies of $Z(t)$ and the standard error in the estimate. Additionally, we give the average tree generation $\tau(t) + 1$ at which the algorithm terminates, the value of t/μ for comparison, the average time in seconds to generate one copy of $Z(t)$, and the fraction of the estimates that are nonzero (i.e. the fraction of iterations in which the algorithm terminated on the chosen path).

Figure 2.3a shows a plot of $\log \bar{Z}(t)$ compared with the tail asymptotic $\log(He^{-\alpha t})$ over the range of values in the table, where H is computed using the population dynamics algorithm [86]. As can be seen, the distribution becomes indistinguishable from the tail asymptotic somewhere in the middle of this range. The terminal generation

Table 2.2: Numerical results for the branching M/M/1 queue, sample size 10,000

Branching $M/M/1$ queue: $\alpha = 4.374$, $\mu = 1.383$						
t	$\bar{Z}(t)$	Std. Err.	t/μ	Terminal gen.	Time	Prop. nonzero
0.5	0.037774	0.001241	0.36	1.39	0.002610	0.967
1	0.003025	0.000123	0.72	1.78	0.007702	0.980
1.5	0.000354	1.07147e-05	1.08	2.16	0.017536	0.983
2	3.90110e-05	1.43477e-06	1.45	2.52	0.029310	0.983
2.5	4.11873e-06	1.16323e-07	1.81	2.90	0.065747	0.985

Table 2.3: Numerical results for C's on the N-simplex, sample size 10,000

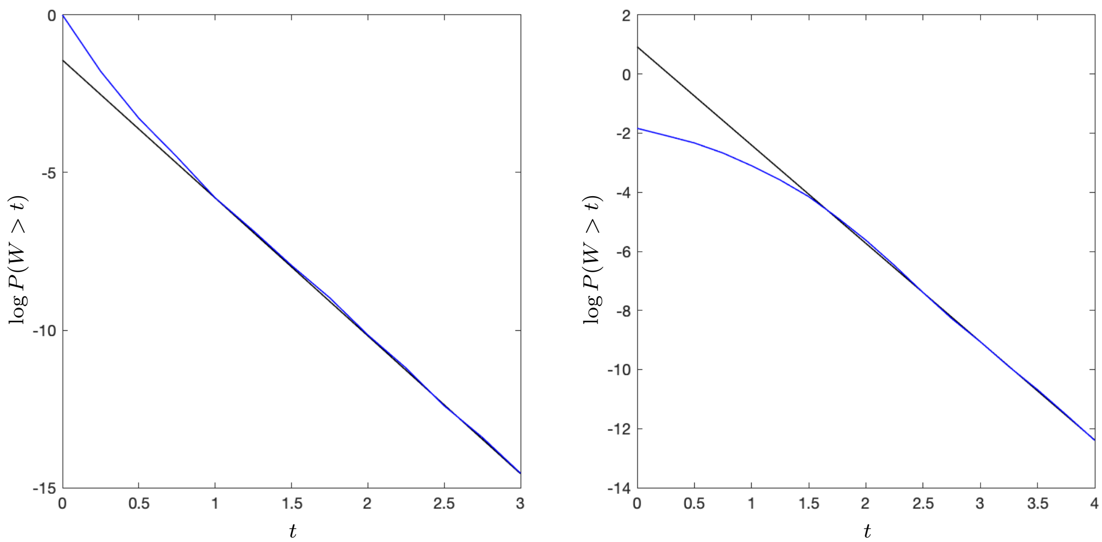
C's on the N-simplex: $\alpha = 3.328$, $\mu = 0.995$						
t	$\bar{Z}(t)$	Std. Err.	t/μ	Terminal gen.	Time	Prop. nonzero
1.5	0.015785	0.000166	1.51	0.33	0.000235	0.998
2	0.003611	3.51666e-05	2.01	0.78	0.000311	0.994
2.5	0.000613	6.60663e-06	2.51	1.33	0.000439	0.994
3	0.000116	1.21042e-06	3.01	1.84	0.000671	0.994
3.5	2.29240e-05	2.35959e-07	3.52	2.33	0.001058	0.992

of each estimator does not converge to t/μ as quickly; the terminal generations listed are greater than t/μ despite the perturbation $Q \geq 1$ which in this case can only cause the process $S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}$ to reach the level t earlier than the random walk $S_{\boldsymbol{i}}$. The large fraction of estimates that are nonzero in this example indicates that in almost all iterations, the level t was first reached on the spine of the tree. In the case of i.i.d. C's, only one offspring of each node on the spine is chosen for the tilt, and this titled branch is then the most likely to be the next step in the spine, making the event that a path off the spine hits t first unlikely.

The second experiment is for C's on the N-simplex, as in Example 2.4.6. We choose $B \sim \text{Gamma}(1/4, 1)$, set α so that $E[B^\alpha] = 1$, and then let $C_i = B\beta_i^{1\alpha}$, where the $\{\beta_i : 1 \leq i \leq N\}$ are Dirichlet$(1, \ldots, 1)$ random variables conditional on N. We let N be uniform over $\{1, 2, 3\}$ independently of B, and we take $Q = 2B$, so that the perturbation is positively correlated with the $\{C_i\}$. As noted in Example 2.4.6, the change of measure tilts B, the distribution of N is the same, and $(Q, \{C_i\})$ is generated the same way through B and $\{\beta_i\}$. Because of the dependence between Q and $\{C_i\}$, the estimator we use is in its most general form (2.4.1). Table 2.3 gives the numerical results for a range of t values and a sample size of 10,000 for each, and Figure 2.3b is a plot of $\log \bar{Z}(t)$ and $\log(He^{-\alpha t})$, where again H is obtained using the population dynamics algorithm.

Again in this experiment, almost every estimator terminated on the spine of the tree. Unlike in the i.i.d. $\{C_i\}$ case, the tilted B influences the C's of every offspring of a node on the spine, so one might expect a significant percentage of estimators to terminate on nodes off the spine but which have the spine in their recent ancestry. We do not see this likely because $Q = 2B$ also is made larger by the tilt but only on the spine, and it seems that it is the perturbation $Y_{\boldsymbol{i}}$ which is causing the process $S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}$ to reach the level t in almost every iteration. This tilted perturbation is likely also causing the terminal tree generation to be smaller than t/μ for these values of t, as it is greater than zero with high probability.

In each example, there is a range of t values in the tail of the distribution for which $P(W > t)$ is not yet indistinguishable from the asymptotic behavior. The importance sampling algorithm presented here provides an efficient method of simulating tail probabilities for these intermediate values of t before the asymptotic behavior dominates.



(a) Log probabilities for the branching $M/M/1$ queue, $H = 0.2390$.

(b) Log probabilities for C's on the N-simplex, $H = 2.5180$.

Figure 2.3: Log probability plots estimated two ways: black lines are the logs of asymptotic approximations $He^{-\alpha t}$ and blue lines are the logs of the estimates $\bar{Z}(t)$.

2.5 Proofs

In this section we provide the proofs to all of our results. To ease its reading, we start first with the proofs of Lemmas 2.1.1 and 2.1.2, which describe the distribution of the tree $\mathcal{T}$ under $\widetilde{P}$. We then give the proof of our main theoretical result and its corollary, Theorem 2.3.1 and Corollary 2.3.1, followed by the proofs of Lemmas 2.4.1 and 2.4.2 which are related to our importance sampling estimator. Finally, we end the paper with a short proof of Theorem 1 in [6] for the non-branching, bounded Q case (Theorem 2.5.1). Throughout the remainder of the paper we assume that Condition 2.0.1 holds for some $\alpha > 0$.

2.5.1 The Distribution of $\mathcal{T}$ Under $\widetilde{P}$

We start with the proof of Lemma 2.1.1, which provides the distribution of the generic branching vectors defining the weighted tree $\mathcal{T}$. The distribution for vectors on the spine is different under $\widetilde{P}$ and P, whereas that of vectors off the spine remains the same.

Proof of Lemma 2.1.1. We will start by deriving an expression for the joint distribution of the vectors $\boldsymbol{\psi}_{\boldsymbol{i}}$ along then spine. To do this, fix $\boldsymbol{i} \in \mathbb{N}_+^k$ and let $B_0, B_1, \ldots, B_k \subseteq \mathbb{N} \times \mathbb{R}^\infty$ be measurable sets. Next, note that the event $\{\boldsymbol{\psi}_{\boldsymbol{i}|r} \in B_r\,, r = 0, \ldots, k,\ \boldsymbol{J}_k = \boldsymbol{i}\}$ is measurable with respect to $\mathcal{G}_{k+1}$, and therefore,

$$\widetilde{P}\left(\boldsymbol{\psi}_{\boldsymbol{i}|r} \in B_r\,, r = 0, \ldots, k,\ \boldsymbol{J}_k = \boldsymbol{i}\right)$$

$$
\begin{aligned}
&= E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}|r} \in B_r\,, r = 0, \ldots, k,\ \boldsymbol{J}_k = \boldsymbol{i}\right) L_{k+1}\right] \\
&= E\left[E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}|r} \in B_r\,, r = 0, \ldots, k,\ \boldsymbol{J}_k = \boldsymbol{i}\right) L_{k+1}\middle|\, \mathcal{G}_k\right]\right] \\
&= E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}|r} \in B_r\,, r = 0, \ldots, k-1,\ \boldsymbol{J}_k = \boldsymbol{i}\right) L_k E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}} \in B_k\right) D_{\boldsymbol{i}}\middle|\, \mathcal{G}_k\right]\right] \\
&= E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}|r} \in B_r\,, r = 0, \ldots, k-1,\ \boldsymbol{J}_k = \boldsymbol{i}\right) L_k\right] E\left[1(\boldsymbol{\psi} \in B_k) D\right] \\
&= E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}|r} \in B_r\,, r = 0, \ldots, k-2,\ \boldsymbol{J}_{k-1} = (\boldsymbol{i}|k-1)\right) L_{k-1}\right. \\
&\qquad \left.\times E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}|k-1} \in B_{k-1},\ \boldsymbol{J}_k = \boldsymbol{i}\right) D_{\boldsymbol{J}_{k-1}}\middle|\, \mathcal{G}_{k-1}\right]\right] E\left[1(\boldsymbol{\psi} \in B_k) D\right].
\end{aligned}
$$

Now note that by letting $\boldsymbol{j} = (\boldsymbol{i}|k-1)$ we obtain

$$
\begin{aligned}
&E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}|k-1} \in B_{k-1},\ \boldsymbol{J}_k = \boldsymbol{i}\right) D_{\boldsymbol{J}_{k-1}}\middle|\, \mathcal{G}_{k-1}\right] \\
&= E\left[1(\boldsymbol{\psi}_{\boldsymbol{j}} \in B_{k-1},\ \text{offspring } i_k \text{ of } \boldsymbol{j} \text{ is chosen}) D_{\boldsymbol{j}}\right] \\
&= E\left[1(\boldsymbol{\psi}_{\boldsymbol{j}} \in B_{k-1}, N_{\boldsymbol{j}} \geq i_k) \cdot \frac{C^{\alpha}_{(\boldsymbol{j}, i_k)}}{D_{\boldsymbol{j}}} \cdot D_{\boldsymbol{j}}\right] \\
&= E\left[1(\boldsymbol{\psi} \in B_{k-1},\ N \geq i_k) C^{\alpha}_{i_k}\right].
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}|r} \in B_r\,, r = 0, \ldots, k-1,\ \boldsymbol{J}_k = \boldsymbol{i}\right) L_k\right] \\
&= E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}|r} \in B_r\,, r = 0, \ldots, k-2,\ \boldsymbol{J}_{k-1} = (\boldsymbol{i}|k-1)\right) L_{k-1}\right] E\left[1(\boldsymbol{\psi} \in B_{k-1},\ N \geq i_k) C^{\alpha}_{i_k}\right] \\
&= E\left[1\left(\boldsymbol{\psi}_{\emptyset} \in B_0,\ \boldsymbol{J}_1 = i_1\right) L_1\right] \prod_{r=2}^{k} E\left[1(\boldsymbol{\psi} \in B_{r-1},\ N \geq i_r) C^{\alpha}_{i_r}\right] \\
&= \prod_{r=1}^{k} E\left[1(\boldsymbol{\psi} \in B_r,\ N \geq i_r) C^{\alpha}_{i_r}\right].
\end{aligned}
$$

We conclude that

$$
\widetilde{P}\left(\boldsymbol{\psi}_{\boldsymbol{i}|r} \in B_r\,, r = 0, \ldots, k,\ \boldsymbol{J}_k = \boldsymbol{i}\right) = E\left[1(\boldsymbol{\psi} \in B_k) D\right] \prod_{r=1}^{k} E\left[1(\boldsymbol{\psi} \in B_{r-1},\ N \geq i_r) C^{\alpha}_{i_r}\right]. \tag{2.5.1}
$$

In particular, by setting $B_r = \mathbb{N} \times \mathbb{R}^{\infty}$ for all $r = 0, 1, \ldots, k$, we obtain the first expression in the statement of the lemma, i.e.,

$$
\widetilde{P}(\boldsymbol{J}_k = \boldsymbol{i}) = \prod_{r=1}^{k} E[1(N \geq i_r) C^{\alpha}_{i_r}].
$$

Similarly, by setting $B_k = B$ and $B_r = \mathbb{N} \times \mathbb{R}^\infty$ for all $r = 0, 1, \dots, k-1$, we obtain the third expression:

$$\widetilde{P}(\boldsymbol{\psi}_{\boldsymbol{i}} \in B | \boldsymbol{i} = \boldsymbol{J}_k) = E[1(\boldsymbol{\psi} \in B) D] = E\left[1((N, Q, C_1, C_2, \dots) \in B) \sum_{j=1}^{N} C_j^\alpha\right].$$

To obtain the corresponding expression for nodes off the spine (the second expression in the statement of the lemma) note that the same conditioning approach used for a node on the spine gives, for any $\boldsymbol{i} \in \mathbb{N}_+^k$ and any measurable $B \subseteq \mathbb{N} \times \mathbb{R}^\infty$,

$$\begin{aligned}
\widetilde{P}\left(\boldsymbol{\psi}_{\boldsymbol{i}} \in B; \boldsymbol{J}_k \neq \boldsymbol{i}\right) &= E\left[E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}} \in B; \boldsymbol{J}_k \neq \boldsymbol{i}\right) L_{k+1} \middle| \mathcal{G}_k\right]\right] \\
&= E\left[L_k E\left[1\left(\boldsymbol{\psi}_{\boldsymbol{i}} \in B; \boldsymbol{J}_k \neq \boldsymbol{i}\right) D_{\boldsymbol{J}_k} \middle| \mathcal{G}_k\right]\right] \\
&= E\left[L_k E[1(\boldsymbol{\psi}_{\boldsymbol{i}} \in B) | \mathcal{G}_k] E\left[1\left(\boldsymbol{J}_k \neq \boldsymbol{i}\right) D_{\boldsymbol{J}_k} \middle| \mathcal{G}_k\right]\right] \\
&= P(\boldsymbol{\psi} \in B) E\left[1(\boldsymbol{J}_k \neq \boldsymbol{i}) L_{k+1}\right] \\
&= P(\boldsymbol{\psi} \in B) \widetilde{P}(\boldsymbol{J}_k \neq \boldsymbol{i}).
\end{aligned}$$

Therefore,

$$\widetilde{P}\left(\boldsymbol{\psi}_{\boldsymbol{i}} \in B | \boldsymbol{J}_k \neq \boldsymbol{i}\right) = P(\boldsymbol{\psi} \in B) = P((N, Q, C_1, C_2, \dots) \in B).$$

The conditional independence of the vectors $\{\boldsymbol{\psi}_{\boldsymbol{i}} : \boldsymbol{i} \in A_k\}$ given $\mathcal{G}_{k-1}$ follows from the branching property under P. This completes the proof. □

We now give the proof of Lemma 2.1.2, which gives the distribution of the random walk defined by the nodes along the spine.

Proof of Lemma 2.1.2. Recall that $\hat{X}_k = X_{\boldsymbol{J}_k} = \log C_{\boldsymbol{J}_k}$ and $\xi_k = Y_{\boldsymbol{J}_k} = \log Q_{\boldsymbol{J}_k}$. By conditioning on all possible paths that could be chosen to define $\boldsymbol{J}_k$ we obtain

$$\begin{aligned}
&\widetilde{P}\left(\hat{X}_1 \leq x_1, \dots, \hat{X}_k \leq x_k,\, \xi_k \leq y\right) \\
&= \sum_{\boldsymbol{i} \in \mathbb{N}_+^k} \widetilde{P}\left(\hat{X}_1 \leq x_1, \dots, \hat{X}_k \leq x_k,\, \xi_k \leq y,\, \boldsymbol{J}_k = \boldsymbol{i}\right) \\
&= \sum_{\boldsymbol{i} \in \mathbb{N}_+^k} \widetilde{P}\left(C_{\boldsymbol{i}|1} \leq e^{x_1}, \dots, C_{\boldsymbol{i}|k} \leq e^{x_k},\, Q_{\boldsymbol{i}|k} \leq e^{y},\, \boldsymbol{J}_k = \boldsymbol{i}\right) \\
&= \sum_{\boldsymbol{i} \in \mathbb{N}_+^k} \widetilde{P}(Q \leq e^y) \prod_{r=1}^{k} E\left[1(C_{i_r} \leq e^{x_r}, N \geq i_r) C_{i_r}^\alpha\right] \\
&= \widetilde{P}(Q \leq e^y) \sum_{i_1=1}^{\infty} E\left[1(C_{i_1} \leq e^{x_1},\, N \geq i_1) C_{i_1}^\alpha\right] \cdots \sum_{i_k=1}^{\infty} E\left[1(C_{i_k} \leq e^{x_k},\, N \geq i_k) C_{i_k}^\alpha\right]
\end{aligned}$$

$$= E\left[1(Q \leq e^y)\sum_{i=1}^N C_i^\alpha\right]\prod_{r=1}^k G(x_r),$$

where in the third equality we used (2.5.1) and the independence of $Q_{\boldsymbol{i}|k}$ and $\boldsymbol{\psi}_{\boldsymbol{i}|k-1}$. To compute the mean of the $\hat{X}_i$'s note that

$$\widetilde{E}\left[|\hat{X}_1|\right] = \int_{-\infty}^\infty |x| G(dx) = \int_{-\infty}^\infty |x| E\left[\sum_{i=1}^N 1(\log C_i \in dx) C_i^\alpha\right] = E\left[\sum_{i=1}^N C_i^\alpha |\log C_i|\right].$$

Now choose $0 < \beta < \alpha$ such that $E\left[\sum_{i=1}^N C_i^\beta\right] < 1$ and note that since $\log^- C_i = 0$ when $C_i > 1$,

$$\widetilde{E}\left[\hat{X}_1^-\right] = E\left[\sum_{i=1}^N C_i^\alpha \log^- C_i\right] \leq \sup_{0\leq x\leq 1} x^{\alpha-\beta}|\log x| \widetilde{E}\left[\sum_{i=1}^N C_i^\beta\right] < \infty.$$

For the positive part note that by Condition 2.0.1(a) we have $E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right] \in (0,\infty)$, and therefore

$$\widetilde{E}\left[\hat{X}_1^+\right] = E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right] + \widetilde{E}\left[\hat{X}_1^-\right] < \infty.$$

Finally, since both $\widetilde{E}\left[\hat{X}_1^-\right]$ and $\widetilde{E}\left[\hat{X}_1^+\right]$ are finite, we have $\widetilde{E}\left[\hat{X}_1\right] = E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right] \in (0,\infty)$. □

2.5.2 Proof of Theorem 2.3.1

We now move on to the proof of our main theorem, which is obtained by applying renewal theory to compute the limit of

$$\widetilde{E}\left[1(\boldsymbol{J}_{\tau(t)} = \gamma(t)) e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right].$$

The proof will rely on several preliminary results, the first of which establishes the almost sure finiteness of $\tau(t)$ under $\widetilde{P}$ (Lemma 2.1.3). Throughout this subsection we assume all parts of Condition 2.0.1 hold.

Proof of Lemma 2.1.3. The sequence $\{V_k\}$ satisfies the strong law of large numbers

$$V_n/n \to \mu > 0 \quad \widetilde{P}\text{-a.s.}$$

Note that Condition 2.0.1(c) ensures that ξ_0 has the same support under $\widetilde{P}$ that under P, hence, since $P(Q > 0) > 0$, there must exist an $\epsilon > 0$ such that $P(Q > \epsilon) > 0$, and therefore, $\widetilde{P}(Q > \epsilon) > 0$. Hence, under $\widetilde{P}$, the random times $T_0 = \inf\{i \geq 0 : \xi_i > \log\epsilon\}$ and $T_{k+1} = \inf\{i > T_k : \xi_i > \log\epsilon\}$ are finite $\widetilde{P}$-a.s. for all $k \geq 0$; moreover,

$T_k \to +\infty$ $\widetilde{P}$-a.s. as $k \to \infty$. Focusing on subsequences along the indexes $\{T_k : k \geq 0\}$ gives

$$\liminf_{k\to\infty} \frac{V_{T_k} + \xi_{T_k}}{T_k} \geq \liminf_{k\to\infty} \frac{V_{T_k} + \log \epsilon}{T_k} = \mu \qquad \widetilde{P}\text{-a.s}$$

Since

$$\sup_{n\geq 0} (V_n + \xi_n) \geq \sup_{k\geq 0} (V_{T_k} + \xi_{T_k}) \qquad \widetilde{P}\text{-a.s.},$$

and $\widetilde{P}\left(\sup_{k\geq 0}(V_{T_k} + \xi_{T_k}) > y\right) = 1$ for all y, the result follows. □

The next thing we need to establish is an upper bound for $\widetilde{E}\left[e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right]$, since this quantity will appear in various places throughout the proof of Theorem 2.3.1.

Lemma 2.5.1. *Under Condition 2.0.1, we have for any $t \in \mathbb{R}$,*

$$\widetilde{E}\left[e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right] \leq \sum_{n=0}^{\infty} \widetilde{E}\left[u(t - V_n)\right], \tag{2.5.2}$$

where $u(x) = e^{\alpha x} P\left(Y > x\right)$ is d.R.i. on $\mathbb{R}$. Moreover,

$$\limsup_{t\to\infty} \widetilde{E}\left[e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right] \leq \frac{E[Q^\alpha]}{\alpha\mu}. \tag{2.5.3}$$

Proof. Note that

$$\begin{aligned}
&\widetilde{E}\left[e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right] \\
&= \widetilde{E}\left[e^{\alpha t} D^{-1}_{\boldsymbol{J}_0} 1(\xi_0 > t)\right] + \sum_{n=1}^{\infty} \widetilde{E}\left[1\left(\max_{0\leq k\leq n-1} V_k + \xi_k \leq t < V_n + \xi_n\right) e^{-\alpha(V_n - t)} D^{-1}_{\boldsymbol{J}_n}\right] \\
&= \widetilde{E}\left[e^{\alpha t} D^{-1}_{\boldsymbol{J}_0} 1(\xi_0 > t)\right] \\
&\quad + \sum_{n=1}^{\infty} \widetilde{E}\left[1\left(\max_{0\leq k\leq n-1} V_k + \xi_k \leq t\right) e^{-\alpha(V_n - t)} \widetilde{E}\left[1\left(V_n + \xi_n > t\right) D^{-1}_{\boldsymbol{J}_n} \middle| \mathcal{G}_n\right]\right] \\
&\leq \sum_{n=0}^{\infty} \widetilde{E}\left[u(t - V_n)\right],
\end{aligned}$$

where

$$u(x) = e^{\alpha x} \widetilde{E}\left[1(\xi_0 > x) D^{-1}_{\boldsymbol{J}_0}\right] = e^{\alpha x} P(Y > x).$$

We will now show that u is d.R.i. on $(-\infty, \infty)$, and we start by proving that u is integrable. To see this note that

$$\int_{-\infty}^{\infty} u(x)dx = E\left[\int_{-\infty}^{\infty} e^{\alpha x} 1(Y > x)\, dx\right] = \frac{E[Q^\alpha]}{\alpha} < \infty.$$

Now note that for any $h > 0$,

$$\sum_{n=-\infty}^{\infty} \sup_{y \in (nh,(n+1)h]} u(y) \leq \sum_{n=-\infty}^{\infty} e^{\alpha(n+1)h} P(Y > nh)$$
$$\leq \sum_{n=-\infty}^{\infty} \int_{(n-1)h}^{nh} e^{2\alpha h} e^{\alpha x} P(Y > x) dx$$
$$= e^{2\alpha h} \int_{-\infty}^{\infty} u(x) dx < \infty,$$

so by Lemma 1.1.1(i), u is d.R.i.

To complete the proof use the two-sided renewal theorem (see Corollary 1.1.3 in the Introduction or Theorem 4.2 in [9]), to obtain

$$\lim_{t \to \infty} \sum_{n=0}^{\infty} \widetilde{E}\left[u(t - V_n)\right] = \frac{1}{\mu} \int_{-\infty}^{\infty} u(x) dx = \frac{E[Q^\alpha]}{\alpha \mu}. \qquad \square$$

Corollary 2.5.2. *There exists a constant $0 < B < \infty$ such that for any $t \in \mathbb{R}$,*

$$e^{\alpha t} P(W > t) \leq B\, e^{-\alpha(-t)^+}.$$

Proof. By (2.5.3), there is a $t_0 > 0$ such that $\sup_{t \geq t_0} \widetilde{E}\left[e^{-\alpha(V_{\tau(t)} - t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right] \leq 2E[Q^\alpha]/(\alpha\mu) < \infty$. Since for $t \leq 0$ we have the trivial bound $e^{\alpha t} P(W > t) \leq e^{-\alpha(-t)^+}$, and for $t \geq t_0$ we have $e^{\alpha t} P(W > t) \leq \widetilde{E}\left[e^{-\alpha(V_{\tau(t)} - t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right] \leq 2E[Q^\alpha]/(\alpha\mu)$, we can take $B = \max\{e^{\alpha t_0}, 2E[Q^\alpha]/(\alpha\mu)\}$ to obtain the stated inequality. $\square$

We are now ready to move on to the application of the Markov renewal theorem. Let $\mathcal{S}^{(m)}$ denote the state space of weighted trees of height m having a path identified as its spine, and define the Markov chain $\{M_n^{(m)} : n \geq m\}$ in $\mathcal{S}^{(m)}$ as follows:

$$M_m^{(m)} = \left\{ \boldsymbol{\psi_i} : \boldsymbol{i} \in \bigcup_{k=0}^{m-1} A_k \right\} \cup \{\boldsymbol{J}_0, \boldsymbol{J}_1, \dots, \boldsymbol{J}_m\},$$

and

$$M_{m+n}^{(m)} = \left\{ \boldsymbol{\psi_i} : \boldsymbol{i} \in \mathcal{T}_n^{(m)} \right\} \cup \{\boldsymbol{J}_n, \boldsymbol{J}_{n+1}, \dots, \boldsymbol{J}_{n+m}\}$$

for $n > 0$. Recall that

$$\mathcal{T}_n^{(m)} = \bigcup_{k=0}^{m-1} A_{k,\boldsymbol{J}_n} \qquad \text{and} \qquad A_{k,\boldsymbol{i}} = \{(\boldsymbol{i},\boldsymbol{j}) \in \mathcal{T} : |\boldsymbol{j}| = k\}.$$

Note that with this notation we can write

$$
\begin{aligned}
&\widetilde{E}\left[1(\boldsymbol{J}_{\tau(t)} = \gamma(t))e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right] \\
&= \sum_{n=0}^{m-2} \widetilde{E}\left[1(|\gamma(t)| = n, \boldsymbol{J}_{\tau(t)} = \gamma(t))e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right] \\
&\quad + \widetilde{E}\left[1(|\gamma(t)| \geq m-1, \boldsymbol{J}_{\tau(t)} = \gamma(t))e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right] \\
&= \sum_{n=0}^{m-2} \widetilde{E}\left[1(\boldsymbol{J}_n = \gamma(t))e^{-\alpha(V_n-t)}D^{-1}_{\boldsymbol{J}_n}\right] + \widetilde{E}\left[K\left(M_m^{(m)}, t\right)\right],
\end{aligned}
$$

where

$$
K\left(M_m^{(m)}, t\right) := \widetilde{E}\left[1(|\gamma(t)| \geq m-1, \gamma(t) = \boldsymbol{J}_{\tau(t)})e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\boldsymbol{J}_{\tau(t)}} \middle| M_m^{(m)}\right].
$$

The key idea behind the proof of Theorem 2.3.1 is that $\widetilde{E}\left[K\left(M_m^{(m)}, t\right)\right]$ can be analyzed using the Markov renewal theorem (Theorem 2.1 in [4]). However, the use of this theorem is not immediate, since, as mentioned earlier, the perturbations make it difficult to identify clear regeneration points. In comparison, when the perturbations are not random (i.e., Q's are constant), it suffices to focus on the generations where the ladder heights of the random walk $\{V_k : k \geq 0\}$ occur, since the crossing of level t can only happen at these times. To solve this problem, our approach relies on the observation that although the crossing of level t does not need to coincide with a ladder height of $\{V_k : k \geq 0\}$, and the perturbed (branching) random walk does not regenerate when the ladder heights of $\{V_k + \xi_k : k \geq 0\}$ occur, we can ignore the effect of the perturbations by looking at a long enough stretch of the history of the branching random walk along its spine. This history is what the Markov chain $\{M_{m+n}^{(m)} : n \geq 0\}$ includes.

Our first technical result in this section will define a function that will appear in the derivation of a lower bound for $\widetilde{E}\left[K\left(M_m^{(m)}, t\right)\right]$. Before we state it, we will need to define the following random variables. We use

$$
W_{\boldsymbol{i}} \doteq \bigvee_{r=0}^{\infty} \bigvee_{(\boldsymbol{i},\boldsymbol{j}) \in A_{|\boldsymbol{i}|+r}} \left(S_{(\boldsymbol{i},\boldsymbol{j})} - S_{\boldsymbol{i}} + Y_{(\boldsymbol{i},\boldsymbol{j})}\right), \qquad \boldsymbol{i} \in \mathcal{T}, \tag{2.5.4}
$$

to define the maximum of the perturbed branching random walk rooted at node $\boldsymbol{i}$. Note that if $\boldsymbol{i}$ is not part of the spine, then $W_{\boldsymbol{i}}$ has the same distribution under both $\widetilde{P}$ and P. Now use the $W_{\boldsymbol{i}}$ to define

$$
Z_k \doteq \xi_k \vee \max_{(\boldsymbol{J}_k, i) \neq \boldsymbol{J}_{k+1}} \left(S_{(\boldsymbol{J}_k,i)} - V_k + W_{(\boldsymbol{J}_k,i)}\right), \qquad k \geq 0.
$$

Note that the Z_k is the maximum of the perturbation at the spine node $\boldsymbol{J}_k$ and all the branching random walks that are rooted at sibling nodes of $\boldsymbol{J}_k$. Intuitively, since under $\widetilde{P}$ only the spine has positive drift, the probability that any path that coalesces with the spine outside of $\mathcal{T}_k^{(m)}$ has a very small chance of ever reaching level t for sufficiently large t and m. The function h_m will be used to quantify how rare this event is.

Lemma 2.5.3. *Under Condition 2.0.1, the function*

$$h_m(x) = \widetilde{E}\left[1(Z_0 > x)e^{-\alpha(V_{m+1}-x)^+}\right]$$

is d.R.i. on $\mathbb{R}$ for any $m \geq 2$, and satisfies

$$\int_{-\infty}^{\infty} h_m(x)\,dx = \widetilde{E}\left[\frac{e^{-\alpha(V_{m+1}-Z_0)^+}}{\alpha} + (Z_0 - V_{m+1})^+\right] < \infty.$$

Proof. We start by showing that h_m is integrable, for which we note that

$$\begin{aligned}\int_{-\infty}^{\infty} h_m(x)\,dx &= \widetilde{E}\left[\int_{-\infty}^{Z_0} e^{-\alpha(V_{m+1}-x)^+}dx\right] = \widetilde{E}\left[\int_{V_{m+1}-Z_0}^{\infty} e^{-\alpha y^+}dy\right]\\ &= \widetilde{E}\left[\int_{V_{m+1}-Z_0}^{(V_{m+1}-Z_0)^+} dy + \int_{(V_{m+1}-Z_0)^+}^{\infty} e^{-\alpha y}dy\right]\\ &= \widetilde{E}\left[(Z_0 - V_{m+1})^+ + \frac{e^{-\alpha(V_{m+1}-Z_0)^+}}{\alpha}\right].\end{aligned}$$

To see that $\widetilde{E}\left[(Z_0 - V_{m+1})^+\right]$ is finite first note that

$$\widetilde{E}\left[(Z_0 - V_{m+1})^+\right] \leq \widetilde{E}\left[Z_0^+ + (-V_{m+1})^+\right] \leq \widetilde{E}\left[Z_0^+\right] + (m+1)\widetilde{E}\left[(-V_1)^+\right].$$

By Lemma 2.1.2 we have $\widetilde{E}\left[(-V_1)^+\right] = \widetilde{E}\left[\hat{X}_1^-\right] < \infty$. For $\widetilde{E}\left[Z_0^+\right]$, recall $Q = e^{\xi_0}$ and write

$$\begin{aligned}\widetilde{E}\left[Z_0^+\right] &= \int_0^{\infty} \widetilde{E}\left[\widetilde{P}\left(\xi_0 \vee \bigvee_{i \neq \boldsymbol{J}_1} (\log C_i + W_i) > t \middle| \boldsymbol{\psi}_\emptyset\right)\right] dt\\ &= \int_0^{\infty} \widetilde{E}\left[1(\xi_0 > t) + 1(\xi_0 \leq t)\widetilde{P}\left(\bigvee_{i \neq \boldsymbol{J}_1} 1(\log C_i + W_i > t) \middle| \boldsymbol{\psi}_\emptyset\right)\right] dt\\ &\leq \int_0^{\infty} \widetilde{E}\left[1(\xi_0 > t) + 1\left(Q^\alpha \leq e^{\alpha t} < \sum_{i \neq \boldsymbol{J}_1} C_i^\alpha\right)\right] dt\\ &\quad + \int_0^{\infty} \widetilde{E}\left[1\left(Q^\alpha \vee \sum_{i \neq \boldsymbol{J}_1} C_i^\alpha \leq e^{\alpha t}\right) \sum_{i \neq \boldsymbol{J}_1} \widetilde{P}\left(\log C_i + W_i > t \middle| C_i\right)\right] dt\\ &= \widetilde{E}\left[\xi_0^+ + \left(\frac{1}{\alpha}\log\left(\sum_{i \neq \boldsymbol{J}_1} C_i^\alpha\right) - \xi_0^+\right)^+\right] \end{aligned} \tag{2.5.5}$$

$$+\int_0^\infty \widetilde{E}\left[1\left(Q^\alpha \vee \sum_{i\neq \boldsymbol{J}_1} C_i^\alpha \leq e^{\alpha t}\right)\sum_{i\neq \boldsymbol{J}_1} \overline{F}(t-\log C_i)\right]dt, \tag{2.5.6}$$

where $\overline{F}(x) = P(W > x)$ and (2.5.5) is equal to $\widetilde{E}\left[\frac{1}{\alpha}\log^+\left(Q^\alpha \vee \sum_{i\neq \boldsymbol{J}_1} C_i^\alpha\right)\right]$. By Corollary 2.5.2, there exists a constant $B < \infty$ such that $\overline{F}(x) \leq Be^{-\alpha x}$ for $x \geq 0$. It follows that (2.5.6) is bounded from above by

$$\begin{aligned}
&B\int_0^\infty \widetilde{E}\left[1\left(Q^\alpha \vee \sum_{i\neq \boldsymbol{J}_1} C_i^\alpha \leq e^{\alpha t}\right)\sum_{i\neq \boldsymbol{J}_1} e^{-\alpha(t-\log C_i)}\right]dt \\
&= B\widetilde{E}\left[\sum_{i\neq \boldsymbol{J}_1} C_i^\alpha \int_{\frac{1}{\alpha}\log^+\left(Q^\alpha\vee\sum_{i\neq \boldsymbol{J}_1} C_i^\alpha\right)}^\infty e^{-\alpha t}dt\right] \\
&= \frac{B}{\alpha}\widetilde{E}\left[\sum_{i\neq \boldsymbol{J}_1} C_i^\alpha e^{-\log^+\left(Q^\alpha\vee\sum_{i\neq \boldsymbol{J}_1} C_i^\alpha\right)}\right] \leq \frac{B}{\alpha}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\widetilde{E}\left[Z_0^+\right] &\leq \widetilde{E}\left[\frac{1}{\alpha}\log^+\left(Q^\alpha \vee \sum_{i\neq \boldsymbol{J}_1} C_i^\alpha\right)\right] + \frac{B}{\alpha} \\
&\leq \frac{1}{\alpha}E\left[\sum_{i=1}^N C_i^\alpha \log^+\left(Q^\alpha \vee \sum_{i=1}^N C_i^\alpha\right)\right] + \frac{B}{\alpha} < \infty,
\end{aligned}$$

with finiteness provided by Condition 2.0.1(e).

It remains to show that h_m is d.R.i., for which we note that for any $h > 0$,

$$\begin{aligned}
&\sum_{n=-\infty}^\infty \sup_{y\in(nh,(n+1)h]} h_m(y) \\
&\leq \sum_{n=-\infty}^\infty \widetilde{E}\left[1(Z_0 > nh)e^{-\alpha(V_{m+1}-(n+1)h)^+}\right] \\
&\leq \sum_{n=-\infty}^\infty \int_{(n-1)h}^{nh} \widetilde{E}\left[1(Z_0 > x)e^{-\alpha(V_{m+1}-x-2h)^+}\right]dx \\
&= \widetilde{E}\left[(Z_0 - V_{m+1} + 2h)^+ + \frac{e^{-\alpha(V_{m+1}-2h-Z_0)^+}}{\alpha}\right] < \infty,
\end{aligned}$$

so by Lemma 1.1.1(i), h_m is d.R.i. □

We are now ready to formulate the expression to which we will apply the Markov renewal theorem.

Lemma 2.5.4. *For any $t \in \mathbb{R}$ and $m \geq 2$ we have*

$$\sum_{k=0}^\infty \widetilde{E}\left[g\left(M_{m+k}^{(m)}, t - V_k\right)\right] \geq \widetilde{E}\left[K\left(M_m^{(m)}, t\right)\right]$$

$$\geq \sum_{k=0}^{\infty} \widetilde{E}\left[g\left(M_{m+k}^{(m)}, t-V_k\right)\right] - B\sum_{k=0}^{\infty} \widetilde{E}\left[h_m(t-V_k)\right],$$

where $B < \infty$ is the constant from Corollary 2.5.2,

$$g\left(M_m^{(m)}, t\right) = 1\left(\max_{\boldsymbol{i}\prec \boldsymbol{J}_{m-1}} S_{\boldsymbol{i}} + Y_{\boldsymbol{i}} \leq t < V_{m-1} + \xi_{m-1}\right) e^{-\alpha(V_{m-1}-t)} D_{\boldsymbol{J}_{m-1}}^{-1}, \tag{2.5.7}$$

and

$$h_m(x) = \widetilde{E}\left[1(Z_0 > x) e^{-\alpha(V_{m+1}-x)^+}\right].$$

Proof. Start by noting that

$$\begin{aligned}
K\left(M_m^{(m)}, t\right) &= 1(|\gamma(t)| = m-1, \gamma(t) = \boldsymbol{J}_{\tau(t)}) e^{-\alpha(V_{m-1}-t)} D_{\boldsymbol{J}_{m-1}}^{-1} \\
&\quad + \widetilde{E}\left[1(|\gamma(t)| \geq m, \gamma(t) = \boldsymbol{J}_{\tau(t)}) e^{-\alpha(V_{\tau(t)}-t)} D_{\boldsymbol{J}_{\tau(t)}}^{-1} \,\middle|\, M_m^{(m)}\right] \\
&= g\left(M_m^{(m)}, t\right) \\
&\quad + \widetilde{E}\left[1\left(\Delta_{t,\tau(t)}\right) 1(|\gamma(t)| \geq m, \gamma(t) = \boldsymbol{J}_{\tau(t)}) e^{-\alpha(V_{\tau(t)}-t)} D_{\boldsymbol{J}_{\tau(t)}}^{-1} \,\middle|\, M_m^{(m)}\right],
\end{aligned}$$

where the $\Delta_{t,\tau(t)}$ is the event that $\xi_0 \leq t$ and no subtree rooted at any of the sibling nodes of $\boldsymbol{J}_1$ reaches level t before the spine does. Simply ignoring the indicator $1(\Delta_{t,\tau(t)})$ and regenerating at node $\boldsymbol{J}_1$ yields the inequality

$$\begin{aligned}
\widetilde{E}\left[K\left(M_m^{(m)}, t\right)\right] &\leq \widetilde{E}\left[g\left(M_m^{(m)}, t\right)\right] + \widetilde{E}\left[K\left(M_{m+1}^{(m)}, t-V_1\right)\right] \\
&\leq \sum_{k=0}^{n-1} \widetilde{E}\left[g\left(M_m^{(m)}, t-V_k\right)\right] + \widetilde{E}\left[K\left(M_{m+n}^{(m)}, t-V_n\right)\right].
\end{aligned}$$

To further bound the last expectation, let $s \in \mathbb{R}$, and note that on $\{\gamma(s) = \boldsymbol{J}_{\tau(s)}\}$, no random walk on a path other than the chosen one reaches level s before $V_k + \xi_k$. Hence by ignoring these branches up to level m and restarting the branching process at $\boldsymbol{J}_m$ with initial value V_m, we have that

$$\begin{aligned}
K\left(M_m^{(m)}, s\right) &= \widetilde{E}\left[1(|\gamma(s)| \geq m-1, \gamma(s) = \boldsymbol{J}_{\tau(s)}) e^{-\alpha(V_{\tau(s)}-s)} D_{\boldsymbol{J}_{\tau(s)}}^{-1} \,\middle|\, M_m^{(m)}\right] \\
&\leq \widetilde{E}\left[1(\gamma(s-V_m) = \boldsymbol{J}_{\tau(s-V_m)}) e^{-\alpha(V_{\tau(s-V_m)}-(s-V_m))} \,\middle|\, V_m\right] \\
&= e^{\alpha(s-V_m)} \overline{F}(s-V_m),
\end{aligned}$$

where $\overline{F}(x) = P(W > x)$ and we used (2.1.5). Moreover, by Corollary 2.5.2 we have that $e^{\alpha(s-V_m)}\overline{F}(s - V_m) \leq Be^{-\alpha(V_m - s)^+}$, so we obtain for any $s \in \mathbb{R}$,

$$K\left(M_m^{(m)}, s\right) \leq Be^{-\alpha(V_m - s)^+}. \tag{2.5.8}$$

Now replace V_m with $V_{m+n} - V_n$ and $s = t - V_n$ in (2.5.8) to obtain that

$$\widetilde{E}\left[K\left(M_{m+n}^{(m)}, t - V_n\right)\right] \leq B\widetilde{E}\left[e^{-\alpha(V_{m+n} - t)^+}\right].$$

To obtain a lower bound note that $\Delta_{t,\tau(t)}^c \subseteq \{Z_0 > t\}$, since the event $\{Z_0 > t\}$ states that either $\xi_0 > t$ or at least one of the subtrees rooted at a sibling node of $\boldsymbol{J}_1$ reaches level t at some point (even if this happens after the spine does). Hence, we obtain the following lower bound:

$$\begin{aligned}
&\widetilde{E}\left[K\left(M_m^{(m)}, t\right)\right] \\
&\geq \widetilde{E}\left[g\left(M_m^{(m)}, t\right)\right] + \widetilde{E}\left[K\left(M_{m+1}^{(m)}, t - V_1\right)\right] - \widetilde{E}\left[1(Z_0 > t)K\left(M_{m+1}^{(m)}, t - V_1\right)\right] \\
&\geq \widetilde{E}\left[g\left(M_m^{(m)}, t\right)\right] + \widetilde{E}\left[g\left(M_{m+1}^{(m)}, t - V_1\right)\right] + \widetilde{E}\left[K\left(M_{m+2}^{(m)}, t - V_2\right)\right] \\
&\quad - \widetilde{E}\left[1(Z_1 > t - V_1)K\left(M_{m+2}^{(m)}, t - V_2\right)\right] - \widetilde{E}\left[1(Z_0 > t)K\left(M_{m+1}^{(m)}, t - V_1\right)\right] \\
&\geq \sum_{k=0}^{n-1} \widetilde{E}\left[g\left(M_{m+k}^{(m)}, t - V_k\right)\right] + \widetilde{E}\left[K\left(M_{m+n}^{(m)}, t - V_n\right)\right] \\
&\quad - \sum_{k=0}^{n-1} \widetilde{E}\left[1(Z_k > t - V_k)K\left(M_{m+k+1}^{(m)}, t - V_{k+1}\right)\right].
\end{aligned}$$

To provide a bound for the last sum, note that by replacing V_m with $V_{m+k+1} - V_{k+1}$ and $s = t - V_{k+1}$ in (2.5.8) we obtain

$$\begin{aligned}
\widetilde{E}\left[1(Z_k > t - V_k)K\left(M_{m+k+1}^{(m)}, t - V_{k+1}\right)\right] &\leq \widetilde{E}\left[1(Z_k > t - V_k)Be^{-\alpha(V_{m+k+1} - t)^+}\right] \\
&= B\widetilde{E}\left[h_m(t - V_k)\right],
\end{aligned}$$

where $h_m(x) = \widetilde{E}\left[1(Z_0 > x)e^{-\alpha(V_{m+1} - x)^+}\right]$.

We have thus shown that for any $n \geq 1$

$$\begin{aligned}
-B\sum_{k=0}^{n-1} \widetilde{E}\left[h_m(t - V_k)\right] \leq \widetilde{E}\left[K\left(M_m^{(m)}, t\right)\right] - \sum_{k=0}^{n-1} \widetilde{E}\left[g\left(M_{m+k}^{(m)}, t - V_k\right)\right] \\
\leq B\widetilde{E}\left[e^{-\alpha(V_{m+n} - t)^+}\right].
\end{aligned}$$

Monotone convergence and the observation that $V_n \to \infty$ $\widetilde{P}$-a.s. immediately yields

$$\widetilde{E}\left[K\left(M_m^{(m)}, t\right)\right] \leq \sum_{k=0}^{\infty} \widetilde{E}\left[g\left(M_{m+k}^{(m)}, t - V_k\right)\right].$$

To obtain the lower bound note that by Lemma 2.5.3 we have that h_m is nonnegative and d.R.i. on $\mathbb{R}$, and therefore, $\sum_{k=0}^{\infty} \widetilde{E}\left[h_m(t - V_k)\right] < \infty$ for all $t \in \mathbb{R}$. It follows that

$$\widetilde{E}\left[K\left(M_m^{(m)}, t\right)\right] \geq \sum_{k=0}^{\infty} \widetilde{E}\left[g\left(M_{m+k}^{(m)}, t - V_k\right)\right] - B\sum_{k=0}^{\infty} \widetilde{E}\left[h_m(t - V_k)\right].$$

This completes the proof. □

In order to connect our framework with the notation in the Markov renewal theorem from [4], recall that the

$$\hat{X}_n = V_n - V_{n-1}, \qquad n \geq 1,$$

define the increments of the random walk along the spine, and note that $\{(M_{m+n}^{(m)}, \hat{X}_n) : n \geq 0\}$ is a time-homogeneous Markov process that only depends on the past through $\{M_{m+n}^{(m)} : n \geq 0\}$. Hence, we can define a transition kernel $\mathbf{P}$ according to

$$\mathbf{P}\left(M_{m+n}^{(m)},\, A \times B\right) \doteq \widetilde{P}\left(M_{m+n+1}^{(m)} \in A,\, \hat{X}_{n+1} \in B \middle| M_{m+n}^{(m)}, \hat{X}_n\right)$$

for any measurable sets $A \subseteq \mathcal{S}^{(m)}$ and $B \subseteq \mathbb{R}$. Thus, $\{(M_{m+n}^{(m)}, V_n) : n \geq 0\}$ is a Markov random walk in the sense of [4]. Furthermore, by the way the process $\{M_{m+n}^{(m)} : n \geq 0\}$ was constructed, it is mth-order stationary, in the sense that for each $n \geq 0$ the law under $\widetilde{P}$ of $\left(M_{m+n}^{(m)}, \ldots, M_{2m+n}^{(m)}\right)$ is the same as that of $\left(M_m^{(m)}, \ldots, M_{2m}^{(m)}\right)$, from which it follows that the unique stationary distribution for the chain $\{M_{m+n}^{(m)} : n \geq 0\}$ is given by

$$\eta_m(\cdot) \doteq \widetilde{E}\left[\frac{1}{m}\sum_{n=0}^{m-1} 1\left(M_{m+n}^{(m)} \in \cdot\right)\right] = \widetilde{P}\left(M_m^{(m)} \in \cdot\right). \tag{2.5.9}$$

The idea is now to use Theorem 2.1 in [4], which states that provided that

(i) $\{(M_{m+n}^{(m)}, V_n) : n \geq 0\}$ is a non-arithmetic and Harris recurrent Markov random walk, and

(ii) $g : \mathcal{S}^{(m)} \times \mathbb{R} \to \mathbb{R}$ is a measurable function such that $g(M, \cdot)$ is Lebesgue-a.e. continuous for η_m-a.e. M, and g is d.R.i. in the sense that

$$\int_{\mathcal{S}^{(m)}} \sum_{n=-\infty}^{\infty} \sup_{y \in (n, n+1]} |g(M, y)|\, \eta_m(dM) < \infty,$$

then, it will follow that

$$\lim_{t\to\infty} \widetilde{E}\left[\sum_{n=0}^{\infty} g\left(M_{n+m}^{(m)}, t - V_n\right)\right] = \frac{1}{\widetilde{E}\left[\hat{X}_1\right]} \int_{\mathcal{S}^{(m)}} \int_{\mathbb{R}} g(M,x)\, dx\, \eta_m(dM).$$

The non-arithmeticity of $\{(M_{m+n}^{(m)}, V_n) : n \geq 0\}$ follows from the non-arithmeticity of $\{V_n : n \geq 1\}$, which is ensured by Assumption 2.0.1(d) (see Lemma 2.1.2). To see that $\{M_{m+n}^{(m)} : n \geq 0\}$ is Harris recurrent note that by construction, $M_{m+n}^{(m)}$ is independent of $\{M_{n+2m}^{(m)}, M_{n+2m+1}^{(m)}, \dots\}$ for all $n \geq 0$, and therefore, by letting $\mathbf{Q}$ denote the transition kernel of $\{M_{m+n}^{(m)} : n \geq 0\}$ and $\mathbf{Q}^r$ its corresponding r-step transition kernel, we have

$$\mathbf{Q}^m \left(M_{m+n}^{(m)}, A\right) = \widetilde{P}\left(M_{n+2m}^{(m)} \in A \middle| M_{m+n}^{(m)}\right) = \widetilde{P}\left(M_m^{(m)} \in A\right) = \eta_m(A),$$

which satisfies the definition of a Harris chain (see Chapter VII Section 3 in [7]) with regeneration set $R = \mathcal{S}^{(m)}$, probability measure $\lambda = \eta_m$ and $\epsilon = 1$.

The last ingredient before applying the Markov renewal theorem of [4] to our situation is to show that the function g defined by (2.5.7) satisfies the necessary conditions. The corresponding result is given by the following lemma.

Lemma 2.5.5. *Let $g : \mathcal{S}^{(m)} \times \mathbb{R} \to \mathbb{R}$ be defined by* (2.5.7)*. Then, under Condition 2.0.1, $g(M,\cdot)$ is Lebesgue-a.e. continuous for η_m-a.e. M, and g is d.R.i. Moreover,*

$$\int_{\mathcal{S}^{(m)}} \int_{-\infty}^{\infty} g(M,x)\, dx\, \eta_m(dM) = \frac{1}{\alpha} \widetilde{E}\left[D_{\boldsymbol{J}_{m-1}}^{-1} \left(e^{\alpha\xi_{m-1}} - e^{\alpha\left(\max_{\boldsymbol{i} \prec \boldsymbol{J}_{m-1}} (S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}) - V_{m-1}\right)}\right)^+\right] < \infty.$$

Proof. We start by showing that $g(M,\cdot)$ is Lebesgue-a.e. continuous. To see this, let $M \in \mathcal{S}^{(m)}$, identify its m generations, all its weights, and its spine. Then note that

$$g\,(M,t) = 1(a(M) \leq t \leq b(M))c(M)e^{\alpha t}$$

for some fixed numbers $a(M), b(M)$ and $c(M)$. Therefore, it is Lebesgue-a.e. continuous.

It remains to show that g is d.R.i., for which we note that for any $M \in \mathcal{S}^{(m)}$ for which we have identified its generations, weights, and spine, we have

$$\begin{aligned}
&\sum_{n=-\infty}^{\infty} \sup_{y \in (n,n+1]} |g(M,y)| \\
&= \sum_{n=-\infty}^{\infty} \sup_{y \in (n,n+1]} 1\left(\max_{\boldsymbol{i} \prec \boldsymbol{J}_{m-1}} S_{\boldsymbol{i}} + Y_{\boldsymbol{i}} \leq y < V_{m-1} + \xi_{m-1}\right) e^{-\alpha(V_{m-1} - y)} D_{\boldsymbol{J}_{m-1}}^{-1}
\end{aligned}$$

$$
\begin{aligned}
&\leq D_{\boldsymbol{J}_{m-1}}^{-1} e^{-\alpha V_{m-1}} \sum_{n=-\infty}^{\infty} 1\left(V_{m-1}+\xi_{m-1}>n, \max_{\boldsymbol{i}\prec \boldsymbol{J}_{m-1}} S_{\boldsymbol{i}}+Y_{\boldsymbol{i}} \leq n+1\right) e^{\alpha(n+1)} \\
&\leq D_{\boldsymbol{J}_{m-1}}^{-1} e^{-\alpha V_{m-1}} \sum_{n=-\infty}^{\infty} \int_{n+1}^{n+2} 1\left(V_{m-1}+\xi_{m-1}+2>x, \max_{\boldsymbol{i}\prec \boldsymbol{J}_{m-1}} S_{\boldsymbol{i}}+Y_{\boldsymbol{i}} \leq x\right) e^{\alpha x}\, dx \\
&= D_{\boldsymbol{J}_{m-1}}^{-1} e^{-\alpha V_{m-1}} \int_{-\infty}^{\infty} 1\left(\max_{\boldsymbol{i}\prec \boldsymbol{J}_{m-1}} S_{\boldsymbol{i}}+Y_{\boldsymbol{i}} \leq x < V_{m-1}+\xi_{m-1}+2\right) e^{\alpha x}\, dx \\
&= D_{\boldsymbol{J}_{m-1}}^{-1} e^{-\alpha V_{m-1}} \frac{1}{\alpha}\left(e^{\alpha(V_{m-1}+\xi_{m-1}+2)} - e^{\alpha\left(\max_{\boldsymbol{i}\prec \boldsymbol{J}_{m-1}} S_{\boldsymbol{i}}+Y_{\boldsymbol{i}}\right)}\right)^{+} \\
&= \frac{1}{\alpha} D_{\boldsymbol{J}_{m-1}}^{-1}\left(e^{\alpha(\xi_{m-1}+2)} - e^{\alpha\left(\max_{\boldsymbol{i}\prec \boldsymbol{J}_{m-1}} S_{\boldsymbol{i}}+Y_{\boldsymbol{i}}-V_{m-1}\right)}\right)^{+}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\int_{\mathcal{S}^{(m)}} \sum_{n=-\infty}^{\infty} \sup_{y\in(n,n+1]} |g(M,y)|\, \eta_m(dM) \\
&\leq \frac{1}{\alpha}\widetilde{E}\left[D_{\boldsymbol{J}_{m-1}}^{-1}\left(e^{\alpha(\xi_{m-1}+2)} - e^{\alpha\left(\max_{\boldsymbol{i}\prec \boldsymbol{J}_{m-1}}(S_{\boldsymbol{i}}+Y_{\boldsymbol{i}})-V_{m-1}\right)}\right)^{+}\right] \\
&\leq \frac{1}{\alpha}\widetilde{E}\left[D_{\boldsymbol{J}_{m-1}}^{-1} e^{\alpha(\xi_{m-1}+2)}\right] = \frac{e^{2\alpha}}{\alpha} E[Q^\alpha] < \infty,
\end{aligned}
$$

which implies that g is d.R.i.

To complete the proof, note that essentially the same steps followed above give that

$$
\int_{\mathcal{S}^{(m)}} \int_{-\infty}^{\infty} g(M,x)\, dx\, \eta_m(dM) = \frac{1}{\alpha}\widetilde{E}\left[D_{\boldsymbol{J}_{m-1}}^{-1}\left(e^{\alpha\xi_{m-1}} - e^{\alpha\left(\max_{\boldsymbol{i}\prec \boldsymbol{J}_{m-1}}(S_{\boldsymbol{i}}+Y_{\boldsymbol{i}})-V_{m-1}\right)}\right)^{+}\right]
$$

and that the right hand side is finite. □

We are finally ready to prove Theorem 2.3.1.

Proof of Theorem 2.3.1. From the derivations at the beginning of the subsection and Lemma 2.5.4 we have that for any $m \geq 2$,

$$
\begin{aligned}
\widetilde{E}\left[1(\gamma(t)=\boldsymbol{J}_{\tau(t)}) e^{-\alpha(V_{\tau(t)}-t)} D_{\boldsymbol{J}_{\tau(t)}}^{-1}\right] &\leq \sum_{k=0}^{m-1} \widetilde{E}\left[1(|\gamma(t)|=k, \gamma(t)=\boldsymbol{J}_{\tau(t)}) e^{-\alpha(V_k-t)} D_{\boldsymbol{J}_k}^{-1}\right] \\
&\quad + \sum_{n=0}^{\infty} \widetilde{E}\left[g\left(M_{m+n}^{(m)}, t-V_n\right)\right].
\end{aligned}
$$

To see that each of the first m expectations converges to zero as $t\to\infty$, note that

$$
\begin{aligned}
\widetilde{E}\left[1(|\gamma(t)|=k, \gamma(t)=\boldsymbol{J}_{\tau(t)}) e^{-\alpha(V_k-t)} D_{\boldsymbol{J}_k}^{-1}\right] &\leq \widetilde{E}\left[1(V_k+\xi_k>t) e^{-\alpha(V_k-t)} D_{\boldsymbol{J}_k}^{-1}\right] \\
&= \widetilde{E}\left[u(t-V_k)\right],
\end{aligned}
$$

where $u(x) = e^{\alpha x}\widetilde{E}\left[1(\xi_0 > x)D_{\boldsymbol{J}_0}^{-1}\right] = e^{\alpha x}P(Y > x)$. Since u is bounded and integrable on $(-\infty, \infty)$, it follows from the bounded convergence theorem that

$$\limsup_{t\to\infty} \widetilde{E}\left[1(|\gamma(t)| = k, \gamma(t) = \boldsymbol{J}_{\tau(t)})e^{-\alpha(V_k - t)}D_{\boldsymbol{J}_k}^{-1}\right] \le \widetilde{E}\left[\limsup_{t\to\infty} u(t - V_k)\right] = 0.$$

Now use the Markov renewal theorem (Theorem 2.1 in [4]) and Lemma 2.5.5 to obtain that

$$\begin{aligned}
&\limsup_{t\to\infty} \widetilde{E}\left[1(\gamma(t) = \boldsymbol{J}_{\tau(t)})e^{-\alpha(V_{\tau(t)} - t)}D_{\boldsymbol{J}_{\tau(t)}}^{-1}\right] \\
&\le \sum_{k=0}^{m-1} \lim_{t\to\infty} \widetilde{E}\left[1(|\gamma(t)| = k, \gamma(t) = \boldsymbol{J}_{\tau(t)})e^{-\alpha(V_k - t)}D_{\boldsymbol{J}_k}^{-1}\right] \\
&\quad + \frac{1}{\widetilde{E}\left[\hat{X}_1\right]} \int_{\mathcal{S}^{(m)}} \int_{-\infty}^{\infty} g(M, x)\, dx\, \eta_m(dM) \\
&= \frac{1}{\alpha\mu}\widetilde{E}\left[D_{\boldsymbol{J}_{m-1}}^{-1}\left(e^{\alpha\xi_{m-1}} - e^{\alpha\left(\max_{\boldsymbol{i}\prec \boldsymbol{J}_{m-1}}(S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}) - V_{m-1}\right)}\right)^+\right] \doteq H_{m-1},
\end{aligned} \tag{2.5.10}$$

where $\mu = \widetilde{E}\left[\hat{X}_1\right] = E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right] > 0$.

To obtain a lower bound use Lemma 2.5.4 again to obtain that

$$\widetilde{E}\left[1(\gamma(t) = \boldsymbol{J}_{\tau(t)})e^{-\alpha(V_{\tau(t)} - t)}D_{\boldsymbol{J}_{\tau(t)}}^{-1}\right] \ge \sum_{n=0}^{\infty} \widetilde{E}\left[g\left(M_{m+n}^{(m)}, t - V_n\right)\right] - B\sum_{n=0}^{\infty} \widetilde{E}\left[h_m(t - V_n)\right].$$

Now use the two-sided renewal theorem (Lemma 1.1.3) and Lemma 2.5.3 to obtain that

$$\begin{aligned}
\lim_{t\to\infty} \sum_{n=0}^{\infty} \widetilde{E}\left[h_m(t - V_n)\right] &= \frac{1}{\mu}\int_{-\infty}^{\infty} h_m(x)\, dx \\
&= \frac{1}{\mu}\widetilde{E}\left[\frac{e^{-\alpha(V_{m+1} - Z_0)^+}}{\alpha} + (Z_0 - V_{m+1})^+\right] < \infty.
\end{aligned}$$

It follows that

$$\begin{aligned}
\liminf_{t\to\infty} \widetilde{E}\left[1(\gamma(t) = \boldsymbol{J}_{\tau(t)})e^{-\alpha(V_{\tau(t)} - t)}D_{\boldsymbol{J}_{\tau(t)}}^{-1}\right] &\ge H_{m-1} \\
&\quad - \frac{B}{\mu}\widetilde{E}\left[\frac{e^{-\alpha(V_{m+1} - Z_0)^+}}{\alpha} + (Z_0 - V_{m+1})^+\right].
\end{aligned} \tag{2.5.11}$$

Since $V_m \to \infty$ $\widetilde{P}$-a.s., we have that

$$\lim_{m\to\infty} \widetilde{E}\left[\frac{e^{-\alpha(V_{m+1} - Z_0)^+}}{\alpha} + (Z_0 - V_{m+1})^+\right] = 0,$$

and we conclude that

$$\lim_{t\to\infty} e^{\alpha t}P(W>t) = \lim_{t\to\infty} \widetilde{E}\left[1(\gamma(t)=\boldsymbol{J}_{\tau(t)})e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right] = \lim_{m\to\infty} H_m \doteq H.$$

The positivity of H under Condition 2.0.2 follows from representation (2.3.1) and Theorem 3.4 in [64]. ☐

2.5.3 Proof of Corollary 2.3.1

Here we provide a proof of Corollary 2.3.1, which relies on the following lemma.

Lemma 2.5.6. *For any* $0<\gamma<\alpha$,

$$\widetilde{E}\left[\bigvee_{i\neq \boldsymbol{J}_1} e^{\gamma(S_i - V_1)}\right] \leq \widetilde{E}[N].$$

Proof. Recall that $\widetilde{E}[N] = E\left[N\sum_{i=1}^N C_i^\alpha\right]$. If $\mathcal{F}_1 = \sigma(\boldsymbol{\psi}_\emptyset)$, then for $1\leq j\leq N$,

$$P\left(\mathbf{J}_1 = j \middle| \mathcal{F}_1\right) = \frac{C_j^\alpha}{\sum_{i=1}^N C_i^\alpha}.$$

For any $x\in\mathbb{R}$, we have that

$$\begin{aligned}
\widetilde{P}\left(\bigvee_{i\in B_1} S_i - V_1 > x\right) &= \sum_{j=1}^\infty \widetilde{P}\left(\bigvee_{i\neq j} S_i - S_j > x, j\leq N, \mathbf{J}_1 = j\right) \\
&= E\left[\sum_{j=1}^N 1\left(\frac{\bigvee_{i\neq j} C_i}{C_j} > e^x, \mathbf{J}_1 = j\right)\sum_{i=1}^N C_i^\alpha\right] \\
&= E\left[\sum_{j=1}^N 1\left(\frac{\bigvee_{i\neq j} C_i}{C_j} > e^x\right) P\left(\mathbf{J}_1 = j \middle| \mathcal{F}_1\right)\sum_{i=1}^N C_i^\alpha\right] \\
&= E\left[\sum_{j=1}^N 1\left(\frac{\bigvee_{i\neq j} C_i}{C_j} > e^x\right) C_j^\alpha\right].
\end{aligned}$$

Then,

$$\begin{aligned}
\widetilde{E}\left[e^{\gamma\left(\bigvee_{i\in B_1} S_i - V_1\right)}\right] &= \int_0^\infty \widetilde{P}\left(e^{\gamma\left(\bigvee_{i\in B_1} S_i - V_1\right)} > x\right) dx \\
&= \int_0^\infty \widetilde{P}\left(\bigvee_{i\in B_1} S_i - V_1 > \frac{\log x}{\gamma}\right) dx \\
&= \int_0^\infty E\left[\sum_{j=1}^N 1\left(\frac{\bigvee_{i\neq j} C_i}{C_j} > x^{1/\gamma}\right) C_j^\alpha\right] dx
\end{aligned}$$

$$
\begin{aligned}
&= E\left[\sum_{j=1}^{N} C_j^{\alpha} \int_0^{\bigvee_{i\neq j} C_i^{\gamma}/C_j^{\gamma}} dx\right] \\
&= E\left[\sum_{j=1}^{N} C_j^{\alpha-\gamma} \bigvee_{i\neq j} C_i^{\gamma}\right] \\
&\leq E\left[\sum_{j=1}^{N} \bigvee_{i=1}^{N} C_i^{\alpha}\right] \leq E\left[N\sum_{i=1}^{N} C_i^{\alpha}\right] = \widetilde{E}[N],
\end{aligned}
$$

where in the last line we used the fact that if $C_i \geq C_j$, then $C_j^{\alpha-\gamma}C_i^{\gamma} = \left(\frac{C_j}{C_i}\right)^{\alpha-\gamma} C_i^{\alpha} \leq C_i^{\alpha}$ and hence

$$
C_j^{\alpha-\gamma} \bigvee_{i\neq j} C_i^{\gamma} \leq C_j^{\alpha-\gamma} \bigvee_{i=1}^{N} C_i^{\gamma} = C_j^{\alpha-\gamma}\left(\bigvee_{i=1}^{N} C_i\right)^{\gamma} \leq \bigvee_{i=1}^{N} C_i^{\alpha}. \qquad \square
$$

We now complete the proof of Corollary 2.3.1.

Proof of Corollary 2.3.1. From (2.5.10) and (2.5.11) in the Proof of Theorem 2.3.1, it suffices to quantify the convergence

$$
\lim_{m\to\infty} \widetilde{E}\left[\frac{e^{-\alpha(V_{m+1}-Z_0)^+}}{\alpha} + (Z_0 - V_{m+1})^+\right] = 0.
$$

If $\alpha > 1$, let $(\alpha/2)\vee 1 < \beta < \alpha$ such that $\rho_\beta = E\left[\sum_{i=1}^{N} C_i^{\beta}\right] < 1$, and if $0 < \alpha \leq 1$, let $\alpha/(1+\epsilon) < \beta < \alpha$ such that $\rho_\beta < 1$, where ϵ is as in Condition 2.0.2. We have that

$$
\begin{aligned}
\widetilde{E}\left[e^{-\alpha(V_{m+1}-Z_0)^+}\right] &\leq \widetilde{E}\left[e^{-(\alpha-\beta)(V_{m+1}-Z_0)^+}\right] \leq \widetilde{E}\left[e^{-(\alpha-\beta)(V_{m+1}-Z_0)}\right] \\
&= \widetilde{E}\left[e^{-(\alpha-\beta)(V_{m+1}-V_1)}e^{(\alpha-\beta)(Z_0-V_1)}\right] \\
&= \rho_\beta^m \widetilde{E}\left[e^{(\alpha-\beta)(Z_0-V_1)}\right],
\end{aligned}
$$

since $V_{m+1} - V_1$ is independent of $Z_0 - V_1$ and $\widetilde{E}\left[e^{-(\alpha-\beta)(V_{m+1}-V_1)}\right] = \widetilde{E}\left[e^{-(\alpha-\beta)V_m}\right] = \rho_\beta^m$. Also,

$$
\widetilde{E}\left[(Z_0 - V_{m+1})^+\right] \leq \frac{\widetilde{E}\left[e^{(\alpha-\beta)(Z_0-V_{m+1})}\right]}{\alpha-\beta} = \frac{\rho_\beta^m}{\alpha-\beta}\widetilde{E}\left[e^{(\alpha-\beta)(Z_0-V_1)}\right].
$$

Showing that $\widetilde{E}\left[e^{(\alpha-\beta)(Z_0-V_1)}\right] < \infty$ will imply the result with $c = \rho_\beta$.

To that end, first note that

$$
\widetilde{E}\left[e^{(\alpha-\beta)(Z_0-V_1)}\right] \leq \widetilde{E}\left[e^{(\alpha-\beta)(\xi_0-V_1)}\right] + \widetilde{E}\left[\bigvee_{i\in B_1} e^{(\alpha-\beta)(S_i-V_1+W_i)}\right]. \tag{2.5.12}
$$

For the first term on the right hand side, note that since $P(\boldsymbol{J}_1 = i | \boldsymbol{\psi}_\emptyset) = C_i^\alpha / D_\emptyset$ for any $i \leq N_\emptyset$ and $D_\emptyset = L_1$,

$$
\begin{aligned}
\widetilde{E}\left[e^{(\alpha-\beta)(\xi_0 - V_1)} \right] &= E\left[\sum_{i=1}^{N} e^{(\alpha-\beta)(Y - X_i)} 1(\boldsymbol{J}_1 = i) L_1 \right] \\
&= E\left[Q^{\alpha-\beta} \sum_{i=1}^{N} C_i^{\beta-\alpha} P(\boldsymbol{J}_1 = i | \boldsymbol{\psi}_\emptyset) L_1 \right] \\
&= E\left[Q^{\alpha-\beta} \sum_{i=1}^{N} C_i^{\beta} \right] \\
&\leq \left(E\left[Q^\alpha \right] \right)^{(\alpha-\beta)/\alpha} \left(E\left[\left(\sum_{i=1}^{N} C_i^{\beta} \right)^{\alpha/\beta} \right] \right)^{\beta/\alpha},
\end{aligned}
$$

where the last line follows from Hölder's inequality. In the case that $\alpha > 1$ and $(\alpha/2) \vee 1 < \beta < \alpha$, in particular $\beta > 1$ implies that

$$
E\left[\left(\sum_{i=1}^{N} C_i^{\beta} \right)^{\alpha/\beta} \right] \leq E\left[\left(\sum_{i=1}^{N} C_i \right)^{\alpha} \right] < \infty.
$$

When $\alpha \leq 1$ and $\beta > \alpha/(1+\epsilon)$, we have that $(1+\epsilon)\beta/\alpha > 1$ and hence

$$
E\left[\left(\sum_{i=1}^{N} C_i^{\beta} \right)^{\alpha/\beta} \right] = E\left[\left(\sum_{i=1}^{N} \left(C_i^{\alpha/(1+\epsilon)} \right)^{(1+\epsilon)\beta/\alpha} \right)^{\alpha/\beta} \right] \leq E\left[\left(\sum_{i=1}^{N} C_i^{\alpha/(1+\epsilon)} \right)^{1+\epsilon} \right] < \infty.
$$

In either case, $\widetilde{E}\left[e^{(\alpha-\beta)(\xi_0 - V_1)} \right] < \infty$.

For the second term in (2.5.12), let $\gamma = 2(\alpha - \beta)$, and note that by assumption $\beta > \alpha/2$ and so $\gamma < \alpha$. We then have that

$$
\begin{aligned}
\widetilde{E}\left[\bigvee_{i \in B_1} e^{(\alpha-\beta)(S_i - V_1 + W_i)} \right] &\leq \widetilde{E}\left[\bigvee_{i \in B_1} e^{(\alpha-\beta)(S_i - V_1)} \bigvee_{i \in B_1} e^{(\alpha-\beta) W_i} \right] \\
&\leq \left(\widetilde{E}\left[\bigvee_{i \in B_1} e^{\gamma(S_i - V_1)} \right] \right)^{1/2} \left(\widetilde{E}\left[\bigvee_{i \in B_1} e^{\gamma W_i} \right] \right)^{1/2} \\
&\leq \left(\widetilde{E}[N] \right)^{1/2} \left(\widetilde{E}[N] E\left[e^{\gamma W} \right] \right)^{1/2} \\
&= \widetilde{E}[N] \left(E\left[e^{\gamma W} \right] \right)^{1/2} < \infty,
\end{aligned}
$$

where $E\left[e^{\gamma W} \right] < \infty$ since $\gamma < \alpha$, and we used Lemma 2.5.6. □

2.5.4 The Importance Sampling Estimator $Z(t)$

The last part of the paper contains the proofs of Lemmas 2.4.1 and 2.4.2 in Section 2.4. The first of these establishes the asymptotic behavior of $\tau(t)$ as $t \to \infty$, and the second one proves the strong efficiency of our proposed estimator.

Proof of Lemma 2.4.1. Since $\tau(y)$ is monotone nondecreasing in y, $\lim_{y\to\infty}\tau(t) = \sup_y \tau(y)$ exists, and for any $k > 0$ and $x \in \mathbb{R}$,

$$\widetilde{P}\left(\lim_{y\to\infty}\tau(y) > k\right) \geq \widetilde{P}(\tau(x) > k) = \widetilde{P}\left(\max_{j\leq k} V_j + \xi_j \leq x\right).$$

Letting $x \to \infty$, we see that $\widetilde{P}(\lim_{y\to\infty}\tau(y) > k) = 1$, and since this is true for all k, $\lim_{y\to\infty}\tau(y) = \infty$ $\widetilde{P}$-a.s. It remains to show that $\tau(t)/t \to 1/\mu$ $\widetilde{P}$-a.s. provided $\widetilde{E}\left[\xi_0^+\right] < \infty$ and $\widetilde{P}(\xi_0 > -\infty) > 0$, which are implied by Condition 2.0.1.

Start by noting that for any $t > 0$,

$$\frac{1}{\tau(t)+1}\max_{0\leq k<\tau(t)}(V_k + \xi_k) \leq \frac{t}{\tau(t)+1} \leq \frac{V_{\tau(t)} + \xi_{\tau(t)}}{\tau(t)+1}.$$

To obtain an upper bound for $t/(\tau(t)+1)$ note that

$$\limsup_{t\to\infty}\frac{t}{\tau(t)+1} \leq \limsup_{n\to\infty}\frac{V_n + \xi_n^+}{n} \leq \mu + \limsup_{n\to\infty}\frac{\xi_n^+}{n} = \mu,$$

where $\limsup_{n\to\infty}\xi_n^+/n = 0$ $\widetilde{P}$-a.s. since $\widetilde{E}[\xi_0^+] < \infty$ and $\limsup_{n\to\infty} V_n/n = \mu$ $\widetilde{P}$-a.s. by the strong law of large numbers since $\widetilde{E}[|V_1|] < \infty$. To obtain a lower bound let $m_n = n/\log n$ and note that since $t/(\tau(t)+1) > 0$ we have

$$\begin{aligned}
\liminf_{t\to\infty}\frac{t}{\tau(t)+1} &\geq \liminf_{n\to\infty}\left(\frac{1}{n}\max_{0\leq k<n}(V_k+\xi_k)\right)^+ \geq \liminf_{n\to\infty}\frac{1}{n}\max_{m_n\leq k<n}(V_k+\xi_k)^+ \\
&\geq \liminf_{n\to\infty}\frac{1}{n}\max_{m_n\leq k<n}\left((\mu k+\xi_k)^+ - (\mu k - V_k)^+\right) \\
&\geq \liminf_{n\to\infty}\frac{1}{n}\max_{m_n\leq k<n}(\mu k+\xi_k)^+ - \limsup_{n\to\infty}\frac{1}{n}\max_{m_n\leq k<n}(\mu k - V_k)^+ \\
&\geq \liminf_{n\to\infty}\frac{1}{n}\max_{m_n\leq k<n}(\mu k+\xi_k)^+ - \limsup_{n\to\infty}\left(\mu - \frac{V_n}{n}\right)^+.
\end{aligned}$$

Since the strong law of large numbers gives that $\limsup_{n\to\infty}(\mu - V_n/n)^+ = 0$ $\widetilde{P}$-a.s., it only remains to show that $\liminf_{n\to\infty} n^{-1}\max_{m_n\leq k<n}(\mu k+\xi_k)^+ \geq \mu$ $\widetilde{P}$-a.s. To show that this is indeed the case, fix $0 < \epsilon < \mu$, define $M_\epsilon = \lceil e^{(1-\epsilon/2)^{-1}}\rceil$, and note that

$$\begin{aligned}
&\sum_{n=3}^{\infty}\widetilde{P}\left(\frac{1}{n}\max_{m_n\leq k<n}(\mu k+\xi_k)^+ - \mu < -\epsilon\right) \\
&= \sum_{n=3}^{\infty}\widetilde{P}\left(\max_{m_n\leq k<n}(\mu k+\xi_k)^+ < (\mu-\epsilon)n\right) \\
&\leq M_\epsilon + \sum_{n=M_\epsilon+1}^{\infty}\widetilde{P}\left(\max_{\lceil(1-\epsilon/2)n\rceil\leq k<n}(\mu k+\xi_k)^+ < (\mu-\epsilon)n\right)
\end{aligned}$$

$$\leq M_\epsilon + \sum_{n=M_\epsilon+1}^{\infty} \prod_{k=\lceil(1-\epsilon/2)n\rceil}^{n} \widetilde{P}\left(\mu k + \xi_k < (\mu-\epsilon)n\right)$$
$$\leq M_\epsilon + \sum_{n=M_\epsilon+1}^{\infty} \prod_{k=\lceil(1-\epsilon/2)n\rceil}^{n} \widetilde{P}\left(\xi_0 < -(\epsilon/2)n\right)$$
$$\leq M_\epsilon + \sum_{n=M_\epsilon+1}^{\infty} \widetilde{P}(2\xi_0 < -\epsilon n)^{(\epsilon/2)n-1}.$$

Since by assumption we have that $\widetilde{P}(\xi_0 > -\infty) > 0$, then there exists $n_0 > M_\epsilon$ such that $\widetilde{P}(2\xi_0 < -\epsilon n) < 1$ for all $n \geq n_0$, which shows that the series above converges. Finally, use the Borel-Cantelli lemma to conclude that

$$\liminf_{n\to\infty} \frac{1}{n} \max_{m_n \leq k < n} (\mu k + \xi_k)^+ - \mu = 0 \qquad \widetilde{P}\text{-a.s.},$$

which in turn implies that

$$\lim_{t\to\infty} \frac{t}{\tau(t)} = \mu \qquad \widetilde{P}\text{-a.s.} \quad \square$$

Proof of Lemma 2.4.2. From Theorem 2.3.1,

$$P(W > t)^2 \sim H^2 e^{-2\alpha t} \quad \text{as} \quad t \to \infty$$

for $H^2 > 0$, and so

$$\limsup_{t\to\infty} \frac{\widetilde{\mathrm{Var}}(Z(t))}{P(W>t)^2} \leq \limsup_{t\to\infty} \frac{\widetilde{E}\left[1(\boldsymbol{J}_{\tau(t)} = \gamma(t)) e^{-2\alpha V_{\tau(t)}} D^{-2}_{\boldsymbol{J}_{\tau(t)}}\right]}{H^2 e^{-2\alpha t}}$$
$$\leq H^{-2} \limsup_{t\to\infty} \widetilde{E}\left[e^{-2\alpha(V_{\tau(t)}-t)} D^{-2}_{\boldsymbol{J}_{\tau(t)}}\right].$$

Now, by an argument analogous to that in the proof of Lemma 2.5.1, we have that

$$\widetilde{E}\left[e^{-2\alpha(V_{\tau(t)}-t)} D^{-2}_{\boldsymbol{J}_{\tau(t)}}\right] \leq \sum_{n=0}^{\infty} \widetilde{E}\left[v(t - V_n)\right],$$

where $v(x) = e^{2\alpha x} E[1(Y > x) D^{-1}]$, which is integrable since

$$\int_{-\infty}^{\infty} v(x)\, dx = \int_{-\infty}^{\infty} e^{2\alpha x} E\left[1(Y > x) D^{-1}\right] dx$$
$$= E\left[\int_{-\infty}^{Y} e^{2\alpha x} D^{-1}\, dx\right] = \frac{1}{2\alpha} E\left[Q^{2\alpha} D^{-1}\right] < \infty.$$

Then, v is d.R.i. by the same argument as in the proof of Lemma 2.5.1 for the function u, and hence

$$\limsup_{t\to\infty} \widetilde{E}\left[e^{-2\alpha(V_{\tau(t)}-t)} D_{\boldsymbol{J}_{\tau(t)}}^{-2}\right] \leq \limsup_{t\to\infty} \sum_{n=0}^{\infty} \widetilde{E}\left[v(t-V_n)\right] = \int_{-\infty}^{\infty} v(x)\, dx < \infty.$$

The proof is the same for the estimator $1(\boldsymbol{J}_{\tau(t)} = \gamma(t))e^{-\alpha V_{\tau(t)}}$ in the case of independent Q. $\square$

2.5.5 The Bounded Perturbations in the Non-Branching Case

We end the paper with a short proof of Theorem 1 in [6] for the non-branching case $N \equiv 1$ and bounded Q, which establishes the exponential asymptotic behavior of $P(W > x)$. As mentioned earlier, a similar approach could be used for the branching case with bounded Q, however, since our goal was not to establish the exponential asymptotic itself (for which the implicit renewal theorem on trees in [64] can be used), but rather shed some light into the event leading to the constant, we do not pursue this idea any further. In the $N \equiv 1$ case, W is the maximum of a negative drift perturbed random walk, that is

$$W \stackrel{\mathcal{D}}{=} \sup_{n\geq 0} \left(V_n + \xi_n\right).$$

Theorem 2.5.1 (Theorem 1 in [6])**.** *Suppose that $N \equiv 1$ and one of the following holds:*

(a) $\{(\xi_{i-1}, \hat{X}_i) : i \geq 1\}$ are i.i.d., or,

(b) $\{\xi_i : i \geq 0\}$ is a stationary sequence, independent of the i.i.d. sequence $\{\hat{X}_i : i \geq 1\}$.

In either case, assume that $P(\xi_0 \leq c) = 1$ for some constant c and $P(\xi_0 > 0) > 0$; we allow the possibility that $P(\xi_0 = -\infty) > 0$ but assume that $P(\hat{X}_1 = -\infty) = 0$. Assume further that $E\left[e^{\alpha \hat{X}_1}\right] = 1$ and $E\left[\hat{X}_1 e^{\alpha \hat{X}_1}\right] \in (0,\infty)$ for some $\alpha > 0$, and that the measure $P(\hat{X}_1 \in dx)$ is non-arithmetic. Then,

$$P(W > x) \sim He^{-\alpha x}, \qquad x \to \infty,$$

for some constant $0 < H < \infty$.

Proof. Define the filtration $\mathcal{H}_n = \sigma(\hat{X}_i : 1 \leq i \leq n)$ for $n \geq 0$ and $\mathcal{H}_0 = \sigma(\varnothing)$. Let $T(x) = \inf\{n \geq 1 : V_n > x\}$ and note that it is a stopping time with respect to $\{\mathcal{H}_n : n \geq 0\}$. Also let $\tau(x) = \inf\{n \geq 1 : V_n + \xi_n > x\}$. Since the perturbations are bounded, we have $\tau(x) \geq T(x - c)$, and since the drift of $V_n = \hat{X}_1 + \cdots + \hat{X}_n$ is positive under $\widetilde{P}$, then $\widetilde{P}(T(t) < \infty) = 1$ for all $t \geq 0$. Now let $y = x - c$ and write,

$$\begin{aligned} &P(W > x) \\ &= P(T(y) \leq \tau(x) < \infty) \end{aligned}$$

$$= P\left(T(y) < \infty, \sup_{k\ge 0}\left(V_{T(y)-1} + V_{T(y)+k} + \xi_{T(y)+k}\right) > x\right)$$
$$= E\left[1(T(y) < \infty) E\left[1\left(V_{T(y)} + \max\left\{\xi_{T(y)}, \sup_{k\ge 1}\left(V_{T(y)+k} + \xi_{T(y)+k}\right)\right\} > x\right)\middle| \mathcal{H}_{T(y)}\right]\right].$$

Since $\max\left\{\xi_{T(y)}, \sup_{k\ge 1}(V_{T(y)+k} + \xi_{T(y)+k})\right\}$ is independent of $\mathcal{H}_{T(y)}$ and has the same distribution as W, we have that

$$E\left[1\left(V_{T(y)} + \max\left\{\xi_{T(y)}, \sup_{k\ge 1}\left(V_{T(y)+k} + \xi_{T(y)+k}\right)\right\} > x\right)\middle| \mathcal{H}_{T(y)}\right] = \overline{F}(x - V_{T(y)}),$$

where $\overline{F}(t) = P(W > t)$. Hence,

$$\begin{aligned} e^{\alpha x} P(W > x) &= e^{\alpha x} E\left[1(T(y) < \infty)\overline{F}(x - V_{T(y)})\right] \\ &= e^{\alpha x} E\left[1(T(y) < \infty)\overline{F}(x - V_{T(y)}) e^{-\alpha V_{T(y)}} L_{T(y)}\right] \\ &= e^{\alpha x} \widetilde{E}\left[1(T(y) < \infty)\overline{F}(x - V_{T(y)}) e^{-\alpha V_{T(y)}}\right] \\ &= \widetilde{E}\left[\overline{F}(x - V_{T(y)}) e^{-\alpha(V_{T(y)} - x)}\right] \\ &= \widetilde{E}\left[\overline{F}(c - B(x-c)) e^{-\alpha B(x-c)}\right] e^{\alpha c}, \end{aligned}$$

where $B(t) = V_{T(t)} - t \ge 0$ is the overshoot process of the random walk $\{V_n : n \ge 1\}$. Since $P(\hat{X}_1 \in dx)$ is non-arithmetic, so is $\widetilde{P}(\hat{X}_1 \in dx) = E[1(\hat{X}_1 \in dx)e^{\alpha \hat{X}_1}]$, and hence by Theorem 2.1 in Chapter VIII of [7], $B(t)$ converges in $\widetilde{P}$-distribution as $t \to \infty$ to an a.s. finite limit $B(\infty)$, and therefore,

$$\lim_{x\to\infty} \widetilde{E}\left[\overline{F}(c - B(x-c)) e^{-\alpha B(x-c)}\right] e^{\alpha c} = \widetilde{E}\left[\overline{F}(c - B(\infty)) e^{-\alpha B(\infty)}\right] \doteq H.$$

To see that $H > 0$, note that $\overline{F}(t) \ge \sup_{n\ge 0} P(V_n + \xi_n > t, \xi_n > 0) \ge \sup_{n\ge 0} P(V_n > t)P(\xi_0 > 0)$. The condition $E[\hat{X}_1 e^{\alpha \hat{X}_1}] > 0$ implies that $P(\hat{X}_1 > 0) > 0$, and in particular there must be some $\epsilon > 0$ such that $P(\hat{X}_1 > \epsilon) > 0$. For any $t > 0$, if $k > t/\epsilon$, then

$$P(V_k > t) \ge P\left(\hat{X}_1 > \frac{t}{k}, \ldots, \hat{X}_k > \frac{t}{k}\right) \ge P(\hat{X}_1 > \epsilon)^k > 0.$$

Since $P(\xi_0 > 0) > 0$ as well, we the have $\overline{F}(t) > 0$ for each $t > 0$. In particular, since $B(\infty) \ge 0$ a.s., $\overline{F}(c - B(\infty)) \ge \overline{F}(c) > 0$. Therefore, $B(\infty) < \infty$ a.s. and $\overline{F}(c - B(\infty)) > 0$ a.s. imply together that $H > 0$. $\square$

CHAPTER 3

Efficient Hybrid Estimation for Tail Events on Trees

In this chapter, we again consider the max-type distributional recursion

$$R \stackrel{\mathcal{D}}{=} Q \vee \bigvee_{i=1}^{N} C_i R_i, \tag{3.0.1}$$

where the random vector $\boldsymbol{\psi} \doteq (Q, N, \{C_i\})$ has arbitrary dependence with $Q \geq 0$, $C_i \geq 0$ for all i, and $N \in \mathbb{N}_+ = \{1, 2, \ldots\}$, and $\{R_i\}$ are i.i.d. copies of R. Letting $Y = \log Q$, $X_i = \log C_i$, and $W = \log R_i$, this is equivalent to what is known as the high-order Lindley equation

$$W \stackrel{\mathcal{D}}{=} \max\left\{Y, \max_{1\leq i\leq N}(X_i + W_i)\right\}, \tag{3.0.2}$$

where $\{W_i\}$ are i.i.d. copies of W. Its *special endogenous solution* can be constructed through a marked Galton-Watson process that lives on a tree $\mathcal{T}$ constructed as in Section 1.1.4. We recall that $\mathcal{T}$ is constructed from the collection $\{\boldsymbol{\psi}_{\boldsymbol{i}} = (Q_{\boldsymbol{i}}, N_{\boldsymbol{i}}, C_{(\boldsymbol{i},1)}, C_{(\boldsymbol{i},2)}, \ldots) : \boldsymbol{i} \in U\}$ of i.i.d. copies of $\boldsymbol{\psi}$, where $U = \bigcup_{k=0}^{\infty} \mathbb{N}_+^k$ is the collection of finite strings of positive integers endowed with length-lexicographic ordering. A_n denotes the set of nodes in the nth generation, as determined by $\{N_{\boldsymbol{i}} : \boldsymbol{i} \in U\}$. Each note $\boldsymbol{i} \neq \emptyset$ in the tree $\mathcal{T}$ is assigned the weight $C_{\boldsymbol{i}}$ and the cumulative weight $\Pi_{\boldsymbol{i}}$, where

$$\Pi_\emptyset = 1, \qquad \Pi_i = C_i, \quad i \in A_1, \qquad \text{and} \qquad \Pi_{\boldsymbol{i}} = \Pi_{\boldsymbol{i}|n-1} C_{\boldsymbol{i}}, \quad \boldsymbol{i} \in A_n, \quad n \geq 2.$$

Recall that the special endogenous solution to (3.0.1) is then given by

$$R = \bigvee_{\boldsymbol{i}\in\mathcal{T}} \Pi_{\boldsymbol{i}} Q_{\boldsymbol{i}}. \tag{3.0.3}$$

Letting $S_{\boldsymbol{i}} = \log \Pi_{\boldsymbol{i}}$ and $Y_{\boldsymbol{i}} = \log Q_{\boldsymbol{i}}$ for each $\boldsymbol{i} \in \mathcal{T}$, the endogenous solution to (3.0.2) satisfies

$$W = \bigvee_{\boldsymbol{i}\in\mathcal{T}} (S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}). \tag{3.0.4}$$

Setting $X_{\boldsymbol{i}} = \log C_{\boldsymbol{i}}$, the increments $X_{(\boldsymbol{i}|1)}, X_{(\boldsymbol{i}|2)}, \ldots, X_{\boldsymbol{i}}$ of $S_{\boldsymbol{i}}$ are independent (but not necessarily identically distributed), and hence W is the maximum of a branching random walk with perturbation. The previous chapter was partially devoted to showing that

$$P(W > t) \sim He^{-\alpha t}, \qquad t \to \infty, \tag{3.0.5}$$

for some $H \geq 0$, whenever $\alpha > 0$ can be found such that

$$E\left[\sum_{i=1}^{N} C_i^{\alpha}\right] = E\left[\sum_{i=1}^{N} e^{\alpha X_i}\right] = 1 \quad \text{and} \quad E\left[\sum_{i=1}^{N} C_i^{\alpha} \log C_i\right] = E\left[\sum_{i=1}^{N} e^{\alpha X_i} X_i\right] \in (0, \infty), \tag{3.0.6}$$

among other assumptions. This result was also established using implicit renewal theory in [62, 64]. Also key to establishing (3.0.5) is the contraction condition

$$\rho_{\beta} = E\left[\sum_{i=1}^{N} C_i^{\beta}\right] = E\left[\sum_{i=1}^{N} e^{\beta X_i}\right] < 1 \qquad \text{for some} \qquad 0 < \beta < \alpha. \tag{3.0.7}$$

When this holds, the random walk $S_{\boldsymbol{i}}$ has negative drift along its branches (see (1.1.18)). This in particular ensures that $R < \infty$ a.s. and hence $W < \infty$ a.s. (see Lemma 3.1 in [64]).

It is of practical interest to estimate the tail probabilities $P(W > t)$, and in Chapter 2, a strongly efficient importance sampling algorithm for simulating the tail probabilities $P(W > t)$ is proposed, which significantly generalizes Siegmund's algorithm [92] for rare event simulation for the maximum of a negative-drift random walk on $\mathbb{R}$. The algorithm is based on defining a new probability measure $\widetilde{P}$ that tilts the weights and perturbations on one randomly chosen path down the tree and its nodes' immediate offspring while leaving the distribution of the weights and perturbations on all other branches unchanged. We will refer to this special set of nodes as the *spine* of $\mathcal{T}$.

The algorithm in Chapter 2 proceeds by constructing an unbiased estimator for $P(W > t)$ by generating the tree $\mathcal{T}$ under the change of measure until the random time

$$\gamma(t) = \inf\left\{\boldsymbol{i} \in \mathcal{T} : S_{\boldsymbol{i}} + Y_{\boldsymbol{i}} > t\right\},$$

where the infimum is taken according to $\prec$ and where $|\gamma(t)|$ is guaranteed to be finite because $S_{\boldsymbol{i}}$ restricted to the spine is a positive-drift random walk. In fact, if $\tau(t)$ denotes the generation at which $S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}$ restricted to the spine first reaches the level t, then $\tau(t) \sim t/\mu$ $\widetilde{P}$-a.s. as $t \to \infty$, where $\mu = E\left[\sum_{i=1}^{N} e^{\alpha X_i} X_i\right]$, a result analogous to the case of a random walk on $\mathbb{R}$ (See Lemma 2.4.1 in Chapter 2 and Lemma 5.7 in Chapter XIII of [7]).

While this algorithm has many nice properties, obtaining a sample of size n of estimates of $P(W > t)$ requires the construction of n trees up to generation approximately t/μ, a task that necessitates generating on the order of $n(E[N])^{t/\mu}$ copies of the random vector $\boldsymbol{\psi}$. When $E[N]$ and t are large, this may be prohibitively slow. However, the

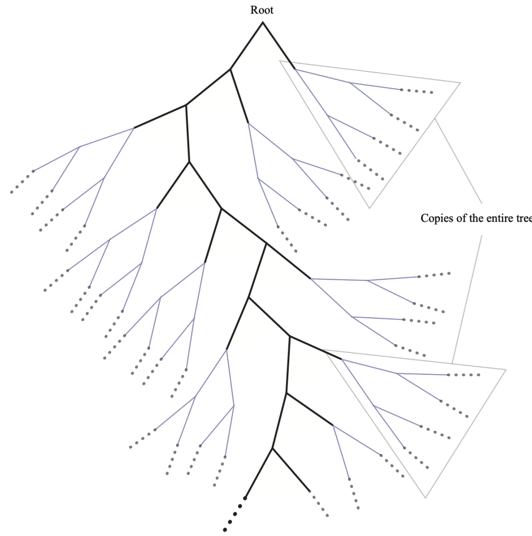


Figure 3.1: The tree $\mathcal{T}$ under $\widetilde{P}$ decomposed as the spine (black) and independent copies of $\mathcal{T}$ under P.

fact that the change of measure $\widetilde{P}$ applies a tilt only to the spine of $\mathcal{T}$ suggests that this is the path on which $S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}$ is most likely to hit the level t first. The idea of the algorithm presented herein is that the rare event probability $P(W > t)$ can be approximated by only generating the spine and then accounting for the other (untilted) branches using some approximation method. Figure 3.1 shows the spine decomposition of $\mathcal{T}$ and gives visual intuition for this idea. We present the resulting "hybrid" estimator for $P(W > t)$ in terms of general (simulated, random) approximations $\hat{F}_k$ for the CDFs $P(W^{(k)} \leq \cdot)$, where $W^{(k)} = \max_{|\boldsymbol{i}| \leq k}(S_{\boldsymbol{i}} + Y_{\boldsymbol{i}})$ is endogenous solution W truncated at some finite level of $\mathcal{T}$.

The main theoretical result of this chapter (Theorem 3.1.5) concerns the rate of convergence to zero of the relative bias of this estimator with respect to $P(W > t)$. In Section 3.1.2, we give an example of efficient approximations $\{\hat{F}_k : k \geq 0\}$ generated with the *population dynamics algortihm* analyzed in [3, 24, 84, 86]. We remark that with this particular hybrid estimator, the number of copies of $\boldsymbol{\psi}$ that need to be generated for an estimate of $P(W > t)$ is *linear* in t, importantly with no dependence on the distribution of N either under P or $\widetilde{P}$ (see Remark 3.1.2).

Throughout, we assume that $\boldsymbol{\psi} = (Q, N, \{C_i\})$ satisfies the following conditions.

Condition 3.0.1. *For some $\alpha > 0$,*

$$E\left[\sum_{i=1}^{N} C_i^\alpha\right] = 1 \quad \textit{and} \quad 0 < E\left[\sum_{i=1}^{N} C_i^\alpha \log C_i\right] < \infty.$$

Furthermore, for this α,

(a) The probability measure $\eta(dx) = E\left[\sum_{i=1}^{N} C_i^\alpha 1(\log C_i \in dx)\right]$ is nonlattice,

(b) $P\left(\sum_{i=1}^{N} C_i^{\alpha} = 0\right) = 0$,

(c) For some $\alpha/2 < \beta < \alpha$, $\rho_\beta = E\left[\sum_{i=1}^{N} C_i^{\beta}\right] < 1$,

(d) $N \geq 1$ *a.s.,* $E[N] < \infty$, *and* $E\left[N \sum_{i=1}^{N} C_i^{\alpha}\right] < \infty$,

(e) $P(Q > 0) > 0$ *and* $E[Q^{\alpha}] < \infty$.

For the inclusion of conditions (a) and $N \geq 1$ a.s., see the discussion in Remark 2.1.1(a). Note that the other conditions in part (d) included here are absent from the previous chapter. Having $E\left[N \sum_{i=1}^{N} C_i^{\alpha}\right] < \infty$ in particular ensures that $\widetilde{E}[N] < \infty$ under the change of measure. Generally, having finite mean of the offspring distribution is desirable when simulating Galton-Watson processes. However, in the analysis of this chapter, these finite moments are used explicitly.

We recall the change of measure $\widetilde{P}$ defined in Chapter 2. Namely, we start from the root node of $\mathcal{T}$ and construct a *chosen path* by selecting one offspring at random with probability proportional to their weights to the power of α, and then proceed in this manner down the tree. We begin to choose the path by $\boldsymbol{J}_0 = \emptyset$ and then recursively define the indices by

$$\boldsymbol{J}_k = (\boldsymbol{J}_{k-1}, i) \quad \text{w.p.} \quad \frac{C^{\alpha}_{(\boldsymbol{J}_{k-1}, i)}}{\sum_{j=1}^{N_{\boldsymbol{J}_{k-1}}} C^{\alpha}_{(\boldsymbol{J}_{k-1}, j)}}, \qquad 1 \leq i \leq N_{\boldsymbol{J}_{k-1}}, \qquad k \geq 1.$$

The sequence $\{\boldsymbol{J}_k : k \geq 0\}$ gives the nodes in the spine of the tree $\mathcal{T}$, with $\boldsymbol{J}_k$ denoting the chosen node in the kth generation.

Now define the process $\{L_k : k \geq 0\}$ by

$$L_0 = 1, \qquad L_k = \prod_{r=0}^{k-1} D_{\boldsymbol{J}_r}, \quad k \geq 1,$$

where $D_{\boldsymbol{i}} = \sum_{j=1}^{N_{\boldsymbol{i}}} C^{\alpha}_{(\boldsymbol{i}, j)}$ for $\boldsymbol{i} \in \mathcal{T}$ with generic copy D, and define the filtration

$$\mathcal{G}_0 = \sigma(\varnothing), \qquad \mathcal{G}_k = \sigma\left(\{\boldsymbol{\psi}_{\boldsymbol{i}} : \boldsymbol{i} \in A_j, j < k\} \cup \{\boldsymbol{J}_j : j \leq k\}\right), \quad k \geq 1.$$

Recall that L_k is a mean-one nonnegative martingale with respect to $\mathcal{G}_k$, and $\widetilde{P}$ is defined by

$$\widetilde{P}(A) = E\left[1(A) L_k\right], \qquad A \in \mathcal{G}_k, \quad k \geq 0,$$

on the σ-field $\mathcal{G}_\infty$, the smallest σ-field containing all of $\{\mathcal{G}_k : k \geq 0\}$. Here we recall the properties of the measure $\widetilde{P}$ given in Lemma 2.1.1 of Chapter 2

To identify the random walk restricted to the chosen path, define

$$\hat{X}_k = X_{\boldsymbol{J}_k} = \log C_{\boldsymbol{J}_k}, \qquad \xi_k = Y_{\boldsymbol{J}_k} = \log Q_{\boldsymbol{J}_k}, \qquad k \geq 0$$

and let $V_0 = 0$ and $V_k = \hat{X}_1 + \cdots + \hat{X}_k$ for $k \geq 1$. As we recall from Lemma 2.1.2 of Chapter 2, $\{V_k : k \geq 0\}$ is a positive-drift random walk with i.i.d. increments under $\widetilde{P}$, and for each k, V_k is independent of ξ_k. Furthermore,

$$\mu \doteq \widetilde{E}\left[\hat{X}_1\right] = E\left[\sum_{i=1}^{N} C_i^\alpha \log C_i\right] \in (0, \infty).$$

3.1 A Hybrid Importance Sampler

In Chapter 2, we derived the representation

$$P(W > t) = \widetilde{E}\left[1(\gamma(t) = \boldsymbol{J}_{\tau(t)}) e^{-\alpha V_{\tau(t)}} D^{-1}_{\boldsymbol{J}_{\tau(t)}}\right], \tag{3.1.1}$$

where $\tau(t) = \inf\{k \geq 1 : V_k + \xi_k > t\}$ is the first generation of the random walk along the spine that exceeds the level t. This suggests estimating $P(W > t)$ by

$$Z(t) = 1(\gamma(t) = \boldsymbol{J}_{\tau(t)}) e^{-\alpha V_{\tau(t)}} D^{-1}_{\boldsymbol{J}_{\tau(t)}},$$

which is unbiased when sampled under $\widetilde{P}$. Generating a single copy of $Z(t)$ requires simulating the entire tree $\mathcal{T}$ in length-lexicographic order until the node $\gamma(t)$ to determine whether or not the event $\{\gamma(t) = \boldsymbol{J}_{\tau(t)}\}$ occurs. Note that the remaining part of the estimator, namely $e^{-\alpha V_{\tau(t)}} D^{-1}_{\boldsymbol{J}_{\tau(t)}}$, only depends on the spine of $\mathcal{T}$, and the change of measure $\widetilde{P}$ was constructed so that the branch to most likely reach the level t first is part of the spine. Our approach to reducing the complexity of this agorithm is therefore to appropriately condition on the spine in (3.1.1) in order to find a representation that will allow us to generate only the spine of $\mathcal{T}$, while accounting for the probability that $\gamma(t)$ is not on the spine with some available approximation. The new estimator will be biased, but the computational complexity will be significantly reduced.

It was shown in Chapter 2 that the estimator $Z(t)$ has bounded relative error, namely

$$\limsup_{t \to \infty} \frac{\widetilde{\mathrm{Var}}\,(Z(t))}{P(W > t)^2} < \infty,$$

where $\widetilde{\mathrm{Var}}$ denotes variance under the measure $\widetilde{P}$, when Condition 5.1.1 holds along with the following (see Lemma 2.4.2).

Condition 3.1.1. *With $\alpha > 0$ as in Condition 5.1.1, $E\left[Q^{2\alpha}\left(\sum_{i=1}^{N} C_i^{\alpha}\right)^{-1}\right] < \infty$.*

Let $W^{(k)} = \max_{|\boldsymbol{j}|\le k}(S_{\boldsymbol{j}} + Y_{\boldsymbol{j}})$ for each $k \ge 0$, that is, $W_{\boldsymbol{i}}^{(k)}$ is a copy of W truncated at the kth generation. Denote its CDF with respect to P by

$$F_k(x) = P\left(W^{(k)} \le x\right).$$

Note in particular that $F_0(x) = P(Y \le x)$. For each $k \ge 1$ let

$$B_k^{\prec} = \{\boldsymbol{i} \in A_k : (\boldsymbol{i}|k-1) = \boldsymbol{J}_{k-1}, \boldsymbol{i} \prec \boldsymbol{J}_k\},$$
$$B_k^{\succ} = \{\boldsymbol{i} \in A_k : (\boldsymbol{i}|k-1) = \boldsymbol{J}_{k-1}, \boldsymbol{i} \succ \boldsymbol{J}_k\},$$

and let $B_k = B_k^{\prec} \cup B_k^{\succ}$. Because the distribution of the branching vector on nodes off the spine is unchanged by the change of measure (Lemma 2.1.1), we can obtain the following result that gives a representation for $P(W > t)$ in terms of the expectation of something that only depends on the spine and the functions $\{F_k : k \ge 0\}$.

Theorem 3.1.1. *For any $t > 0$,*

$$P(W > t) = \widetilde{E}\left[e^{-\alpha V_{\tau(t)}} D_{\boldsymbol{J}_{\tau(t)}}^{-1} \prod_{j=1}^{\tau(t)} \prod_{\boldsymbol{i}\in B_j^{\prec}} F_{\tau(t)-j}(t - S_{\boldsymbol{i}}) \prod_{\boldsymbol{j}\in B_j^{\succ}} F_{\tau(t)-j-1}(t - S_{\boldsymbol{j}})\right],$$

with the conventions $F_{-1}(x) \equiv 1$ and $\prod_{j=1}^{0} x_j \equiv 1$ for any values $\{x_j\}$.

If the functions $\{F_k : k \ge 0\}$ were known, the above representation would give an unbiased conditional Monte Carlo estimator for $P(W > t)$ that only requires computation of the spine up to generation $\tau(t) + 1$. In the case that approximations to these CDFs can be generated, one can form an estimator to approximate $P(W > t)$.

Suppose there exist estimators

$$\{\hat{F}_k : k \ge 0\}$$

for the CDFs $\{F_k : k \ge 0\}$. Based on Theorem 3.1.1, it is then natural to consider the estimator

$$\hat{Z}(t) = e^{-\alpha V_{\tau(t)}} D_{\boldsymbol{J}_{\tau(t)}}^{-1} \prod_{j=1}^{\tau(t)} \prod_{\boldsymbol{i}\in B_j^{\prec}} \hat{F}_{\tau(t)-j}(t - S_{\boldsymbol{i}}) \prod_{\boldsymbol{j}\in B_j^{\succ}} \hat{F}_{\tau(t)-j-1}(t - S_{\boldsymbol{j}}), \tag{3.1.2}$$

where again we use the conventions $\hat{F}_{-1}(x) \equiv 1$ and $\prod_{j=1}^{0} x_j \equiv 1$. Let $\tilde{\boldsymbol{\psi}}$ denote the distribution of the random vector $\boldsymbol{\psi}$ under the tilted measure $\widetilde{P}$ on the spine, that is

$$\tilde{\boldsymbol{\psi}} \sim \widetilde{P}\left(\boldsymbol{\psi}_{\boldsymbol{i}} \in \cdot \,|\, \boldsymbol{J}_{|\boldsymbol{i}|} = \boldsymbol{i}\right) = E\left[1(\boldsymbol{\psi} \in \cdot) \sum_{i=1}^{N} C_i^{\alpha}\right].$$

Then Table 3.1 gives the algorithm for generating one copy of $\hat{Z}$ conditional on knowing $\{\hat{F}_k : k \geq 0\}$.

Remark 3.1.1. By Lemma 2.4.1 in Chapter 2, $\tau(t) \sim t/\mu$ as $t \to \infty$ $\widetilde{P}$-a.s. Hence the computational complexity of generating a copy of $\hat{Z}(t)$ conditional on knowing $\{\hat{F}_k : k \geq 0\}$, measured by the number of independent copies of $\tilde{\psi}$ required to be generated, is asymptotically of order t/μ. Notably, this does not depend on $E[N]$ or $\widetilde{E}[N]$.

Of course, in implementation there is the practical consideration that in general it is not possible to have a countably infinite collection $\{\hat{F}_k : k \geq 0\}$ simulated a priori. Considering that $\tau(t) \sim t/\mu$ a.s., a natural solution is to choose some $K > t/\mu$ and generate $\{\hat{F}_k : 0 \leq k \leq K\}$, then set

$$\hat{F}_j = \hat{F}_K \quad \text{for} \;\; j \geq K.$$

We require the collection of estimators $\{\hat{F}_k : k \geq 0\}$ to satisfy certain properties.

Condition 3.1.2. *(a) For any $x \in \mathbb{R}$ and $k \geq 0$, $\hat{F}_k(x) \leq P(Y \leq x)$ a.s.*

(b) For some $0 < \lambda < \alpha$,

$$\delta_\lambda = \sup_{k \geq 0} E\left[d_1\left(F_{\lambda,k}, \hat{F}_{\lambda,k}\right)\right] = \sup_{k \geq 0} E\left[\int_0^1 \left|F_{\lambda,k}^{-1}(x) - \hat{F}_{\lambda,k}^{-1}(x)\right| dx\right]$$

can be made arbitrarily small, where

$$F_{\lambda,k}(x) = F_k\left(\lambda^{-1}\log x\right) \quad \text{and} \quad \hat{F}_{\lambda,k}(x) = \hat{F}_k\left(\lambda^{-1}\log x\right),$$

and where d_1 denotes the Wasserstein-1 distance.

Part (a) of the above condition says that the estimators preserve the stochastic ordering $P(Y > x) \leq P(W^{(k)} > x)$ for any k. Part (b) says that they are consistent in the sense that the Wasserstein-1 distance between $(R^{(k)})^\lambda$ and $(\hat{R}^{(k)})^\lambda$ vanishes uniformly in k for some λ, where $R^{(k)} = e^{W^{(k)}}$ and $\hat{R}^{(k)} = e^{\hat{W}^{(k)}}$, $\hat{W}^{(k)} \sim \hat{F}_k$. One reason to consider the Wasserstein distance after this transformation is that moments are not guaranteed otherwise, however we do have the following.

Lemma 3.1.2. *If W solves (3.0.2) and $0 < \gamma < \alpha$, then $E\left[e^{\gamma W}\right] < \infty$.*

Additionally, when $\rho_\gamma = E\left[\sum_{i=1}^N e^{\gamma X_i}\right] < 1$, the map

$$\Phi_\gamma(q, n, \{c_i\}, \{x_i\}) = q^\gamma \vee \bigvee_{i=1}^n c_i^\gamma x_i$$

Table 3.1: Hybrid Importance Sampling Algorithm

Input: $t > 0$ and $\{\hat{F}_k : k \geq 0\}$
Output: A single copy of $\hat{Z}$
Generate $(N, Q, C_1, \dots, C_N) \stackrel{\mathcal{D}}{=} \tilde{\boldsymbol{\psi}}$
Choose $j \in \{1, \dots, N\}$ w.p. C_j^α / D and set $\boldsymbol{J}_1 \leftarrow j$
Set $S_j \leftarrow \log C_j$ for $j = 1, \dots, N$
Initialize $V_0 \leftarrow 0$, $V_1 \leftarrow S_{\boldsymbol{J}_1}$, $\xi_0 \leftarrow \log Q$, $k \leftarrow 0$, $\boldsymbol{J}_0 \leftarrow \emptyset$
while $V_k + \xi_k \leq t$ **do**
    Update $k \leftarrow k + 1$
    Generate $(N_{\boldsymbol{J}_k}, Q_{\boldsymbol{J}_k}, C_{(\boldsymbol{J}_k,1)}, \dots, C_{(\boldsymbol{J}_k, N_{\boldsymbol{J}_k})}) \stackrel{\mathcal{D}}{=} \tilde{\boldsymbol{\psi}}$
    Choose $j \in \{1, \dots, N_{\boldsymbol{J}_k}\}$ w.p. $C^\alpha_{(\boldsymbol{J}_k, j)} / D_{\boldsymbol{J}_k}$ and set $\boldsymbol{J}_{k+1} \leftarrow (\boldsymbol{J}_k, j)$
    Set $\xi_k \leftarrow \log Q_{\boldsymbol{J}_k}$ and $S_{(\boldsymbol{J}_k, j)} \leftarrow S_{\boldsymbol{J}_k} + \log C_{(\boldsymbol{J}_k, j)}$ for $j = 1, \dots, N_{\boldsymbol{J}_k}$
    Set $V_{k+1} \leftarrow S_{\boldsymbol{J}_{k+1}}$
end while
if $k = 0$ **then**
    Set $\hat{Z} \leftarrow 1/D$
else
    Compute $\hat{F}_{k-j}(t - S_{\boldsymbol{i}})$ for $\boldsymbol{i} \in B_j^{\prec}$, $j = 1, \dots, k$
    Compute $\hat{F}_{k-j-1}(t - S_{\boldsymbol{i}})$ for $\boldsymbol{i} \in B_j^{\succ}$, $j = 1, \dots, k-1$
    Set $\hat{Z} \leftarrow (e^{-\alpha V_k} / D_{\boldsymbol{J}_k}) \prod_{j=1}^k \prod_{\boldsymbol{i} \in B_j^{\prec}} \hat{F}_{k-j}(t - S_{\boldsymbol{i}}) \prod_{\boldsymbol{j} \in B_j^{\succ}} \hat{F}_{k-j-1}(t - S_{\boldsymbol{j}})$
end if
Output $\hat{Z}$

---

has contractive properties that make the uniformity condition reasonable (see Section 3.1.2). Note that $R^\lambda = e^{\lambda W}$ solves the stochastic recursion

$$R^\lambda \stackrel{\mathcal{D}}{=} \Phi_\lambda \left(Q, N, \{C_i\}, \{R_i^\lambda\}\right)$$

when W solves (3.0.2).

Using estimates in place of the CDFs $\{F_k\}$ introduces bias into the estimate for $P(W > t)$ that ideally can be controlled. Since we seek an estimate of a rare event probability, the relevant measure is relative bias.

Definition 3.1.1. *The relative bias in an estimate $\hat{Z}(t)$ sampled under $\widetilde{P}$ for the probability $P(W > t)$ is given by*

$$\textit{Rel. Bias}\left(\hat{Z}(t)\right) = \left| \frac{\widetilde{E}\left[\hat{Z}(t)\right] - P(W > t)}{P(W > t)} \right|.$$

3.1.1 Properties

In the following results as well as in their proofs, the functions $u, v : \mathbb{R} \to [0, \infty)$ will be important, where

$$u(x) = e^{\alpha x} P(Y > x), \qquad v(x) = e^{2\alpha x} E\left[1(Y > x) D^{-1}\right]. \tag{3.1.3}$$

The function u and v have the following properties, the proofs of which can be found in Chapter 2.

Lemma 3.1.3. *Let u and v be as in* (3.1.3).

(i) Under Condition 5.1.1, u is d.R.i. on $\mathbb{R}$. Futhermore, if $h_n(x) = \widetilde{E}\left[u(x - V_n)\right]$ for each $n \in \mathbb{N}$ and $x \in \mathbb{R}$, then

$$B_h \doteq \sup_{x\in\mathbb{R}} \sum_{n=0}^{\infty} h_n(x) < \infty.$$

(ii) Under Conditions 5.1.1 and 3.1.1, v is d.R.i. on $\mathbb{R}$. Futhermore, if $g_n(x) = \widetilde{E}\left[v(x - V_n)\right]$ for each $n \in \mathbb{N}$ and $x \in \mathbb{R}$, then

$$B_g \doteq \sup_{x\in\mathbb{R}} \sum_{n=0}^{\infty} g_n(x) < \infty.$$

Additionally, we note the following result.

Lemma 3.1.4. *Under Condition 5.1.1 and the additional condition that $\left[\left(\sum_{i=1}^N C_i\right)^\alpha\right] < \infty$ if $\alpha > 1$ and $E\left[\left(\sum_{i=1}^N C_i^{\alpha/(1+\varepsilon)}\right)^{1+\varepsilon}\right] < \infty$ for some $0 < \varepsilon < 1$ if $0 < \alpha \le 1$, we have*

$$\inf_{t\ge 0} e^{\alpha t} P(W > t) > 0.$$

To bound the relative bias in $\hat{Z}(t)$, we assume the tilted density of the random walk increments along the spine are sufficiently nice. Namely, we have the following assumption.

Condition 3.1.3. *The density $\varphi(x)\, dx = \widetilde{P}\left(\hat{X}_1 \in dx\right)$ exists and is bounded on $\mathbb{R}$.*

The following is the main theoretical result of this chapter, which gives a bound on the relative bias in the estimate $\hat{Z}(t)$ for arbitrary estimators $\{\hat{F}_k : k \ge 0\}$ that satisfy Condition 3.1.2.

Theorem 3.1.5. *Assume Condtions 5.1.1, 3.1.1, and 3.1.3, and let $\{\hat{F}_k : k \ge 0\}$ be estimators for $\{F_k : k \ge 0\}$ that satisfy Condition 3.1.2 with δ_λ for some $0 < \lambda < \alpha$.*

Let $t > 0$ and $\alpha/2 < \beta < \alpha$ be such that $\rho_\beta < 1$. Recall that $Y = \log Q$. If $\widetilde{E}\left[(Y^-)^p\right] < \infty$ for $p \in \{1, 2\}$, then for any M, ε, and $\eta > 0$,

$$\textit{Rel. Bias}\left(\hat{Z}(t)\right) \le C_{p,\beta}\left(\frac{\left(\mu^{-1}t + 1\right)^{p/2}}{M^{p/2}} + \left(\frac{\delta_\lambda}{\varepsilon\eta} + P(Q \le \eta)\right) M + \varepsilon\right),$$

where

$$C_{p,\beta} = \left[\frac{2(2-\rho_\beta)E\left[Q^\beta\right]\widetilde{E}[N]\sqrt{c(\gamma,\beta)}}{(1-\rho_\beta)^2} \sup_{x\in\mathbb{R}} e^{-(\alpha-\beta)x}\varphi(x)\right] \vee \left[2B_h\widetilde{E}[N]\left(2 \vee \beta^{-1}\right)\right]$$

$$\vee \left[\sqrt{B_g}\left(2^{p/2} \vee \left(\frac{\left(1+e^{-\theta}\right)^{(p-1)/2}}{(1-e^{-\theta})^{p/2}} + \sqrt{2}\sum_{r=1}^{p}\left(\widetilde{E}\left[\left(3\mu^{-1}Y^{-}\right)^{r}\right]\right)^{1/2}\right)\right)\right] \cdot \sup_{t>0} \frac{e^{-\alpha t}}{P(W>t)} < \infty,$$

$$\theta = \frac{\mu^2(\alpha-\beta)^2}{24\left(\mu(\alpha-\beta)+3\rho_\beta e^{\mu(\alpha-\beta)}\right)},$$

$$c(\gamma,\beta) = \begin{cases} \left((1-\rho_\beta)^{-1}E\left[Q^\beta\right]\right)^{\gamma/\beta} & \textit{if } \gamma \le \beta, \\ (1-\rho_\gamma)^{-1}E\left[Q^\gamma\right] & \textit{if } \gamma > \beta, \end{cases}$$

$$\gamma = 2(\alpha-\beta).$$

It is clear from the above theorem that the relative bias in the estimate $\hat{Z}(t)$ can be made arbitrarily small uniformly in t on any bounded interval $[0,T]$ since

$$\limsup_{M\to\infty}\limsup_{\varepsilon,\eta\to 0}\limsup_{\delta_\lambda\to\infty}\sup_{0\le t\le T} \text{Rel. Bias}\left(\hat{Z}(t)\right) = 0.$$

3.1.2 The Population Dynamics Algorithm

The population dynamics algorithm provides a way to efficiently generate dependent samples from the distribution of $W^{(k)}$, and hence provide approximate empirical distribution functions with which to approximate F_k. We provide a brief description of the algorithm here as it pertains to the recursion (3.0.2), however it is applicable to many other recursions with a branching structure. For a more detailed analysis see [86]. The algorithm uses a bootstrapping approach to produce a sample $\hat{W}_1^{(j,m)},\ldots,\hat{W}_m^{(j,m)}$ of size m of random variables approximately distributed according to F_j for all $0 \le j \le k$, and it proceeds as in Table 3.2.

If we are able to simulate $\mathcal{P}^{(k,m)} = \{\hat{W}_i^{(k,m)} : 1 \le i \le m\}$ according to this algorithm, then we can use the empirical CDFs

$$\hat{F}_{k,m}(x) \doteq \frac{1}{m}\sum_{i=1}^{m} 1(\hat{W}_i^{(k,m)} \le x).$$

The idea is to generate the functions $\hat{F}_{0,m},\hat{F}_{1,m},\ldots,\hat{F}_{K,m}$ for some $K \ge 1$, and since $\tau(t)$ may be any positive integer with positive $\widetilde{P}$-probability, we will approximate F_k with $\hat{F}_{k\wedge K,m}$. Additionally, it makes sense to choose the intitial distribution $\mu_0 = P(Y \in \cdot)$. Then by construction, each $\hat{W}_i^{(k,m)}$ is lower bounded by a copy of $Y = \log Q$, and hence for any m and K the collection $\{\hat{F}_{k\wedge K,m} : k \ge 0\}$ satisfies Condition 3.1.2(a). Note that the empirical CDFs $\{\hat{F}_{k\wedge K,m} : k \ge 0\}$ are stored as the samples $\{\mathcal{P}^{(k\wedge K,m)} : k \ge 0\}$, and given any $x \in \mathbb{R}$, $\hat{F}_{k\wedge K,m}(x)$ can be calculated by

$$\hat{F}_{k\wedge K,m}(x) = \frac{1}{m}\cdot\left|\left\{w \in \mathcal{P}^{(k\wedge K,m)} : w \le x\right\}\right|.$$

Table 3.2: Population Dynamics Algorithm

Input: $k, m \in \mathbb{N}$ and a probability distribution μ_0 on $\mathbb{R}$
Output: Samples $\mathcal{P}^{(j,m)} = \{\hat{W}_i^{(j,m)} : 1 \leq i \leq m\}$ of size m approximately from the law of $W^{(j)}$, for $1 \leq j \leq k$
Generate an i.i.d. sample $\{\hat{W}_i^{(0,m)} : 1 \leq i \leq m\}$ from μ_0
Output $\mathcal{P}^{(0,m)}$
Initialize $j \leftarrow 1$
while $j \leq k$ **do**
    Generate an i.i.d. sample $\{(Q_i^{(j)}, N_i^{(j)}, C_{(i,1)}^{(j)}, \ldots, C_{(i,N_i^{(j)})}^{(j)}) : 1 \leq i \leq m\}$ from the law of $\boldsymbol{\psi}$
    Initialize $i \leftarrow 1$
    while $i \leq m$ **do**
        Sample $\{\hat{W}_{(i,r)}^{(j-1,m} : 1 \leq r \leq N_i^{(j)}\}$ uniformly with replacement from $\mathcal{P}^{(j-1,m)}$
        Set $\hat{W}_i^{(j,m)} \leftarrow \log Q_i^{(j)} \vee \bigvee_{r=1}^{N_i^{(j)}} (\log C_{(i,r)}^{(j)} + \hat{W}_{(i,r)}^{(j-1,m)})$
    end while
    Output $\mathcal{P}^{(j,m)}$
end while

Remark 3.1.2. Generating the samples $\mathcal{P}^{(0,m)}, \ldots, \mathcal{P}^{(K,m)}$ has a computational complexity of Km, measured in terms of the number of the number of copies of the branching vector $\boldsymbol{\psi}$ that need to be simulated. The bootstraping procedure eliminates dependence on $E[N]$. Consequently, by Remark 3.1.1, the complexity of generating a sample of size n of the estimator $\hat{Z}(t)$ using the population dynamics approximation is asymptotically of order

$$Km + \frac{nt}{\mu}.$$

Notably, there is no dependence on the distribution of N.

For $0 < \lambda < \alpha$ such that $\rho_\lambda = E\left[\sum_{i=1}^{N} C_i^\lambda\right] < 1$, we have

$$\delta_\lambda = \delta_\lambda(K,m) = \sup_{r \geq 0} E\left[d_1\left(e^{\lambda W^{(r)}}, e^{\lambda \hat{W}^{(r\wedge K,m)}}\right)\right],$$

where $W^{(r)} \sim F_r$ and $\hat{W}^{(r\wedge K,m)} \sim \hat{F}_{r\wedge K,m}$. By the following lemma, $\{\hat{F}_{k\wedge K,m} : k \geq 0\}$ also satisfy Condition 3.1.2(b).

Lemma 3.1.6. *Suppose $\alpha/2 < \lambda < \alpha$ is such that $\rho_\lambda < 1$, and suppose there exists $\lambda < \lambda' < \alpha$ such that $\rho_{\lambda'} < 1$. Then for some $\kappa = \kappa(\lambda, \lambda')$,*

$$\delta_\lambda(K,m) \leq \frac{(1+\rho_\lambda)E\left[Q^\lambda\right] + \rho_\lambda}{1-\rho_\lambda} \cdot \rho_\lambda^K + \frac{\kappa\left(E\left[Q^{\lambda'}\right]\right)^{\lambda/\lambda'}}{(1-\rho_\lambda)^2(1-\rho_{\lambda'})^{\lambda/\lambda'}} \cdot m^{-1/2}$$

In the case that $\{\hat{F}_k\} = \{\hat{F}_{k\wedge K,m}\}$ are generated using the population dynamics algorithm, denote the estimator $\hat{Z}(t) = \hat{Z}^{(K,m)}(t)$. The following is a refinement of Theorem 3.1.5 to the case when the population dynamics algorithm is used to approxiamte the CDFs $\{F_k\}$.

Theorem 3.1.7. *Let $\alpha/2 < \lambda, \beta < \alpha$ such that $\rho_\lambda, \rho_\beta < 1$, and suppose $\lambda < \lambda' < \alpha$ such that $\rho_{\lambda'} < 1$. Suppose $\widetilde{E}\left[(Y^-)^p\right] < \infty$ for $p \in \{1,2\}$. Let $C_{p,\beta}$ be the constant in Theorem 3.1.5 and $\kappa = \kappa(\lambda,\lambda')$ be the constant in Lemma 3.1.6. Let*

$$\widetilde{C}_{p,\beta,\lambda,\lambda'} = 2C_{p,\beta} \cdot \left[1 \vee \frac{(1+\rho_\lambda)E\left[Q^\lambda\right] + \rho_\lambda}{1-\rho_\lambda} \vee \frac{\kappa\left(E\left[Q^{\lambda'}\right]\right)^{\lambda/\lambda'}}{(1-\rho_\lambda)^2(1-\rho_{\lambda'})^{\lambda/\lambda'}} \right].$$

(i) If $P(Q \le q) = 0$ for som $q > 0$ (e.g. if $Q \equiv 1$), then

$$\textit{Rel. Bias}\left(\hat{Z}^{(K,m)}(t)\right) \le \frac{\widetilde{C}_{p,\beta,\lambda,\lambda'}}{\tilde{q}\wedge 1} \cdot \left(\left(\frac{t}{\mu}+1\right)^{p/2} m^{-\omega_1 p/2} + m^{\omega_2}\rho_\lambda^K + m^{-\min\{\frac{1}{2}-\omega_1-\omega_2,\omega_2\}} \right)$$

for any $\omega_1, \omega_2 > 0$ such that $\omega_1 + \omega_2 < 1/2$ and any $0 < \tilde{q} < q$.

(ii) If $P(Q \le \eta) \le c\eta^{\zeta_1}(-\log\eta)^{-\zeta_2}$ for some $c > 0$ and $\zeta_1, \zeta_2 \ge 0$, then

$$\begin{aligned}\textit{Rel. Bias}\left(\hat{Z}^{(K,m)}(t)\right) \le \frac{\widetilde{C}_{p,\beta,\lambda,\lambda'}(c\vee 1)}{\omega_3^{\zeta_2}} \cdot \Bigg(&\left(\frac{t}{\mu}+1\right)^{p/2} m^{-\omega_1 p/2} + m^{\omega_2+\omega_3}\rho_\lambda^K \\ &+ m^{-(\zeta_1\omega_3-\omega_1)}(\log m)^{-\zeta_2} + m^{-\min\{\frac{1}{2}-\omega_1-\omega_2-\omega_3,\omega_2\}} \Bigg)\end{aligned}$$

for any $\omega_1, \omega_2, \omega_3 > 0$ such that $\omega_1 + \omega_2 + \omega_3 < 1/2$ and $\omega_1 < \zeta_1\omega_3$.

3.2 Proofs

Here we give the proofs of all previous results. We start with the proofs of Theorem 3.1.1 and Lemma 3.1.4, followed by some auxiliary lemmas and the proof of Theorem 3.1.5. At the end are the proofs of the results pertaining to the Population Dynamics Algorithm, Lemma 3.1.6 and Theorem 3.1.7.

Proof of Theorem 3.1.1. Start by noting that we can write the event $\{\boldsymbol{J}_{\tau(t)} = \gamma(t)\}$ as

$$\begin{aligned}\left\{\boldsymbol{J}_{\tau(t)} = \gamma(t)\right\} &= \left\{ \bigvee_{\boldsymbol{i} \prec \boldsymbol{J}_{\tau(t)}} (S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}) \le t \right\} \\ &= \left\{ \max_{0\le k<\tau(t)-1} V_k + \Xi_k^{(\tau(t)-k)} \le t \right\} \cap \left\{ \max_{\boldsymbol{i}\in B^{\prec}_{\tau(t)}} S_{\boldsymbol{i}} + Y_{\boldsymbol{i}} \le t \right\},\end{aligned}$$

where

$$\Xi_k^{(r)} = \left(\bigvee_{\boldsymbol{i} \in B_{k+1}^{\prec}} S_{\boldsymbol{i}} - V_k + W_{\boldsymbol{i}}^{(r-1)} \right) \vee \left(\bigvee_{\boldsymbol{i} \in B_{k+1}^{\succ}} S_{\boldsymbol{i}} - V_k + W_{\boldsymbol{i}}^{(r-2)} \right), \qquad r \geq 2,$$

and the $W_{\boldsymbol{i}}^{(r)}$ are i.i.d. copies of

$$W^{(r)} = \bigvee_{k=0}^{r} \bigvee_{\boldsymbol{i} \in A_k} (S_{\boldsymbol{i}} + Y_{\boldsymbol{i}}),$$

under the probability P, independent of everything else. In other words, the $W_{\boldsymbol{i}}^{(r)}$ are i.i.d. with common distribution F_r. Next, let $\mathcal{H}_n = \sigma\left((\boldsymbol{J}_k, Q_{\boldsymbol{J}_k}, C_{(\boldsymbol{J}_k,1)}, C_{(\boldsymbol{J}_k,2)}, \ldots) : 0 \leq k \leq n \right)$ and note that

$$P(W > t) = \widetilde{E}\left[1(\boldsymbol{J}_{\tau(t)} = \gamma(t)) e^{-\alpha V_{\tau(t)}} D^{-1}_{\boldsymbol{J}_{\tau(t)}} \right] = \widetilde{E}\left[e^{-\alpha V_{\tau(t)}} D^{-1}_{\boldsymbol{J}_{\tau(t)}} \widetilde{P}\left(\boldsymbol{J}_{\tau(t)} = \gamma(t) \middle| \mathcal{H}_{\tau(t)} \right) \right].$$

To complete the proof note that since the $\{W_{\boldsymbol{i}}^{(r)}\}$ are independent of $\mathcal{H}_n$ for all n, then

$$\begin{aligned}
&\widetilde{P}\left(\boldsymbol{J}_{\tau(t)} = \gamma(t) \middle| \mathcal{H}_{\tau(t)} \right) \\
&= \prod_{k=0}^{\tau(t)-2} \widetilde{P}\left(V_k + \Xi_k^{(\tau(t)-k)} \leq t \middle| \mathcal{H}_{\tau(t)} \right) \prod_{\boldsymbol{i} \in B_{\tau(t)}^{\prec}} \widetilde{P}\left(S_{\boldsymbol{i}} + Y_{\boldsymbol{i}} \leq t \middle| \mathcal{H}_{\tau(t)} \right) \\
&= \prod_{k=0}^{\tau(t)-1} \prod_{\boldsymbol{i} \in B_{k+1}^{\prec}} \widetilde{P}\left(S_{\boldsymbol{i}} + W_{\boldsymbol{i}}^{(\tau(t)-k-1)} \leq t \middle| \mathcal{H}_{\tau(t)} \right) \prod_{j=0}^{\tau(t)-2} \prod_{\boldsymbol{j} \in B_{j+1}^{\succ}} \widetilde{P}\left(S_{\boldsymbol{j}} + W_{\boldsymbol{j}}^{(\tau(t)-j-2)} \leq t \middle| \mathcal{H}_{\tau(t)} \right) \\
&= \prod_{k=1}^{\tau(t)} \prod_{\boldsymbol{i} \in B_k^{\prec}} F_{\tau(t)-k}(t - S_{\boldsymbol{i}}) \prod_{\boldsymbol{j} \in B_k^{\succ}} F_{\tau(t)-k-1}(t - S_{\boldsymbol{j}}),
\end{aligned}$$

where $F_{-1}(x) \equiv 1$ and $\prod_{j=1}^{0} x_j \equiv 1$ for any values $\{x_j\}$. □

Proof of Lemma 3.1.4. By Theorem 3.4 in [64],

$$P(W > t) \sim He^{-\alpha t} \qquad \text{as } \ t \to \infty$$

for $H > 0$. The only way that $e^{\alpha t} P(W > t) = 0$ for finite t is for $P(W > t) = 0$, i.e. W is bounded, and since which would contradict the above asymptotic. Since $e^{\alpha t} P(W > t)$ is positive at infinity, $e^{\alpha t} P(W > t)$ must be uniformly lower bounded. □

3.2.1 Auxiliary Results

Here we collect some lemmas that will be useful in the proof of Theorem 3.1.5. For all that follows in this section and the next, fix $\beta \in (\alpha/2, \alpha)$ such that $\rho_\beta < 1$. Note that the following lemma implies Lemma 3.1.2.

Lemma 3.2.1. *Let $\gamma \in (0,\alpha)$. If $\gamma > \beta$, then $\rho_\gamma < 1$. Furthermore,*

$$E\left[e^{\gamma W}\right] \leq c(\gamma,\beta) < \infty,$$

where

$$c(\gamma,\beta) = \begin{cases} \left((1-\rho_\beta)^{-1} E\left[Q^\beta\right]\right)^{\gamma/\beta} & \text{if } \gamma \leq \beta, \\ (1-\rho_\gamma)^{-1} E\left[Q^\gamma\right] & \text{if } \gamma > \beta. \end{cases}$$

Proof. First note that $e^{\beta W} = R^\beta$ and

$$E\left[R^\beta\right] = E\left[\bigvee_{\boldsymbol{i}\in\mathcal{T}} Q_{\boldsymbol{i}}^\beta \Pi_{\boldsymbol{i}}^\beta\right] \leq \sum_{k=0}^\infty E\left[\sum_{\boldsymbol{i}\in A_k} Q_{\boldsymbol{i}}^\beta \Pi_{\boldsymbol{i}}^\beta\right] = \sum_{k=0}^\infty E\left[Q^\beta\right]\rho_\beta^k = \frac{E\left[Q^\beta\right]}{1-\rho_\beta}, \tag{3.2.1}$$

since $\rho_\beta < 1$. Then, when $\gamma \leq \beta$,

$$E\left[e^{\gamma W}\right] \leq \left(E\left[e^{\beta W}\right]\right)^{\gamma/\beta} \leq \left(\frac{E\left[Q^\beta\right]}{1-\rho_\beta}\right)^{\gamma/\beta}.$$

When $\gamma > \beta$, the convexity of $\theta \mapsto \rho_\theta$ and the fact that $\rho_\alpha = 1$ imply that $\rho_\gamma < 1$. Then, the calculation in (3.2.1) completes the proof. □

The next lemma is a version of Bernstein's inequality that only has moment conditions of the positive parts of the random variables involved.

Lemma 3.2.2. *If $\{\zeta_i\}$ are i.i.d. with $E[\zeta_1] = 0$ and $E\left[\left(\zeta_1^+\right)^n\right] \leq bc^{n-2}n!/2$ for each $n \geq 2$ for some $b, c > 0$, then $S_k = \sum_{i=1}^k \zeta_i$ satisfies*

$$P\left(S_k \geq xk\right) \leq \exp\left(-\frac{kx^2}{2(b+cx)}\right)$$

for all $x \geq 0$.

Proof. By Chernoff's inequality, for any $\lambda > 0$,

$$P\left(S_k \geq xk\right) \leq e^{-\lambda xk} E\left[e^{\lambda S_k}\right] = e^{-\lambda xk}\left(E\left[e^{\lambda\zeta_1}\right]\right)^k.$$

Note that for any $y \in \mathbb{R}$,

$$e^{\lambda y} = 1 + \lambda y + \sum_{n=2}^\infty \frac{\lambda^n y^n}{n!} \leq 1 + \lambda y + \sum_{n=2}^\infty \frac{\lambda^n (y\vee 0)^n}{n!},$$

and hence since $E[\zeta_1] = 0$, for each $\lambda \in (0, c^{-1})$,

$$\begin{aligned} E\left[e^{\lambda\zeta_1}\right] &\le 1 + \sum_{n=2}^{\infty} \frac{\lambda^n E\left[\left(\zeta_1^+\right)^n\right]}{n!} \\ &\le 1 + \sum_{n=2}^{\infty} \frac{\lambda^n b c^{n-2}}{2} = 1 + \frac{b\lambda^2}{2} \sum_{n=0}^{\infty} (\lambda c)^n \\ &= 1 + \frac{b\lambda^2}{2(1-\lambda c)}. \end{aligned}$$

Using $1 + y \le e^y$ we have

$$\begin{aligned} P\left(S_k \ge xk\right) &\le e^{-\lambda x k} \left(1 + \frac{b\lambda^2}{2(1-\lambda c)}\right)^k \\ &\le \exp\left(-k\left(\lambda x - \frac{b\lambda^2}{2(1-\lambda c)}\right)\right). \end{aligned}$$

Since the result is trivial when $x = 0$, suppose $x > 0$ and choose

$$\lambda = \frac{x}{b + cx} = \frac{1}{b/x + c} < \frac{1}{c}$$

to obtain the result. □

Lemma 3.2.3. *Let $\beta \in (0, \alpha)$ such that $\rho_\beta < \infty$. Then there are constants $b, c > 0$ such that for any $x \ge 0$,*

$$\widetilde{P}(\mu k - V_k \ge xk) \le \exp\left(-\frac{kx^2}{2(b + cx)}\right)$$

Proof. For each $n \in \mathbb{N}_+$, we have that

$$\begin{aligned} \widetilde{E}\left[((\mu - V_1)^+)^n\right] &= \int_0^\infty \widetilde{P}\left(\mu - V_1 > t^{1/n}\right) dt \le \int_0^\infty \widetilde{E}\left[e^{-(\alpha-\beta)V_1}\right] e^{-(\alpha-\beta)(t^{1/n} - \mu)} dt \\ &= \rho_\beta \int_0^\infty e^{-(\alpha-\beta)(t^{1/n} - \mu)} dt = \rho_\beta \frac{n}{(\alpha-\beta)^n} e^{\mu(\alpha-\beta)} \int_0^\infty u^{n-1} e^{-u} du \\ &= \rho_\beta \frac{n!}{(\alpha-\beta)^n} e^{\mu(\alpha-\beta)}. \end{aligned}$$

Hence, for any $n \ge 2$,

$$\widetilde{E}\left[((\mu - V_1)^+)^n\right] \le \frac{n! b c^{n-2}}{2}$$

for $b = \rho_\beta e^{\mu(\alpha-\beta)} (\alpha - \beta)^{-2}$ and

$$c = \frac{1}{\alpha - \beta} \sup_{n \ge 3} 2^{1/(n-2)} = \frac{2}{\alpha - \beta}.$$

The result then follows by the previous lemma. □

Lemma 3.2.4. *Let* $p \in \{1,2\}$. *If* $\widetilde{E}\left[(Y^-)^p\right] < \infty$, *then*

$$\widetilde{E}\left[\tau(t)^p\right] \le \left(2\mu^{-1}t\right)^p + \frac{\left(1+e^{-\theta}\right)^{p-1}}{\left(1-e^{-\theta}\right)^p} + 2\sum_{r=1}^{p} \widetilde{E}\left[\left(3\mu^{-1}Y^-\right)^r\right],$$

where

$$\theta = \frac{\mu^2(\alpha-\beta)^2}{24\left(\mu(\alpha-\beta)+3\rho_\beta e^{\mu(\alpha-\beta)}\right)}.$$

Proof. Let $\varepsilon, \delta > 0$ be small such that $\varepsilon + \delta < \mu$., and let $\gamma = \mu - \delta - \varepsilon \in (0,\mu)$. Then for all $k \ge t\delta^{-1}$,

$$\begin{aligned}
\widetilde{P}\left(\tau(t) > k\right) &\le \widetilde{P}\left(V_k + \xi_k \le k\delta\right) \\
&\le \widetilde{P}\left(V_k \le k(\delta+\varepsilon)\right) + \widetilde{P}\left(\xi_k \le -k\varepsilon\right) \\
&\le \widetilde{P}\left(k\mu - V_k \ge k\gamma\right) + \widetilde{P}\left(Y^- \ge k\varepsilon\right).
\end{aligned}$$

First let $p = 1$, and note that

$$\sum_{k=0}^{\infty} \widetilde{P}\left(Y^- \ge k\varepsilon\right) \le \frac{\widetilde{E}\left[Y^-\right]}{\varepsilon} + 1.$$

Using the previous lemma,

$$\begin{aligned}
\widetilde{E}\left[\tau(t)\right] &= \sum_{k=0}^{\infty} \widetilde{P}\left(\tau(t) > k\right) \\
&\le \frac{t}{\delta} - 1 + \sum_{k=\lfloor t\delta^{-1}\rfloor}^{\infty} \widetilde{P}(k\mu - V_k \ge k\gamma) + \sum_{k=\lfloor t\delta^{-1}\rfloor}^{\infty} \widetilde{P}\left(Y^- \ge k\varepsilon\right) \\
&\le \frac{t}{\delta} + \sum_{k=0}^{\infty} \exp\left(-\frac{k\gamma^2}{2(b+c\gamma)}\right) + \frac{\widetilde{E}\left[Y^-\right]}{\varepsilon} \\
&= \frac{t}{\delta} + \left(1 - \exp\left(-\frac{\gamma^2}{2(b+c\gamma)}\right)\right)^{-1} + \frac{\widetilde{E}\left[Y^-\right]}{\varepsilon}.
\end{aligned} \tag{3.2.2}$$

For $p = 2$, we use the tail sum formula

$$E\left[\zeta^2\right] = \sum_{k=0}^{\infty}(2k+1)P(\zeta > k)$$

for a nonnegative discrete random variable ζ. Note that

$$\sum_{k=0}^{\infty}(2k+1)\widetilde{P}\left(Y^- \ge k\varepsilon\right) = 1 + \sum_{k=1}^{\infty}(2k+1)\widetilde{P}\left(Y^- \ge k\varepsilon\right)$$

$$
\begin{aligned}
&= 1 + \sum_{k=0}^{\infty}(2k+3)\widetilde{P}\left(Y^- > k\varepsilon\right) \\
&= 1 + \sum_{k=0}^{\infty}(2k+1)\widetilde{P}\left(Y^- > k\varepsilon\right) + 2\sum_{k=0}^{\infty}\widetilde{P}\left(Y^- > k\varepsilon\right) \\
&= 1 + \frac{\widetilde{E}\left[(Y^-)^2\right]}{\varepsilon^2} + \frac{2\widetilde{E}\left[Y^-\right]}{\varepsilon}.
\end{aligned}
$$

Since $\sum_{k=0}^{n-1}(2k+1) = n^2 - 1$, we have

$$
\begin{aligned}
\widetilde{E}\left[\tau(t)^2\right] &= \sum_{k=0}^{\infty}(2k+1)\widetilde{P}(\tau(t) > k) \\
&\leq \sum_{k=0}^{\lfloor t\delta^{-1}\rfloor - 1}(2k+1) + \sum_{k=\lfloor t\delta^{-1}\rfloor}^{\infty}(2k+1)\widetilde{P}(k\mu - V_k \geq k\gamma) + \sum_{k=\lfloor t\delta^{-1}\rfloor}^{\infty}(2k+1)\widetilde{P}\left(Y^- \geq k\varepsilon\right) \\
&\leq \frac{t^2}{\delta^2} - 1 + \sum_{k=0}^{\infty}(2k+1)\exp\left(-\frac{k\gamma^2}{2(b+c\gamma)}\right) + \sum_{k=0}^{\infty}(2k+1)\widetilde{P}\left(Y^- \geq k\varepsilon\right) \\
&= \frac{t^2}{\delta^2} + \frac{1 + e^{-\gamma^2/(2(b+c\gamma))}}{\left(1 - e^{-\gamma^2/(2(b+c\gamma))}\right)^2} + \frac{\widetilde{E}\left[(Y^-)^2\right]}{\varepsilon^2} + \frac{2\widetilde{E}\left[Y^-\right]}{\varepsilon}. \qquad (3.2.3)
\end{aligned}
$$

Letting $\delta = \mu/2$, $\varepsilon = \mu/3$, and b and c have their values from Lemma 3.2.3 in (3.2.2) and (3.2.3) gives the result. □

Lemma 3.2.5. *Suppose that the density $\varphi(x) = \widetilde{P}\left(\hat{X}_1 \in dx\right)$ exists and is bounded on all of $\mathbb{R}$, and that $\beta \in (0, \alpha)$ such that $\rho_\beta < 1$. Let*

$$
\varphi^{(\infty)}(x) = \sum_{k=0}^{\infty}\varphi^{*(k)}(x).
$$

Then,

(i) $\sup_{x\in\mathbb{R}} e^{-(\alpha-\beta)x}\varphi(x) < \infty$, *and*

(ii) $\sup_{x\in\mathbb{R}} e^{-(\alpha-\beta)x}\varphi^{(\infty)}(x) \leq \frac{2-\rho_\beta}{1-\rho_\beta}\sup_{x\in\mathbb{R}} e^{-(\alpha-\beta)x}\varphi(x)$.

Proof. For part (i), note that

$$
\int_{-\infty}^{\infty} e^{-(\alpha-\beta)x}\varphi(x)\,dx = \widetilde{E}\left[e^{-(\alpha-\beta)\hat{X}_1}\right] = \rho_\beta < \infty,
$$

and so the boundedness of $\varphi(x)$ implies that $e^{-(\alpha-\beta)x}\varphi(x)$ is finite everywhere, hence integrability implies boundedness of $e^{-(\alpha-\beta)x}\varphi(x)$.

For part (ii), first note that

$$
\int_{-\infty}^{\infty} e^{-(\alpha-\beta)x}\varphi^{(\infty)}(x)\,dx = \sum_{k=0}^{\infty}\int_{-\infty}^{\infty} e^{-(\alpha-\beta)x}\varphi^{*(k)}(x)\,dx
$$

$$= \sum_{k=0}^{\infty} \widetilde{E}\left[e^{-(\alpha-\beta)V_k}\right] = \sum_{k=0}^{\infty} \rho_\beta^k = \frac{1}{1-\rho_\beta}.$$

The form of $\varphi^{(\infty)}$ allows us to write

$$\varphi^{(\infty)} * \varphi(x) = \sum_{k=0}^{\infty} \varphi^{*(k)} * \varphi(x) = \sum_{k=0}^{\infty} \varphi^{*(k+1)}(x) = \varphi^{(\infty)}(x) - \varphi(x).$$

hence for any $x \in \mathbb{R}$,

$$\begin{aligned}
e^{-(\alpha-\beta)x}\varphi^{(\infty)}(x) &= e^{-(\alpha-\beta)x}\varphi(x) + e^{-(\alpha-\beta)x}\int_{-\infty}^{\infty} \varphi^{(\infty)}(x-y)\varphi(y)\,dy \\
&= e^{-(\alpha-\beta)x}\varphi(x) + \int_{-\infty}^{\infty} e^{-(\alpha-\beta)(x-y)}\varphi^{(\infty)}(x-y)e^{-(\alpha-\beta)y}\varphi(y)\,dy \\
&\leq \sup_{y\in\mathbb{R}} e^{-(\alpha-\beta)y}\varphi(y) + \int_{-\infty}^{\infty} e^{-(\alpha-\beta)(x-y)}\varphi^{(\infty)}(x-y)\,dy \cdot \sup_{y\in\mathbb{R}} e^{-(\alpha-\beta)y}\varphi(y) \\
&= \left(1 + \frac{1}{1-\rho_\beta}\right)\sup_{y\in\mathbb{R}} e^{-(\alpha-\beta)y}\varphi(y) \\
&= \frac{2-\rho_\beta}{1-\rho_\beta}\sup_{y\in\mathbb{R}} e^{-(\alpha-\beta)y}\varphi(y).
\end{aligned}$$

□

3.2.2 Proof of Theorem 3.1.5

We start by defining a tree coupling as follows. For $k \geq 0$ let

$$\mathbf{X}_k = (\boldsymbol{J}_{k+1}, \boldsymbol{\psi}_{\boldsymbol{J}_k}) = (\boldsymbol{J}_{k+1}, Q_{\boldsymbol{J}_k}, N_{\boldsymbol{J}_k}, C_{(\boldsymbol{J}_k,1)}, C_{(\boldsymbol{J}_k,2)}, \ldots)$$

denote the spine process, i.e. the chosen path along with the associated copy of the branching vector. Let

$$\{\chi_{k,\boldsymbol{i}} : k \geq 0, \boldsymbol{i} \in U\}, \qquad \{\hat{\chi}_{k,\boldsymbol{i}} : k \geq 0, \boldsymbol{i} \in U\}$$

be independent i.i.d. collections of uniform random variables on $[0,1]$. Furthermore, for each $k \geq 0$ let

$$\{\hat{F}_{k,\boldsymbol{i}} : \boldsymbol{i} \in U\}$$

be an i.i.d. collection of copies of $\hat{F}_k$ independent of everything else. Then for each $k \geq 0$ and $\boldsymbol{i} \in U$, set

$$W_{\boldsymbol{i}}^{(k)} = F_k^{-1}(\chi_{k,\boldsymbol{i}}), \qquad \hat{W}_{\boldsymbol{i}}^{(k)} = \hat{F}_{k,\boldsymbol{i}}^{-1}(\hat{\chi}_{k,\boldsymbol{i}}).$$

In addition, let $\{\chi_{\infty,\boldsymbol{i}} : \boldsymbol{i} \in U\}$ be an i.i.d. collection of uniform on $[0,1]$ random variables independent of everything else and for each $\boldsymbol{i} \in U$, set

$$W_{\boldsymbol{i}} = W_{\boldsymbol{i}}^{(\infty)} = F^{-1}(\chi_{\infty,\boldsymbol{i}}).$$

By this construction, for each note $\boldsymbol{i} \in U$, we have independent copies of $W^{(k)}$, W, and the approximation $\hat{W}^{(k)}$ according to $\hat{F}_k$, and all these copies are mutually independent. We suppose that these collections of random variables are constructed on the original probability space, so that once the measure is changed to $\widetilde{P}$, they retain their distribution under P, i.e.

$$\widetilde{P}\left(W_{\boldsymbol{i}}^{(k)} \in \cdot\right) = P\left(W^{(k)} \in \cdot\right), \quad \widetilde{P}\left(\hat{W}_{\boldsymbol{i}}^{(k)} \in \cdot\right) = P\left(\hat{W}^{(k)} \in \cdot\right), \quad \text{and } \widetilde{P}(W_{\boldsymbol{i}} \in \cdot) = P(W \in \cdot)$$

for each $\boldsymbol{i} \in U$. Furthermore, we have that

$$\delta_\lambda = \sup_{k\geq 0} E\left[|e^{\lambda W_{\boldsymbol{i}}^{(k)}} - e^{\lambda \hat{W}_{\boldsymbol{i}}^{(k)}}|\right] = \sup_{k\geq 0} \widetilde{E}\left[|e^{\lambda W_{\boldsymbol{i}}^{(k)}} - e^{\lambda \hat{W}_{\boldsymbol{i}}^{(k)}}|\right]$$

for any $\boldsymbol{i} \in U$.

Remark 3.2.1. There is a technical note to make about the space on which the probability measure $\widetilde{P}$ is defined. The σ-field $\sigma\left(\bigcup_{k=0}^{\infty} \mathcal{G}_k\right)$ needs to be augmented to ensure the measurability of all random variables defined above. However, this does not show up in calculations because of their independence with the collection of branching vectors $\{\boldsymbol{\psi}_{\boldsymbol{i}} : \boldsymbol{i} \in U\}$. It will be convenient later in the proof to also assume that we have copies $\{\hat{X}_k' : k \geq 0\}$ of $\{\hat{X}_k : k \geq 0\}$ equal in distribution under $\widetilde{P}$. The measurability of this collection can be dealt with similarly.

Now, abusing notation, use $\{\mathbf{X}_k : k \geq 0\}$ to define

$$Z(t) = e^{-\alpha V_{\tau(t)}} D_{\boldsymbol{J}_{\tau(t)}}^{-1} \prod_{j=1}^{\tau(t)} \prod_{\boldsymbol{i}\in B_j^{\prec}} F_{\tau(t)-j}(t - S_{\boldsymbol{i}}) \prod_{\boldsymbol{j}\in B_j^{\succ}} F_{\tau(t)-j-1}(t - S_{\boldsymbol{j}}).$$

Furthermore, again abusing notation let

$$\hat{F}_k(x) = \widetilde{P}\left(\hat{W}_{\boldsymbol{i}}^{(k)} \leq x\right)$$

for any $\boldsymbol{i} \in U$, and define

$$\hat{Z}(t) = e^{-\alpha V_{\tau(t)}} D_{\boldsymbol{J}_{\tau(t)}}^{-1} \prod_{j=1}^{\tau(t)} \prod_{\boldsymbol{i}\in B_j^{\prec}} \widetilde{P}\left(\hat{W}_{\boldsymbol{i}}^{(\tau(t)-k)} \leq t - S_{\boldsymbol{i}} \middle| \mathbf{X}_0, \ldots, \mathbf{X}_{\tau(t)}\right)$$
$$\times \prod_{\boldsymbol{j}\in B_j^{\succ}} \widetilde{P}\left(\hat{W}_{\boldsymbol{j}}^{(\tau(t)-j-1)} \leq t - S_{\boldsymbol{j}} \middle| \mathbf{X}_0, \ldots, \mathbf{X}_{\tau(t)}\right)$$

$$= e^{-\alpha V_{\tau(t)}} D^{-1}_{\boldsymbol{J}_{\tau(t)}} \prod_{j=1}^{\tau(t)} \prod_{\boldsymbol{i}\in B_j^{\prec}} \hat{F}_{\tau(t)-j}(t - S_{\boldsymbol{i}}) \prod_{\boldsymbol{j}\in B_j^{\succ}} \hat{F}_{\tau(t)-j-1}(t - S_{\boldsymbol{j}}).$$

It is clear from this construction that $\widetilde{E}[Z(t)] = P(W > t)$ by Theorem 3.1.1 and that $\hat{Z}(t)$ has the same distribution under $\widetilde{P}$ as the estimator in (3.1.2). The key to this construction is that the random variables $Z(t)$ and $\hat{Z}(t)$ are coupled in the sense that they depend on the same spine process. We then can write

$$\begin{aligned} \text{Rel. Bias}\left(\hat{Z}(t)\right) &= \left| \frac{\widetilde{E}\left[\hat{Z}(t)\right] - P(W > t)}{P(W > t)} \right| \\ &\le \frac{\widetilde{E}\left[|\hat{Z}(t) - Z(t)|\right]}{P(W > t)} \\ &\le e^{\alpha t} \widetilde{E}\left[|\hat{Z}(t) - Z(t)|\right] \cdot \sup_{y\ge 0} \frac{1}{e^{\alpha y} P(W > y)}, \end{aligned} \tag{3.2.4}$$

where $\sup_{y\ge 0} \left(e^{\alpha y} P(W > y)\right)^{-1} < \infty$ by Lemma 3.1.4.

Theorem 3.1.5 will be proved with a series of lemmas to follow. For the first, define

$$\begin{aligned} G_0(x, \mathbf{X}_{k-1}) &= \prod_{\boldsymbol{i}\in B_k^{\prec}} F_0\left(x - S_{\boldsymbol{i}}\right), \qquad \text{and} \\ \hat{G}_0(x, \mathbf{X}_{k-1}) &= \prod_{\boldsymbol{i}\in B_k^{\prec}} \hat{F}_0\left(x - S_{\boldsymbol{i}}\right), \end{aligned}$$

and for $r \ge 1$:

$$\begin{aligned} G_r(x, \mathbf{X}_{k-1}) &= \prod_{\boldsymbol{i}\in B_k^{\prec}} F_r\left(x - S_{\boldsymbol{i}}\right) \prod_{\boldsymbol{j}\in B_k^{\succ}} F_{r-1}(x - S_{\boldsymbol{j}}), \qquad \text{and} \\ \hat{G}_r(x, \mathbf{X}_{k-1}) &= \prod_{\boldsymbol{i}\in B_k^{\prec}} \hat{F}_r\left(x - S_{\boldsymbol{i}}\right) \prod_{\boldsymbol{j}\in B_k^{\succ}} \hat{F}_{r-1}(x - S_{\boldsymbol{j}}), \end{aligned}$$

Lemma 3.2.6. *For any integer $M > 0$,*

$$\begin{aligned} e^{\alpha t} \widetilde{E}\left[|\hat{Z}(t) - Z(t)|\right] \le \sum_{k=0}^{M-1} \sum_{n=1}^{M-k} \widetilde{E}\left[u(t - V_k' - V_n) \left| G_{n-1}(t - V_k', \mathbf{X}_0) - \hat{G}_{n-1}(t - V_k', \mathbf{X}_0)\right|\right] \\ + \widetilde{E}\left[e^{-\alpha(V_{\tau(t)} - t)} D^{-1}_{\boldsymbol{J}_{\tau(t)}} 1(\tau(t) > M)\right], \end{aligned}$$

where u is given in (3.1.3) *and* $\{V'_k : k \geq 0\}$ *is a copy of* $\{V_k : k \geq 0\}$ *equal in distribution under* $\widetilde{P}$ *and independent of everything else (see Remark 3.2.1).*

Proof. For $0 \leq j \leq M$, define

$$f_j(t) \\ = \widetilde{E}\left[e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\mathbf{J}_{\tau(t)}} \left|\prod_{j=1}^{\tau(t)} \hat{G}_{\tau(t)-j}(t - V_{j-1}, \mathbf{X}_{j-1}) - \prod_{j=1}^{\tau(t)} G_{\tau(t)-j}(t - V_{j-1}, \mathbf{X}_{j-1})\right| 1(\tau(t) \leq j)\right].$$

We then have that

$$\begin{aligned} & e^{\alpha t}\widetilde{E}\left[|\hat{Z}(t) - Z(t)|\right] \\ &= \widetilde{E}\left[e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\mathbf{J}_{\tau(t)}} \left|\prod_{j=1}^{\tau(t)} \hat{G}_{\tau(t)-j}(t - V_{j-1}, \mathbf{X}_{j-1}) - \prod_{j=1}^{\tau(t)} G_{\tau(t)-j}(t - V_{j-1}, \mathbf{X}_{j-1})\right|\right] \\ &\leq f_M(t) + \widetilde{E}\left[e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\mathbf{J}_{\tau(t)}} 1(\tau(t) > M)\right]. \end{aligned}$$

Next, let

$$I_{\tau(t)}(\mathbf{X}_0, \ldots, \mathbf{X}_{\tau(t)}) = \left|\prod_{j=1}^{\tau(t)} \hat{G}_{\tau(t)-j}(t - V_{j-1}, \mathbf{X}_{j-1}) - \prod_{j=1}^{\tau(t)} G_{\tau(t)-j}(t - V_{j-1}, \mathbf{X}_{j-1})\right|$$

and note that

$$\begin{aligned} f_M(t) &= \widetilde{E}\left[1(\tau(t) = 1)e^{-\alpha(V_1-t)} D^{-1}_{\mathbf{J}_1} \left|\hat{G}_0(t, \mathbf{X}_0) - G_0(t, \mathbf{X}_0)\right|\right] \\ &\quad + \widetilde{E}\left[1(1 < \tau(t) \leq M)e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\mathbf{J}_{\tau(t)}} I_{\tau(t)}(\mathbf{X}_0, \ldots, \mathbf{X}_{\tau(t)})\right] \\ &= \widetilde{E}\left[1(\tau(t) = 1)e^{-\alpha(V_1-t)} D^{-1}_{\mathbf{J}_1} \left|\hat{G}_0(t, \mathbf{X}_0) - G_0(t, \mathbf{X}_0)\right|\right] \\ &\quad + \widetilde{E}\left[\widetilde{E}\left[1(1 < \tau(t) \leq M)e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\mathbf{J}_{\tau(t)}} I_{\tau(t)}(\mathbf{X}_0, \ldots, \mathbf{X}_{\tau(t)})\middle| \mathbf{X}_0\right]\right], \end{aligned}$$

and

$$\begin{aligned} & \widetilde{E}\left[1(1 < \tau(t) \leq M)e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\mathbf{J}_{\tau(t)}} I_{\tau(t)}(\mathbf{X}_0, \ldots, \mathbf{X}_{\tau(t)})\middle| \mathbf{X}_0\right] \\ & \leq \widetilde{E}\Bigg[1(1 < \tau(t) \leq M)e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\mathbf{J}_{\tau(t)}} \\ & \qquad \times\ G_{\tau(t)-1}(t - V_0, \mathbf{X}_0)\left|\prod_{j=1}^{\tau(t)-1} G_{\tau(t)-j-1}(t - V_j, \mathbf{X}_j) - \prod_{j=1}^{\tau(t)-1} \hat{G}_{\tau(t)-j-1}(t - V_j, \mathbf{X}_j)\right|\Bigg| \mathbf{X}_0\Bigg] \\ & \quad + \widetilde{E}\Bigg[1(1 < \tau(t) \leq M)e^{-\alpha(V_{\tau(t)}-t)} D^{-1}_{\mathbf{J}_{\tau(t)}} \end{aligned}$$

$$
\begin{aligned}
&\times \left| G_{\tau(t)-1}(t-V_0,\mathbf{X}_0) - \hat{G}_{\tau(t)-1}(t-V_0,\mathbf{X}_0)\right| \prod_{j=1}^{\tau(t)-1} \hat{G}_{\tau(t)-j-1}(t-V_j,\mathbf{X}_j)\Bigg| \mathbf{X}_0\Bigg] \\
\le\ & \widetilde{E}\Bigg[1(1<\tau(t)\le M)e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\mathbf{J}_{\tau(t)}} \\
&\times \left|\prod_{j=1}^{\tau(t)-1} G_{\tau(t)-j-1}(t-V_j,\mathbf{X}_j) - \prod_{j=1}^{\tau(t)-1} \hat{G}_{\tau(t)-j-1}(t-V_j,\mathbf{X}_j)\right|\Bigg| \mathbf{X}_0\Bigg] \\
&+ \widetilde{E}\left[1(1<\tau(t)\le M)e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\mathbf{J}_{\tau(t)}}\left|G_{\tau(t)-1}(t,\mathbf{X}_0)-\hat{G}_{\tau(t)-1}(t,\mathbf{X}_0)\right|\Big| \mathbf{X}_0\right] \\
=\ & 1(\xi_0\le t) f_{K-1}(t-V_1) \\
&+ \widetilde{E}\left[1(1<\tau(t)\le M)e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\mathbf{J}_{\tau(t)}}\left|G_{\tau(t)-1}(t,\mathbf{X}_0)-\hat{G}_{\tau(t)-1}(t,\mathbf{X}_0)\right|\Big| \mathbf{X}_0\right] \\
\le\ & f_{M-1}(t-V_1) + \widetilde{E}\left[1(1<\tau(t)\le M)e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\mathbf{J}_{\tau(t)}}\left|G_{\tau(t)-1}(t,\mathbf{X}_0)-\hat{G}_{\tau(t)-1}(t,\mathbf{X}_0)\right|\Big| \mathbf{X}_0\right].
\end{aligned}
$$

Now for $1 \le j \le M$ define

$$
\begin{aligned}
g_j(t) &= \widetilde{E}\left[1(\tau(t)=1)e^{-\alpha(V_1-t)}D^{-1}_{\mathbf{J}_1}\left|G_0(t,\mathbf{X}_0)-\hat{G}_0(t,\mathbf{X}_0)\right|\right] \\
&\quad + \widetilde{E}\left[1(1<\tau(t)\le j)e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\mathbf{J}_{\tau(t)}}\left|G_{\tau(t)-1}(t,\mathbf{X}_0)-\hat{G}_{\tau(t)-1}(t,\mathbf{X}_0)\right|\right] \\
&= \widetilde{E}\left[1(1\le\tau(t)\le j)e^{-\alpha(V_{\tau(t)}-t)}D^{-1}_{\mathbf{J}_{\tau(t)}}\left|G_{\tau(t)-1}(t,\mathbf{X}_0)-\hat{G}_{\tau(t)-1}(t,\mathbf{X}_0)\right|\right],
\end{aligned}
$$

and note that we have shown that

$$
f_M(t) \le g_M(t) + \widetilde{E}\left[f_{M-1}(t-V_1)\right].
$$

Since $f_0(t)=0$, iterating we obtain

$$
f_M(t) \le \sum_{k=0}^{M-1} \widetilde{E}\left[g_{M-k}(t-V_k)\right].
$$

We can bound g_j as follows:

$$
\begin{aligned}
g_j(t) &= \sum_{n=1}^{j} \widetilde{E}\left[1(\tau(t)=n)e^{-\alpha(V_n-t)}D^{-1}_{\mathbf{J}_n}\left|G_{n-1}(t,\mathbf{X}_0)-\hat{G}_{n-1}(t,\mathbf{X}_0)\right|\right] \\
&= \sum_{n=1}^{j} \widetilde{E}\left[1\left(\max_{0\le k<n} V_k+\xi_k \le t < V_n+\xi_n\right)e^{-\alpha(V_n-t)}D^{-1}_{\mathbf{J}_n}\left|G_{n-1}(t,\mathbf{X}_0)-\hat{G}_{n-1}(t,\mathbf{X}_0)\right|\right] \\
&\le \sum_{n=1}^{j} \widetilde{E}\left[u(t-V_n)\left|G_{n-1}(t,\mathbf{X}_0)-\hat{G}_{n-1}(t,\mathbf{X}_0)\right|\right].
\end{aligned}
$$

It follows that when $\{V_k' : k \geq 0\}$ is an independent copy of $\{V_k : k \geq 0\}$ then

$$\begin{aligned} f_M(t) &\leq \sum_{k=0}^{M-1} \widetilde{E}\left[\sum_{n=1}^{M-k} \widetilde{E}\left[u(t - V_k' - V_n) \left| G_{n-1}(t - V_k', \mathbf{X}_0) - \hat{G}_{n-1}(t - V_k', \mathbf{X}_0) \right| \middle| V_k' \right] \right] \\ &= \sum_{k=0}^{M-1} \sum_{n=1}^{M-k} \widetilde{E}\left[u(t - V_k' - V_n) \left| G_{n-1}(t - V_k', \mathbf{X}_0) - \hat{G}_{n-1}(t - V_k', \mathbf{X}_0) \right| \right]. \end{aligned} \qquad \square$$

For the next lemma define, for $r, s \geq 1$, the random variables

$$\begin{aligned} Z_s^{(r)} &= \bigvee_{\boldsymbol{i} \in B_{s+1}^{\prec}} (S_{\boldsymbol{i}} + W_{\boldsymbol{i}}^{(r)}) \vee \bigvee_{\boldsymbol{j} \in B_{s+1}^{\succ}} (S_{\boldsymbol{j}} + W_{\boldsymbol{j}}^{(r-1)}), \\ \hat{Z}_s^{(r)} &= \bigvee_{\boldsymbol{i} \in B_{s+1}^{\prec}} (S_{\boldsymbol{i}} + \hat{W}_{\boldsymbol{i}}^{(r)}) \vee \bigvee_{\boldsymbol{j} \in B_{s+1}^{\succ}} (S_{\boldsymbol{j}} + \hat{W}_{\boldsymbol{j}}^{(r-1)}), \end{aligned}$$

and for $r = 0$ define similarly

$$\begin{aligned} Z_s^{(0)} &= \bigvee_{i \in B_1^{\prec}} (S_i + W_i^{(0)}), \\ \hat{Z}_s^{(0)} &= \bigvee_{i \in B_1^{\succ}} (S_i + \hat{W}_i^{(0)}). \end{aligned}$$

Lemma 3.2.7. *For any $\varepsilon > 0$, $k \geq 0$, and $n \geq 1$,*

$$\begin{aligned} &\left| G_{n-1}(t - V_k', \mathbf{X}_0) - \hat{G}_{n-1}(t - V_k', \mathbf{X}_0) \right| \\ &\leq 2\left(G_{n-1}(t + \varepsilon - V_k', \mathbf{X}_0) - G_{n-1}(t - \varepsilon - V_k', \mathbf{X}_0) \right) + \widetilde{P}\left(|Z_0^{(n-1)} - \hat{Z}_0^{(n-1)}| > \varepsilon \middle| \mathbf{X}_0 \right). \end{aligned}$$

Proof. Define

$$U_k^{(n)} = V_k' + Z_0^{(n-1)} \qquad \text{and} \qquad \hat{U}_k^{(n)} = V_k' + \hat{Z}_0^{(n-1)},$$

and note that

$$G_{n-1}(t - V_k', \mathbf{X}_0) = \widetilde{P}\left(Z_0^{(n-1)} \leq t - V_k' \middle| \mathbf{X}_0, V_k' \right) = \widetilde{P}\left(U_k^{(n)} \leq t \middle| \mathbf{X}_0, V_k' \right),$$

and

$$\hat{G}_{n-1}(t - V_k', \mathbf{X}_0) = \widetilde{P}\left(\hat{Z}_0^{(n-1)} \leq t - V_k' \middle| \mathbf{X}_0, V_k' \right) = \widetilde{P}\left(\hat{U}_k^{(n)} \leq t \middle| \mathbf{X}_0, V_k' \right).$$

Observe that for any two random variables ζ_1 and ζ_2, we can write

$$|1(\zeta_1 \leq t) - 1(\zeta_2 \leq t)| = 1(\zeta_1 \leq t < \zeta_2) + 1(\zeta_2 \leq t < \zeta_1).$$

Hence,

$$\begin{aligned}
\left|1\left(U_k^{(n)} \le t\right) - 1\left(\hat{U}_k^{(n)} \le t\right)\right| &= 1\left(U_k^{(n)} \le t < \hat{U}_k^{(n)}, |U_k^{(n)} - \hat{U}_k^{(n)}| \le \varepsilon\right) \\
&\quad + 1\left(\hat{U}_k^{(n)} \le t < U_k^{(n)}, |U_k^{(n)} - \hat{U}_k^{(n)}| \le \varepsilon\right) + 1\left(|U_k^{(n)} - \hat{U}_k^{(n)}| > \varepsilon\right) \\
&\le 1\left(U_k^{(n)} \le t < U_k^{(n)} + \varepsilon\right) + 1\left(U_k^{(n)} - \varepsilon \le t < U_k^{(n)}\right) \\
&\quad + 1\left(|U_k^{(n)} - \hat{U}_k^{(n)}| > \varepsilon\right) \\
&\le 2 \cdot 1\left(U_k^{(n)} - \varepsilon \le t < U_k^{(n)} + \varepsilon\right) + 1\left(|U_k^{(n)} - \hat{U}_k^{(n)}| > \varepsilon\right) \\
&\le 2\left(1\left(U_k^{(n)} - \varepsilon \le t\right) - 1\left(U_k^{(n)} + \varepsilon \le t\right)\right) + 1\left(|U_k^{(n)} - \hat{U}_k^{(n)}| > \varepsilon\right) \\
&\le 2\left(1\left(U_k^{(n)} \le t + \varepsilon\right) - 1\left(U_k^{(n)} \le t - \varepsilon\right)\right) \\
&\quad + 1\left(|Z_0^{(n-1)} - \hat{Z}_0^{(n-1)}| > \varepsilon\right).
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\left|G_{n-1}(t - V_k', \mathbf{X}_0) - \hat{G}_{n-1}(t - V_k', \mathbf{X}_0)\right| \\
&= \left|\widetilde{E}\left[1\left(U_k^{(n)} \le t\right) - 1\left(\hat{U}_k^{(n)} \le t\right)\middle| \mathbf{X}_0, V_k'\right]\right| \\
&\le 2\left(\widetilde{P}\left(U_k^{(n)} \le t + \varepsilon\middle| \mathbf{X}_0, V_k'\right) - \widetilde{P}\left(U_k^{(n)} \le t - \varepsilon\middle| \mathbf{X}_0, V_k'\right)\right) + \widetilde{P}\left(|Z_0^{(n-1)} - \hat{Z}_0^{(n-1)}| > \varepsilon\middle| \mathbf{X}_0\right) \\
&= 2\left(G_{n-1}(t + \varepsilon - V_k', \mathbf{X}_0) - G_{n-1}(t - \varepsilon - V_k', \mathbf{X}_0)\right) + \widetilde{P}\left(|Z_0^{(n-1)} - \hat{Z}_0^{(n-1)}| > \varepsilon\middle| \mathbf{X}_0\right). \qquad \square
\end{aligned}$$

Lemma 3.2.8. *Suppose the density $\varphi(x) = \widetilde{P}\left(\hat{X}_1 \in dx\right)$ is bounded and $\beta \in (\alpha/2, \alpha)$ such that $\rho_\beta < 1$. Then for any $\varepsilon > 0$,*

$$\begin{aligned}
&\sum_{k=0}^{M-1} \sum_{n=1}^{M-k} \widetilde{E}\left[u(t - V_k' - V_n)\left(G_{n-1}(t + \varepsilon - V_k', \mathbf{X}_0) - G_{n-1}(t - \varepsilon - V_k', \mathbf{X}_0)\right)\right] \\
&\le \frac{2(2-\rho_\beta)E\left[Q^\beta\right]\widetilde{E}[N]\sqrt{c(\gamma,\beta)}}{(1-\rho_\beta)^2} \sup_{x \in \mathbb{R}} e^{-(\alpha-\beta)x}\varphi(x) \cdot \varepsilon,
\end{aligned}$$

where $\gamma = 2(\alpha - \beta)$ and $c(\gamma, \beta) < \infty$ is as in Lemma 3.2.1.

Proof. First note that

$$\begin{aligned}
\widetilde{E}\left[u(t - V_k' - V_n)\right| \mathbf{X}_0] &= \widetilde{E}\left[u(t - V_k' - (V_n - V_1) - V_1)\right| V_k', \mathbf{X}_0] \\
&= h_{n-1}(t - V_k' - V_1),
\end{aligned}$$

where h_{n-1} is as in Lemma 3.1.3. Consequently, we have that

$$
\begin{aligned}
&\sum_{k=0}^{M-1}\sum_{n=1}^{M-k} \widetilde{E}\left[u(t-V_k'-V_n)\left(G_{n-1}(t+\varepsilon-V_k',\mathbf{X}_0)-G_{n-1}(t-\varepsilon-V_k',\mathbf{X}_0)\right)\right] \\
&\le \sum_{k=0}^{M-1}\sum_{n=1}^{M-k} \widetilde{E}\left[u(t-V_k'-V_n)\widetilde{P}\left(t-\varepsilon \le V_k'+Z_0^{(n-1)} \le t+\varepsilon \,\middle|\, V_k',\mathbf{X}_0\right)\right] \\
&= \sum_{k=0}^{M-1}\sum_{n=1}^{M-k} \widetilde{E}\left[\widetilde{E}\left[u(t-V_k'-V_n)|\, V_k',\mathbf{X}_0\right]\widetilde{P}\left(t-\varepsilon \le V_k'+Z_0^{(n-1)} \le t+\varepsilon \,\middle|\, V_k',\mathbf{X}_0\right)\right] \\
&= \sum_{k=0}^{M-1}\sum_{n=1}^{M-k} \widetilde{E}\left[h_{n-1}(t-V_k'-V_1)\widetilde{P}\left(t-\varepsilon \le V_k'+Z_0^{(n-1)} \le t+\varepsilon \,\middle|\, V_k',\mathbf{X}_0\right)\right] \\
&= \sum_{k=0}^{M-1}\sum_{n=1}^{M-k} \widetilde{E}\left[h_{n-1}(t-V_k'-V_1)1\left(t-\varepsilon \le V_k'+Z_0^{(n-1)} \le t+\varepsilon\right)\right] \\
&= \sum_{k=0}^{M-1}\sum_{n=1}^{M-k} \widetilde{E}\left[\int_{t-\varepsilon-Z_0^{(n-1)}}^{t+\varepsilon-Z_0^{(n-1)}} h_{n-1}(t-x-V_1)\varphi^{*(k)}(x)\,dx\right] \\
&= \sum_{k=0}^{M-1}\sum_{n=1}^{M-k} \widetilde{E}\left[\int_{-\varepsilon}^{\varepsilon} h_{n-1}(-x+Z_0^{(n-1)}-V_1)\varphi^{*(k)}(x)\,dx\right].
\end{aligned}
$$

Now, note that for $\beta \in (0,\alpha)$ such that $\rho_\beta < 1$,

$$
u(x) = e^{\alpha x}P(Y > x) \le E\left[Q^\beta\right]e^{(\alpha-\beta)x},
$$

and so

$$
h_n(x) = \widetilde{E}\left[u(x-V_n)\right] \le E\left[Q^\beta\right]e^{(\alpha-\beta)x}\widetilde{E}\left[e^{-(\alpha-\beta)V_n}\right] = E\left[Q^\beta\right]e^{(\alpha-\beta)x}\rho_\beta.
$$

It follows that, since $Z_0^{(n-1)}$ is monotone a.s. in n,

$$
\begin{aligned}
&\widetilde{E}\left[\int_{-\varepsilon}^{\varepsilon} h_{n-1}(-x+Z_0^{(n-1)}-V_1)\varphi^{*(k)}(x)\,dx\right] \\
&\le E\left[Q^\beta\right]\rho_\beta^{n-1}\widetilde{E}\left[e^{(\alpha-\beta)(Z_0^{(n-1)}-V_1)}\right]\int_{-\varepsilon}^{\varepsilon} e^{-(\alpha-\beta)x}\varphi^{*(k)}(x)\,dx \\
&\le E\left[Q^\beta\right]\rho_\beta^{n-1}\widetilde{E}\left[e^{(\alpha-\beta)(Z_0^{(\infty)}-V_1)}\right]\int_{-\varepsilon}^{\varepsilon} e^{-(\alpha-\beta)x}\varphi^{*(k)}(x)\,dx.
\end{aligned}
$$

Now let $\gamma = 2(\alpha-\beta)$, and note that $\gamma \in (0,\alpha)$ since $\beta \in (\alpha/2,\alpha)$. Also recall that

$$
Z_0^{(\infty)} = \bigvee_{i \in B_1}(S_i + W_i^{(\infty)}),
$$

where $\widetilde{P}\left(W_i^{(\infty)} \le x\right) = P\left(W_i^{(\infty)} \le x\right) = P(W \le x)$ for each i and $x \in \mathbb{R}$. Then, using Wald's identity,

$$\begin{aligned}
\widetilde{E}\left[e^{(\alpha-\beta)(Z_0^{(\infty)}-V_1)}\right] &= \widetilde{E}\left[\bigvee_{i\in B_1} e^{(\alpha-\beta)(S_i-V_1+W_i^{(\infty)})}\right] \\
&\le \widetilde{E}\left[\bigvee_{i\in B_1} e^{(\alpha-\beta)(S_i-V_1)} \bigvee_{i=1}^{N_\emptyset} e^{(\alpha-\beta)W_i^{(\infty)}}\right] \\
&\le \left(\widetilde{E}\left[\bigvee_{i\in B_1} e^{\gamma(S_i-V_1)}\right]\right)^{1/2} \cdot \left(\widetilde{E}\left[\sum_{i=1}^{N_\emptyset} e^{\gamma W_i^{(\infty)}}\right]\right)^{1/2} \\
&\le \left(\widetilde{E}[N]\right)^{1/2} \cdot \left(\widetilde{E}[N]E\left[e^{\gamma W}\right]\right)^{1/2} \\
&\le \widetilde{E}[N]\sqrt{c(\gamma,\beta)},
\end{aligned}$$

where we used Lemma 2.5.6 and Lemma 3.2.1 (which defines $c(\gamma,\beta)$). Furthermore, from Lemma 3.2.5,

$$\sup_{x\in\mathbb{R}} \sum_{k=0}^{\infty} e^{-(\alpha-\beta)x}\varphi^{*(k)}(x) \le \frac{2-\rho_\beta}{1-\rho_\beta} \sup_{x\in\mathbb{R}} e^{-(\alpha-\beta)x}\varphi(x) < \infty.$$

Finally, we have that

$$\begin{aligned}
&\sum_{k=0}^{M-1}\sum_{n=1}^{M-k} \widetilde{E}\left[\int_{-\varepsilon}^{\varepsilon} h_{n-1}(-x+Z_0^{(n-1)}-V_1)\varphi^{*(k)}(x)\,dx\right] \\
&\le E\left[Q^\beta\right]\widetilde{E}[N]\sqrt{c(\gamma,\beta)}\sum_{k=0}^{M}\sum_{n=1}^{M-k}\rho_\beta^{n-1}\int_{-\varepsilon}^{\varepsilon} e^{-(\alpha-\beta)x}\varphi^{*(k)}(x)\,dx \\
&\le E\left[Q^\beta\right]\widetilde{E}[N]\sqrt{c(\gamma,\beta)} \cdot \frac{1}{1-\rho_\beta} \cdot \int_{-\varepsilon}^{\varepsilon}\sum_{k=0}^{M} e^{-(\alpha-\beta)x}\varphi^{*(k)}(x)\,dx \\
&\le \frac{E\left[Q^\beta\right]\widetilde{E}[N]\sqrt{c(\gamma,\beta)}}{1-\rho_\beta} \cdot \frac{2-\rho_\beta}{1-\rho_\beta}\sup_{x\in\mathbb{R}} e^{-(\alpha-\beta)x}\varphi(x)\cdot 2\varepsilon. \qquad \square
\end{aligned}$$

Lemma 3.2.9. *Recall the function $h_n(x)$ and $B_h = \sup_{x\in\mathbb{R}}\sum_{n=0}^{\infty} h_n(x) < \infty$ from Lemma 3.1.3. For any $\varepsilon > 0$ and $\eta > 0$,*

$$\begin{aligned}
&\sum_{k=0}^{M-1}\sum_{n=1}^{M-k} \widetilde{E}\left[u(t-V_k'-V_n)\widetilde{P}\left(|Z_0^{(n-1)}-\hat{Z}_0^{(n-1)}| > \varepsilon \,\middle|\, \mathbf{X}_0\right)\right] \\
&\le 2B_h\widetilde{E}[N]\left(2\vee\beta^{-1}\right)\left(\frac{\delta}{\eta\varepsilon} + P(Q\le\eta)\right)M.
\end{aligned}$$

Proof. Note that

$$|Z_0^{(n-1)} - \hat{Z}_0^{(n-1)}| \leq \bigvee_{\in B_1^{\prec}} |W_i^{(n-1)} - \hat{W}_i^{(n-1)}| + \bigvee_{i \in B_1^{\succ}} |W_i^{(n-2)} - \hat{W}_i^{(n-2)}|,$$

and hence for any $\eta > 0$,

$$\begin{aligned}
\widetilde{P}\left(|Z_0^{(n-1)} - \hat{Z}_0^{(n-1)}| > \varepsilon \,\middle|\, \mathbf{X}_0\right) &\leq 2 \sup_{r \geq 0} \widetilde{P}\left(\bigvee_{i=1}^{N_\emptyset} |W_i^{(r)} - \hat{W}_i^{(r)}| \,\middle|\, \mathbf{X}_0\right) \\
&\leq 2N_\emptyset \sup_{r \geq 0} P\left(|W^{(r)} - \hat{W}^{(r)}| > \varepsilon\right) \\
&\leq 2N_\emptyset \left[\sup_{r \geq 0} P\left(|W^{(r)} - \hat{W}^{(r)}| > \varepsilon, \hat{R}^{(r)} \wedge R^{(r)} > \eta\right)\right. \\
&\qquad \left. + \sup_{r \geq 0} P\left(\hat{R}^{(r)} \wedge R^{(r)} \leq \eta\right)\right].
\end{aligned}$$

Since $\log(x)$ is Lipschitz on $x > \eta$ with Lipschitz constant η^{-1}, on $\{\hat{R}^{(r)} \wedge R^{(r)} > \eta\}$ we have

$$|\hat{W}^{(r)} - W^{(r)}| = \frac{1}{\beta} |\log(\hat{R}^{(r)})^\beta - \log(R^{(r)})^\beta| \leq \frac{1}{\beta\eta} |(\hat{R}^{(r)})^\beta - (R^{(r)})^\beta|.$$

Furthermore, since $R^{(r)}$ is increasing in r, $R^{(r)} \geq R^{(0)} \stackrel{\mathcal{D}}{=} Q$. By Condition 3.1.2(a), $\hat{R}^{(r)}$ is also lower bounded by copy of Q independent $R^{(0)}$. If $\tilde{Q}$ is an independent copy of Q, then

$$P\left(\hat{R}^{(r)} \wedge R^{(r)} \leq \eta\right) \leq P\left(Q \wedge \tilde{Q} \leq \eta\right) \leq 2P(Q \leq \eta).$$

Hence,

$$\begin{aligned}
\widetilde{P}\left(|Z_0^{(n-1)} - \hat{Z}_0^{(n-1)}| > \varepsilon \,\middle|\, \mathbf{X}_0\right) &\leq 2N_\emptyset \left(\sup_{r \geq 0} P\left(|(\hat{R}^{(r)})^\beta - (R^{(r)})^\beta| > \beta\eta\varepsilon\right) + 2P(Q \leq \eta)\right) \\
&\leq 2N_\emptyset \left(\frac{1}{\beta\eta\varepsilon} \sup_{r \geq 0} E\left[|(\hat{R}^{(r)})^\beta - (R^{(r)})^\beta|\right] + 2P(Q \leq \eta)\right) \\
&= 2N_\emptyset \left(\frac{\delta}{\beta\eta\varepsilon} + 2P(Q \leq \eta)\right).
\end{aligned}$$

Then, recalling the function h_n in Lemma 3.1.3(ii),

$$\begin{aligned}
&\sum_{k=0}^{M-1} \sum_{n=1}^{M-k} \widetilde{E}\left[u(t - V_k' - V_n) \widetilde{P}\left(|Z_0^{(n-1)} - \hat{Z}_0^{(n-1)}| > \varepsilon \,\middle|\, \mathbf{X}_0\right)\right] \\
&\leq 2\left(\frac{\delta}{\beta\eta\varepsilon} + 2P(Q \leq \eta)\right) \sum_{k=0}^{M} \sum_{n=1}^{M-k} \widetilde{E}\left[u(t - V_k' - V_n) N_\emptyset\right]
\end{aligned}$$

$$
\begin{aligned}
&= 2\left(\frac{\delta}{\beta\eta\varepsilon} + 2P(Q \le \eta)\right)\sum_{k=0}^{M}\sum_{n=1}^{M-k}\widetilde{E}\left[\widetilde{E}\left[u(t - V_k' - (V_n - V_1) - V_1)\right|\mathbf{X}_0, V_k'\right]N_\emptyset\right] \\
&= 2\left(\frac{\delta}{\beta\eta\varepsilon} + 2P(Q \le \eta)\right)\sum_{k=0}^{M}\sum_{n=1}^{M-k}\widetilde{E}\left[h_{n-1}(t - V_k' - V_1)N_\emptyset\right] \\
&\le 2\left(\frac{\delta}{\beta\eta\varepsilon} + 2P(Q \le \eta)\right)\sum_{k=0}^{M}\widetilde{E}\left[N_\emptyset\sum_{n=0}^{\infty}h_n(t - V_k' - V_1)\right] \\
&\le 2\left(\frac{\delta}{\beta\eta\varepsilon} + 2P(Q \le \eta)\right)M\widetilde{E}\left[N\right]B_h,
\end{aligned}
$$

where $B_h = \sup_x \sum_{n=0}^{\infty} h_n(x) < \infty$. This completes the proof. □

Lemma 3.2.10. *Let $p \in \{1, 2\}$. If $\widetilde{E}\left[(Y^-)^p\right] < \infty$, then*

$$
\begin{aligned}
&\widetilde{E}\left[e^{-\alpha(V_{\tau(t)} - t)}D^{-1}_{\boldsymbol{J}_{\tau(t)}}1(\tau(t) > M)\right] \\
&\le \sqrt{B_g}\left(2^{p/2} \vee \left(\frac{\left(1 + e^{-\theta}\right)^{(p-1)/2}}{\left(1 - e^{-\theta}\right)^{p/2}} + \sum_{r=1}^{p}\left(\widetilde{E}\left[\left(3\mu^{-1}Y^-\right)^r\right]\right)^{1/2}\right)\right) \cdot \frac{\left(\mu^{-1}t\right)^{p/2} + 1}{M^{p/2}}.
\end{aligned}
$$

Proof. Recall the functions v and g from Lemma 3.1.3. By the Cauchy Schwarz inequality and Markov's inequality,

$$
\begin{aligned}
\widetilde{E}\left[e^{-\alpha(V_{\tau(t)} - t)}D^{-1}_{\boldsymbol{J}_{\tau(t)}}1(\tau(t) > M)\right] &\le \left(\widetilde{E}\left[e^{-2\alpha(V_{\tau(t)} - t)}D^{-2}_{\boldsymbol{J}_{\tau(t)}}\right]\right)^{1/2}\widetilde{P}\left(\tau(t) > M\right)^{1/2} \\
&\le \sqrt{B_g} \cdot \frac{\left(\widetilde{E}\left[\tau(t)^p\right]\right)^{1/2}}{M^{p/2}},
\end{aligned}
$$

where we used the fact that

$$
\begin{aligned}
\widetilde{E}\left[e^{-2\alpha(V_{\tau(t)} - t)}D^{-2}_{\boldsymbol{J}_{\tau(t)}}\right] &= \sum_{n=0}^{\infty}\widetilde{E}\left[1(\tau(t) = n)e^{-2\alpha(V_n - t)}D^{-2}_{\boldsymbol{J}_n}\right] \\
&\le \sum_{n=0}^{\infty}\widetilde{E}\left[e^{-2\alpha(V_n - t)}\widetilde{E}\left[1(V_n + \xi_n > t)D^{-2}_{\boldsymbol{J}_n}\right|\mathcal{G}_n\right]\right] \\
&= \sum_{n=0}^{\infty}\widetilde{E}\left[v(t - V_n)\right] \le \sup_{x\in\mathbb{R}}\sum_{n=0}^{\infty}g_n(x) = B_g.
\end{aligned}
$$

By Lemma 3.2.4,

$$
\begin{aligned}
\left(\widetilde{E}\left[\tau(t)^p\right]\right)^{1/2} &\le \left(\left(2\mu^{-1}t\right)^p + \frac{\left(1 + e^{-\theta}\right)^{p-1}}{\left(1 - e^{-\theta}\right)^p} + 2\sum_{r=1}^{p}\widetilde{E}\left[\left(3\mu^{-1}Y^-\right)^r\right]\right)^{1/2} \\
&\le \left(2^{p/2} \vee \left(\frac{\left(1 + e^{-\theta}\right)^{(p-1)/2}}{\left(1 - e^{-\theta}\right)^{p/2}} + \sqrt{2}\sum_{r=1}^{p}\left(\widetilde{E}\left[\left(3\mu^{-1}Y^-\right)^r\right]\right)^{1/2}\right)\right)\left(\frac{t}{\mu} + 1\right)^{p/2},
\end{aligned}
$$

which gives the result. □

We can now complete the proof of Theorem 3.1.5.

Proof of Theorem 3.1.5. The result follows on combining (3.2.4) with Lemmas 3.2.6, 3.2.7, 3.2.8, 3.2.9, and 3.2.10 □

3.2.3 The Population Dynamics Case

Finally, we prove the results pertaining to the use of the population dynamics in the approximation of the $\{F_k\}$, namely Lemma 3.1.6 and Theorem 3.1.7.

Proof of Lemma 3.1.6. First write

$$\delta_\lambda(K,m) \leq \sup_{r\geq 0} E\left[d_1\left(F_{\lambda,r\wedge K}, \hat{F}_{\lambda,r\wedge K,m}\right)\right] + \sup_{r\geq 0} E\left[d_1\left(F_{\lambda,r}, F_{\lambda,r\wedge K}\right)\right].$$

We start with the first term.

Let $R = e^W$, $R^{(r)} = e^{W^{(r)}}$, and $\hat{R}^{(r,m)} = e^{\hat{W}^{(r,m)}}$. R^λ satsfies the fixed point equation $R^\lambda \stackrel{\mathcal{D}}{=} \Phi_\lambda(Q, N, \{C_i\}, \{R_i^\lambda\})$, where R_i are i.i.d. copies of R, for the map

$$\Phi_\lambda(q, n, \{c_i\}, \{x_i\}) = q^\lambda \vee \bigvee_{i=1}^{n} c_i^\lambda x_i.$$

If $\{(X_i, Y_i) : i \geq 1\}$ is an i.i.d. sequence of vectors in $\mathbb{R}_+ \times \mathbb{R}_+$ independent of $(Q, N, \{C_i\})$, then

$$E\left[\left|Q^\lambda \vee \bigvee_{i=1}^{N} C_i^\lambda X_i - Q^\lambda \vee \bigvee_{i=1}^{N} C_i^\lambda Y_i\right|\right] \leq E\left[\sum_{i=1}^{N} C_i^\lambda |X_i - Y_i|\right] = \rho_\lambda E\left[|X_1 - Y_1|\right].$$

Hence, applying Theorem 2.8 in [86] with $p = 1$ and $q = \lambda'/\lambda$ and noting that $\frac{\alpha}{2} < \lambda < \lambda' < \alpha$ implies that $q < 2$, there exists some $\kappa = \kappa(\lambda, \lambda')$ such that

$$E\left[d_1\left((R^{(r)})^\lambda, (\hat{R}^{(r,m)})^\lambda\right)\right] \leq \kappa \left(\sum_{j=0}^{r} \rho_\lambda^j\right)\left(\sum_{j=0}^{r} \rho_\lambda^{r-j}\left(E\left[|R^{(j)}|^{\lambda'}\right]\right)^{\lambda/\lambda'}\right) \cdot m^{-\min\{1-\lambda/\lambda', 1/2\}}.$$

Note that for all j,

$$E\left[|R^{(j)}|^{\lambda'}\right] \leq E\left[\sum_{k=0}^{\infty} \sum_{\mathbf{i}\in A_k} Q_{\mathbf{i}}^{\lambda'} \Pi_{\mathbf{i}}^{\lambda'}\right] = E\left[Q^{\lambda'}\right] \sum_{k=0}^{\infty} \rho_{\lambda'}^k = \frac{E\left[Q^{\lambda'}\right]}{1-\rho_{\lambda'}}.$$

Since $\lambda'/\lambda < 2$, $\min\{1-\lambda/\lambda', 1/2\} = 1/2$, and so

$$E\left[d_1\left((R^{(r)})^\lambda, (\hat{R}^{(r,m)})^\lambda\right)\right] \leq \kappa \left(\sum_{j=0}^{\infty} \rho_\lambda^j\right)\left(\sum_{j=0}^{\infty} \rho_\lambda^j \left(\frac{E\left[Q^{\lambda'}\right]}{1-\rho_{\lambda'}}\right)^{\lambda/\lambda'}\right) \cdot m^{-1/2}$$

$$= \frac{\kappa \left(E\left[Q^{\lambda'} \right] \right)^{\lambda/\lambda'}}{(1-\rho_\lambda)^2 (1-\rho_{\lambda'})^{\lambda/\lambda'}} \cdot m^{-1/2},$$

and the bound is uniform in r.

For the other term, let U be a uniform random variable over $[0,1]$ and let $R^{(r)} = \left(F_{\lambda,r}^{-1}(U) \right)^{1/\lambda}$ for each $r \geq 0$ and $R = \left(F_{\lambda,\infty}^{-1}(U) \right)^{1/\lambda}$, i.e. $R \stackrel{\mathcal{D}}{=} e^W$ where W solves (3.0.2). Note that $R^{(r)}$ are increasing in r to R. Then we have

$$\begin{aligned} \sup_{r \geq 0} E\left[d_1 \left(F_{\lambda,r}, F_{\lambda, r\wedge K} \right) \right] &= \sup_{r \geq 0} E\left[|(R^{(r)})^\lambda - (R^{(r\wedge K)})^\lambda| \right] \\ &= E\left[|(R^{(K)})^\lambda - R^\lambda| \right]. \end{aligned}$$

By Theorem 2.5 in [86] and its proof,

$$E\left[|(R^{(K)})^\lambda - R^\lambda| \right] \leq \frac{\rho_\lambda E\left[Q^\lambda\right] + E\left[Q^\lambda \vee \bigvee_{i=1}^N C_i^\lambda \right]}{1-\rho_\lambda} \rho_\lambda^K \leq \frac{(1+\rho_\lambda) E\left[Q^\lambda\right] + \rho_\lambda}{1-\rho_\lambda} \rho_\lambda^K .$$

This completes the proof. □

Proof of Theorem 3.1.7. When $P(Q \leq q) = 0$ for some $q > 0$, choose some $0 < \tilde{q} < q$ and set $\eta = \tilde{q}$ in the bound in Theorem 3.1.5. The result then follows by Lemma 3.1.6 and choosing $M = m^{\omega_1}$ and $\varepsilon = m^{-\omega_2}$.

When $P(Q \leq \eta) \leq K\eta^{\zeta_1}(-\log \eta)^{-\zeta_2}$, the result follows by Lemma 3.1.6 and choosing $M = m^{\omega_1}$, $\varepsilon = m^{-\omega_2}$, and $\eta = m^{-\omega_3}$. □

CHAPTER 4

Large Deviations for Interacting Diffusions with Common Small Noise

In this chapter we study large deviation properties of interacting particle systems that are described through a certain collection of stochastic differential equations. Our main interest is in diffusions interacting through the empirical measure of the particle system with both individual and common sources of noises, given by a system of equations of the following form:

$$\begin{aligned} dX_i^n(t) &= b(X_i^n(t), \mu^n(t))\, dt + \sigma(X_i^n(t), \mu^n(t))\, dW_i(t) + \kappa(n)\alpha(X_i^n(t), \mu^n(t))\, dB(t), \\ X_i^n(0) &= x_i^n, \quad \mu^n(t) = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^n(t)}, \quad 1 \le i \le n, \quad t \in [0,T], \end{aligned} \tag{4.0.1}$$

where $\{W_i, i \in \mathbb{N}\}$ are independent m-dimensional Brownian motions, B is a k-dimensional Brownian motion, independent of $\{W_i, i \in \mathbb{N}\}$, $b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^{d\times m}$, and $\alpha : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^{d\times k}$ are appropriate maps, and $\{x_i^n\}_{1\le i\le n} \subset \mathbb{R}^d$ (see Section 4.1.1 for precise conditions on the coefficients and the initial conditions).

We will also study large deviation asymptotics for a second class of models that are given as particle approximations for Feynman-Kac functionals of the form

$$E\left[e^{\int_0^T c(X_1(s))ds} g(X_1(T))\right], \tag{4.0.2}$$

where g and c are bounded and continuous functions and X_1 is given by (1.2.12) (with $i = 1$). Denote by $\mathcal{M}_+(\mathbb{R}^d)$ the space of finite measures on $\mathbb{R}^d$ equipped with the topology of weak convergence, and consider the $\mathcal{C}([0,T] : \mathcal{M}_+(\mathbb{R}^d))$-valued random variables ν^n defined as

$$\nu^n(t) = \frac{1}{n}\sum_{i=1}^n e^{\int_0^t c(X_i^n(s))ds + \kappa(n)\int_0^t \beta(X_i^n(s))dB(s)} \delta_{X_i^n(t)}, \quad t \in [0,T], \tag{4.0.3}$$

where $\{X_i^n\}$ are given by (1.2.14) and β is a bounded and continuous function. Then, as $n \to \infty$, $\langle g, \nu^n(T)\rangle \doteq \int g(x)\, \nu^n(T)(dx)$ converges to the Feynman-Kac functional in (4.0.2) for all choices of sequences $\kappa(n) \to 0$. As a special case of Theorems 4.1.2 and 4.1.4 we obtain large deviation principles for ν^n, for different choices of $\kappa(n)$.

The law of large number behavior of such systems of particles outlined above is described by nonlinear equations of McKean-Vlasov type (cf. [82, 94]). The large deviation behavior of the associated empirical measure process is governed by two types of scaling, one corresponding to mean field asymptotics (as the number of particles $n \to \infty$) and the other to the Freidlin-Wentzell small noise asymptotics (as the noise intensity $\kappa(n) \to 0$).

In the setting where there is no common Brownian motion, i.e. $\kappa(n) = 0$, large deviation principles for the empirical measure have been studied in [33]. A different approach, based on certain variational representations for exponential functionals of finite dimensional Brownian motions [17] and weak convergence arguments, was taken in [22]. The latter paper, in contrast to [33], allowed for degenerate diffusion coefficients and for a mean field interaction in the diffusion coefficient. Large deviation properties of a system related to (4.0.1) were studied recently in [88], in which there is no common noise term but the independent Brownian motions $\{W_i\}$ are made to be small and vanish in the limit. In the systems with common noise that are considered in the current work, one needs to analyze the interplay between the contributions of two distinct sources of noise to non-typical behavior of the empirical measures. In the rate function (see (4.1.3)), this interplay is manifested through certain stochastic control problems in which there are two types of controls that play somewhat different roles in the dynamics. As already noted below (1.2.16) in a simpler setting, the control that arises from the individual noises is random and nonanticipative whereas the control from the common Brownian motion is nonrandom. In game theoretic terminology, the first control arises from the aggregated actions of the n individual players whereas the second control corresponds to the action of a single major agent that impacts the dynamics of all n players.

Our results give a complete characterization of the asymptotic behavior for different choices of $\kappa(n)$. Specifically, taking $\kappa(n) = n^{-1/2}$, Theorems 4.1.1 and 4.1.3 show that rates of decay of $P(\mu^n \in A)$ for non-typical events A are of the form $e^{-nI(A)}$, where the exponent $I(A)$ is described through a stochastic control problem with controls for both the aggregated player and the major agent. However, when $\kappa(n)n^{1/2} \to 0$, the contribution of the common Brownian motion to deviations in the empirical measure becomes negligible and the rate function only involves the aggregated player control. Finally, when $\kappa(n)n^{1/2} \to \infty$, the decay rates of $P(\mu^n \in A)$ are slower , given as $e^{-\kappa(n)^{-2}I(A)}$, and this time the dominating contribution to deviations to the empirical measure are due to the common Brownian motion and the corresponding stochastic control problem is described in terms of nonlinear Markov processes with deterministic controls.

In order to study rates of convergence of Feynman-Kac functionals analogous to those in (4.0.2), we consider the following system of coupled equations:

$$
\begin{aligned}
dX_i^n(t) &= b(X_i^n(t),\mu^n(t))\,dt + \sigma(X_i^n(t),\mu^n(t))\,dW_i(t) + \kappa(n)\alpha(X_i^n(t),\mu^n(t))\,dB(t),\\
dA_i^n(t) &= A_i^n(t)c(X_i^n(t),\mu^n(t))\,dt + A_i^n(t)\gamma^T(X_i^n(t),\mu^n(t))\,dW_i(t) + \kappa(n)A_i^n(t)\beta^T(X_i^n(t),\mu^n(t))\,dB(t),\\
X_i^n(0) &= x_i^n, \quad A_i^n(0) = a_i^n, \quad \mu^n(t) = \frac{1}{n}\sum_{i=1}^n \theta(A_i^n(t))\delta_{X_i^n(t)}, \quad 1 \le i \le n, \quad t \in [0,T],
\end{aligned}
\tag{4.0.4}
$$

where $c : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$, $\gamma : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^m$, and $\beta : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^k$ are suitable maps and $\{(x_i^n, a_i^n)\}_{1\le i\le n} \subset \mathbb{R}^d \times \mathbb{R}_+$ (see Section 4.1.2 for precise conditions). Note that in the special case where $\theta(x) = x$, $\gamma(x,\mu) = 0$, and the coefficients do not depend on the empirical measure (i.e. $b(x,\mu) = b(x)$, and similarly for σ, α, c, β), μ^n reduces to (4.0.3) (with c replaced by $c - \kappa(n)^2\beta^T\beta/2$). In the general case the finite weighted empirical measures $\mu^n(t)$ take the form

$$
\mu^n(t) = \frac{1}{n}\sum_{i=1}^n \theta\left(e^{\int_0^t c_n(X_i^n(s),\mu^n(s))ds + \int_0^t \gamma^T(X_i^n(s),\mu^n(s))\,dW_i(s) + \kappa(n)\int_0^t \beta^T(X_i^n(s),\mu^n(s))\,dB(s)}\right)\delta_{X_i^n(t)}, \tag{4.0.5}
$$

where $c_n = c - \gamma^T\gamma/2 - \kappa(n)^2\beta^T\beta/2$, which covers a broad family of interacting particle models for Feynman-Kac distribution flows (cf. [34]). Our main result is Theorem 4.1.2, which gives a large deviation principle for $\{\mu^n\}$ in $\mathcal{C}([0,T] : \mathcal{M}^+(\mathbb{R}^d))$ under appropriate conditions on the coefficients and the initial conditions.

The LDP results herein have a somewhat similar flavor to those for two-scale stochastic systems, see for example the recent works [58, 93] which analyze the large deviations behavior of reaction-diffusion equations with slow and fast time scales in a particular limiting regime of the parameters, as well as [38] which considers multiple regimes in a finite dimensional problem. As in the problems studied here, in two-scale systems as well there are two natural parameters of interest, one (denoted as δ) representing the speed of the fast system, and the other (denoted as ε) representing the magnitude of the noise in the slow system. Depending on the manner in which δ and ε approach 0 in relation to each other, one expects different forms of large deviation behavior. Specifically the papers [58, 93] considered the regime $\delta/\sqrt{\varepsilon} \to 0$ while the other regimes, namely $\delta/\sqrt{\varepsilon} \to c \in (0,\infty)$ and $\delta/\sqrt{\varepsilon} \to \infty$ were left open and are expected to be more challenging. Although there are formal similarities with the problem studied here, it is not immediately clear whether the methods developed in the current paper can be directly used to study the harder regimes that were left unaddressed in [58, 93].

We now make some comments on proof techniques. For an LDP for μ^n associated with the system in (4.0.1), the goal is to characterize the asymptotics of Laplace functionals of the form on the left side of (4.2.4). Since μ^n is a functional of individual Brownian motions W_i and the common Brownian motion B, using the variational formula for exponential functionals of finite dimensional Brownian motions [17], one can give a stochastic control representation for the Laplace functional of interest (see Theorem 4.2.1) that involves two types of controls. The first type, denoted as u_i^n, captures the deviations from the individual Brownian motions W_i (one control for each i) and the other type, denoted as v^n, is associated with the common Brownian motion B. The two types of controls are scaled differently in the representation, and the analysis of this scaling, which depends on $\kappa(n)$, is key to understanding the different types of large deviation behavior for various choices of $\kappa(n)$. In proving the large deviation upper bound one needs to argue the convergence of the cost on the right side of (4.2.4) associated with near optimal choices of control sequences and to characterize the limits. For this, following [22], we consider certain augmented empirical measures Q^n that include, in addition to particle states, the associated controls and the driving individual noises. The convergence of the costs (along subsequences) is shown by establishing the tightness of the collection (Q^n, v^n). Tightness properties depend crucially on the rate at which $\kappa(n) \to 0$, and the forms of the limit points under different conditions on $\kappa(n)$ reveal the different types of large deviation behavior. Next step is to characterize the form of the limit cost. This is done by establishing that the limit points of Q^n solve certain nonlinear controlled martingale problems. The controls arise from two sources, one is from the limits of v^n (this is the control associated with the common noise); and the other is from the second marginal of Q^n. This characterization leads to the forms of rate functions described previously. In order to prove the lower bound one needs to construct a suitable collection of controls for which the associated costs converge to certain near optimal costs for the limiting stochastic control problems. This time tightness is not enough as one needs to prove convergence of (augmented) empirical measures to a specific limiting measure. The key step in the proof of the lower bound is establishing uniqueness of weak solutions of stochastic differential equations associated with certain controlled nonlinear Markov processes. Such results are given in Lemmas 4.2.4 and 4.3.4. With such a uniqueness result one can then construct the desired sequence of controls and controlled processes on certain infinite product path spaces such that the associated state processes and costs converge in an appropriate manner.

Proofs for the large deviation asymptotics of Feynman-Kac measures as in (4.0.4) rely on analyzing the properties of θ. One may attempt to deduce this result as a corollary of large deviation results for (4.0.1) by first establishing an LDP for the empirical measure of $(X_i^n(\cdot), A_i^n(\cdot))$. However, with this approach, the conditions needed appear to be too restrictive (see Remark 4.1.2(a)). We will instead analyze the weighted empirical measure μ^n in (4.0.5) directly via variational representations for Laplace functionals associated with μ^n. We prove the result under two different types of conditions. The first set of conditions requires in particular that $\gamma = 0$ and θ is a Lipschitz function (e.g. $\theta(x) = x$). When $\theta(x) = x$, and $\gamma = 0$ is violated, a large deviation principle is not available even in the most elementary settings (see Remark 4.1.2(c)). The second set of conditions allows γ to be more general but imposes logarithmic growth conditions on θ.

The paper is organized as follows. Section 4.1 introduces the models, gives our precise assumptions, and presents the main results. In particular, Section 4.1.1 considers the empirical measure problem while Section 4.1.2 presents results for interacting particle models for Feynman-Kac functionals. The first two sections consider the case where the common noise intensity $\kappa(n)$ is of order $n^{-1/2}$, and in section 4.1.3 we present results for other choices of $\kappa(n)$ (i.e. of larger or smaller order than $n^{-1/2}$). Sections 4.2 through 4.4 contain the proofs of our main results. The two appendices contain proofs of some auxiliary results.

The following are some notational conventions followed throughout this chapter. We will denote by $\mathcal{C}([0,T]:\mathbb{R}^d)$ the space of continuous functions from $[0,T]$ to $\mathbb{R}^d$, equipped with the sup-norm topology corresponding to the distance

$$d(\psi_1,\psi_2) = \sup_{0\le t\le T} \|\psi_1(t)-\psi_2(t)\| \quad \text{for} \quad \psi_1,\psi_2 \in \mathcal{C}([0,T]:\mathbb{R}^d).$$

For a Polish space S, $\mathcal{C}(S)$ will denote the space of continuous functions from S into $\mathbb{R}$, and $\mathcal{C}_b(S)$ will denote the space of continuous and bounded functions from S into $\mathbb{R}$. We denote by $L^2([0,T],\mathbb{R}^k)$ the space of functions from $[0,T]$ into $\mathbb{R}^k$ that are square integrable with respect to Lebesgue measure. Let $\mathcal{P}(S)$ denote the space of all probability measures on S equipped with the usual weak convergence topology. If S is a product space of the form $S_1\times\cdots\times S_k$, then for $\Theta\in\mathcal{P}(S)$ and $i=1,\ldots,k$, we denote by $[\Theta]_i$ the ith marginal of Θ, which is a probability measure on S_i. Notations $[\Theta]_{(i_1,\ldots,i_r)}$, for $1\le r\le k$ and $1\le i_1<i_2<\cdots<i_r\le k$, will be interpreted in a similar manner. Let $\mathcal{M}_+(S)$ denote the space of finite positive measures on S, also with the topology of weak convergence. In particular, for $\gamma_n,\gamma\in\mathcal{M}_+(S)$, $\gamma_n\to\gamma$ under this topology if and only if for every $f\in\mathcal{C}_b(S)$, $\int f\,d\gamma_n\to\int f\,d\gamma$. For $\gamma\in\mathcal{M}_+(S)$ and a γ-integrable function $f:S\to\mathbb{R}$, we will denote $\int_S f(x)\,\gamma(dx)$ as $\langle f,\gamma\rangle$. $\mathcal{C}^k(\mathbb{R}^d)$ [resp. $\mathcal{C}^k_c(\mathbb{R}^d)$] will denote the space of functions [resp. functions with compact support] from $\mathbb{R}^d$ to $\mathbb{R}$ that are continuously differentiable up to order k. For a bounded map $f:S\to\mathbb{R}$, we denote $\sup_{x\in S}|f(x)|$ as $\|f\|_\infty$.

4.1 Main Results

In this section we introduce the models of interest, state our precise assumptions, and present the main results.

4.1.1 Diffusions Interacting Through the Empirical Distribution

Consider a filtered probability space $(\Omega,\mathcal{F},P,\{\mathcal{F}_t\})$ where the filtration satisfies the usual conditions. Let $\{W_i\}_{i=1}^\infty$ be an iid collection of m-dimensional Brownian motions on this space. Also, let B be a k-dimensional Brownian motion that is independent of the collection $\{W_i\}_{i=1}^\infty$. We assume that, for every s, $\{W_i(t)-W_i(s), B(t)-B(s), i\ge 1, t\ge s\}$ is independent of $\mathcal{F}_s$, so that W_i and B are $\{\mathcal{F}_t\}$-martingales.

Consider, for $n\in\mathbb{N}$, a collection of stochastic processes $\{X_i^n\}_{i=1}^n$ with sample paths in $\mathcal{C}([0,T]:\mathbb{R}^d)$ given by the system of equations in (4.0.1) where $\kappa:\mathbb{N}\to\mathbb{R}_+$ satisfies $\kappa(n)\to 0$ as $n\to\infty$, and b,σ, and α are suitable coefficients.

We will make the following assumption on the initial conditions.

Condition 4.1.1. *There exists $\xi_0 \in \mathcal{P}(\mathbb{R}^d)$ such that for all ξ_0-integrable $f : \mathbb{R}^d \to \mathbb{R}$,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(x_i^n) = \langle f, \xi_0 \rangle.$$

Furthermore, $\sup_{n\geq 1} \frac{1}{n}\sum_{i=1}^{n} \|x_i^n\|^2 < \infty$.

We will require the coefficients b, α, and σ to be Lipschitz continuous. In order to state this condition precisely, we recall the bounded-Lipschitz metric on the space of measures. Recall that $\mathcal{M}_+(\mathbb{R}^d)$ denotes the space of positive measures on $\mathbb{R}^d$ equipped with the weak topology. This topology can be metrized by the bounded Lipschitz metric

$$d_{BL}(\nu_1, \nu_2) = \sup_{f \in BL(\mathbb{R}^d)} \left|\langle f, \nu_1\rangle - \langle f, \nu_2\rangle\right|, \quad \nu_i \in \mathcal{M}_+(\mathbb{R}^d), \quad i = 1, 2,$$

where

$$BL(\mathbb{R}^d) = \left\{ f \in \mathcal{C}(\mathbb{R}^d) : \|f\|_\infty \leq 1 \text{ and } f \text{ is Lipschitz with Lipschitz constant bounded by } 1 \right\}.$$

The following is the main condition on the coefficients.

Condition 4.1.2. *The map b is Lipschitz and the maps σ, α are bounded and Lipschitz from $\mathbb{R}^d \times \mathcal{M}_+(\mathbb{R}^d)$ to $\mathbb{R}^d$, $\mathbb{R}^{d\times m}$, and $\mathbb{R}^{d\times k}$ respectively. Namely, there is a $K \in (0, \infty)$ such that for each $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^d)$,*

(a) $\|\sigma(x,\mu)\|^2 + \|\alpha(x,\mu)\|^2 \leq K^2$, *and*

(b) $\|b(x,\mu) - b(y,\nu)\| + \|\sigma(x,\mu) - \sigma(y,\nu)\| + \|\alpha(x,\mu) - \alpha(y,\nu)\| \leq K\left(\|x-y\| + d_{BL}(\mu,\nu)\right)$.

For Theorem 4.1.1 we can replace $\mathcal{M}_+(\mathbb{R}^d)$ with $\mathcal{P}(\mathbb{R}^d)$ in the above condition, however it is convenient to formulate the condition as above in order to have a common set of conditions for Theorems 4.1.1 and 4.1.2. For the LDP we will assume in addition that the diffusion coefficient σ depends on the state of the system only through the empirical measure:

Condition 4.1.3. *For $x \in \mathbb{R}^d$ and $\mu \in \mathcal{M}_+(\mathbb{R}^d)$, $\sigma(x,\mu) = \sigma(\mu)$.*

Under Condition 4.1.2 it follows by standard arguments that for each n there is a unique pathwise solution of (4.0.1). Abusing notation, let μ^n be a random variable with values in $\mathcal{P}(\mathcal{C}([0,T] : \mathbb{R}^d))$ defined as $\mu^n \doteq \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^n}$. Note that $\mu^n(s)$ is the (random) marginal distribution at time instant s associated with μ^n. We will occasionally denote the map $t \mapsto \mu^n(t)$, as $\mu^n(\cdot)$ which is viewed as a $\mathcal{P}(\mathbb{R}^d)$-valued stochastic process with continuous sample paths or, equivalently, a random variable with values in $\mathcal{C}([0,T] : \mathcal{P}(\mathbb{R}^d))$.

Our first main result gives a large deviation principle for μ^n in $\mathcal{P}(\mathcal{C}([0,T]:\mathbb{R}^d))$. We begin by introducing the associated rate function. This function will be described in terms of solutions to certain controlled McKean-Vlasov equations which we now introduce. Recall the Polish spaces $\mathcal{R}$ and $\mathcal{R}_1$ of relaxed controls from the Introduction.

Given $\varphi \in L^2([0,T]:\mathbb{R}^k)$ and a continuous map $\nu : [0,T] \to \mathcal{P}(\mathbb{R}^d)$, consider the controlled nonlinear SDE $\mathcal{S}_1[\varphi,\nu]$, on some filtered probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}, \{\bar{\mathcal{F}}_t\})$, equipped with an m-dimensional $\bar{\mathcal{F}}_t$-Brownian motion W:

$$\mathcal{S}_1[\varphi,\nu] \doteq \begin{cases} d\bar{X}(t) = b(\bar{X}(t),\nu(t))\,dt + \left(\int_{\mathbb{R}^m} \sigma(\bar{X}(t),\nu(t))y\,\rho_t(dy)\right)\,dt + \sigma(\bar{X}(t),\nu(t))\,dW(t) \\ \qquad\qquad + \alpha(\bar{X}(t),\nu(t))\varphi(t)\,dt, \\ \bar{X}(t) \sim \nu(t), \quad t \in [0,T], \quad \nu(0) = \xi_0, \end{cases} \tag{4.1.1}$$

where $\xi_0 \in \mathcal{P}(\mathbb{R}^d)$ is as in Condition 4.1.1. In the above equation ρ is an $\mathcal{R}_1$-valued random variable such that $\rho([0,t]\times A)$ is $\bar{\mathcal{F}}_t$-measurable for every $A \in \mathcal{B}(\mathbb{R}^m)$ and $t \in [0,T]$, and $\bar{X}$ is an $\bar{\mathcal{F}}_t$-adapted stochastic process with sample paths in $\mathcal{C}([0,T]:\mathbb{R}^d)$. The notation $\bar{X}(t) \sim \nu(t)$ signifies that $\bar{X}(t)$ has probability distribution $\nu(t)$, i.e. $\bar{P} \circ \bar{X}(t)^{-1} = \nu(t)$. We note that $\mathcal{S}_1[\varphi,\nu]$ is driven by two types of controls, the control φ is a deterministic function whereas ρ represents a random control in the dynamics.

A triple $(\bar{X},\rho,W)$ that solves $\mathcal{S}_1[\varphi,\nu]$ for a given φ and ν can be viewed as a $\mathcal{Z}_1$-valued random variable, where

$$\mathcal{Z}_1 \doteq \mathcal{X} \times \mathcal{R}_1 \times \mathcal{W},\ \mathcal{X} \doteq \mathcal{C}([0,T]:\mathbb{R}^d),\ \text{and}\ \mathcal{W} \doteq \mathcal{C}([0,T]:\mathbb{R}^m).$$

The distribution of $(\bar{X},\rho,W)$ on $\mathcal{Z}_1$ is an element of $\mathcal{P}(\mathcal{Z}_1)$ and is called a *weak solution* of the controlled SDE $\mathcal{S}_1[\varphi,\nu]$. Define

$$\mathcal{P}_2(\mathcal{Z}_1) \doteq \left\{\Theta \in \mathcal{P}(\mathcal{Z}_1) : E_\Theta\left[\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\,\rho(dy\,dt)\right] < \infty\right\},$$

where in the above display E_Θ denotes expectation on $(\mathcal{Z}_1, \mathcal{B}(\mathcal{Z}_1), \Theta)$ and, abusing notation, ρ is the second coordinate map on $(\mathcal{Z}_1, \mathcal{B}(\mathcal{Z}_1))$, i.e.

$$\rho(x,r,w) \doteq r, \quad (x,r,w) \in \mathcal{Z}_1.$$

Note that, the above expectation can be written as

$$E_\Theta\left[\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\,\rho(dy\,dt)\right] = \int_{\mathcal{R}_1}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\,r(dy\,dt)\,[\Theta]_2(dr).$$

For $\Theta \in \mathcal{P}(\mathcal{Z}_1)$, let $\nu_\Theta : [0,T] \to \mathcal{P}(\mathbb{R}^d)$ be defined as

$$\nu_\Theta(t)(B) \doteq \Theta\left\{(x,r,w) \in \mathcal{Z}_1 : x(t) \in B\right\}, \quad B \in \mathcal{B}(\mathbb{R}^d).$$

Note that if Θ is a weak solution of $\mathcal{S}_1[\varphi,\nu]$, then $\nu(t) = \nu_\Theta(t)$ for all $t \in [0,T]$. For a given $\varphi \in L^2([0,T]:\mathbb{R}^k)$, let $\mathcal{E}_1[\varphi]$ denote the subset of $\mathcal{P}_2(\mathcal{Z}_1)$ given as

$$\mathcal{E}_1[\varphi] \doteq \left\{\Theta \in \mathcal{P}_2(\mathcal{Z}_1) : \Theta \text{ is a weak solution to } \mathcal{S}_1[\varphi,\nu_\Theta]\right\}. \tag{4.1.2}$$

Then the candidate rate function for the LDP for μ^n is

$$I_1(\nu) \doteq \inf_{\varphi \in L^2([0,T]:\mathbb{R}^k)} \left\{ \inf_{\Theta \in \mathcal{E}_1[\varphi]:[\Theta]_1=\nu} E_\Theta\left[\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2 \rho(dy\,dt)\right] + \frac{1}{2\lambda^2}\int_0^T \|\varphi(t)\|^2\,dt \right\}, \tag{4.1.3}$$

for $\nu \in \mathcal{P}(\mathcal{X})$, where $\lambda \in (0,\infty)$ is introduced below.

The following is the first main result of this work. It gives an LDP in the case $\kappa(n)$ is of the order $n^{-1/2}$. Later in Section 4.1.3 we will consider the large deviation behavior when $\kappa(n)$ is of smaller or higher order than $n^{-1/2}$. Part 1 below gives a law of large numbers result while part 2 establishes a large deviation principle. Denote the element $\delta_{\{0\}}(dy)\,dt$ of $\mathcal{R}$ as r^o.

Theorem 4.1.1. *Suppose that Conditions 4.1.1, 4.1.2 hold and that $\kappa(n) \to 0$ as $n \to \infty$.*

(i) There is a $\mu^* \in \mathcal{P}(\mathcal{X})$ such that $\mu^n \to \mu^*$ in probability. Furthermore, μ^* can be characterized as the first marginal $[\Theta]_1$ of Θ, where Θ is the unique element in $\mathcal{P}(\mathcal{Z}_1)$ that is a weak solution of $\mathcal{S}_1[0,\nu_\Theta]$ and satisfies $[\Theta]_2 = \delta_{r^o}$.

(ii) Suppose in addition that Condition 4.1.3 is satisfied and that $\sqrt{n}\kappa(n) \to \lambda \in (0,\infty)$. Then $\{\mu^n\}_{n\in\mathbb{N}}$ satisfies a large deviation principle on $\mathcal{P}(\mathcal{X})$ with speed n and rate function I_1.

Proof of Theorem 4.1.1 will be given in Section 4.2.

Remark 4.1.1. *Since the map $\nu \mapsto \{t \mapsto \nu(t)\}$ from $\mathcal{P}(\mathcal{X})$ to $\mathcal{C}([0,T]:\mathcal{P}(\mathbb{R}^d))$ is a continuous map, we have by the contraction principle that $\mu^n(\cdot)$ regarded as a sequence of random variables with values in $\mathcal{C}([0,T]:\mathcal{P}(\mathbb{R}^d))$ satisfies an LDP as well.*

4.1.2 Interacting Particle Systems for Feynman-Kac Functionals

In this section we consider a setting where the interaction term is given in terms of a weighted empirical measure of the states of the particles and where the weights are governed by another system of stochastic equations. Let $(\Omega,\mathcal{F},P,\{\mathcal{F}_t\})$,

$\{W_i\}$, B be as in Section 4.1.1. Consider for $n \in \mathbb{N}$, a collection of stochastic processes $\{(X_i^n, A_i^n)\}_{i=1}^n$ with sample paths in $\mathcal{C}([0,T] : \mathbb{R}^d \times \mathbb{R}_+)$ given by the system of equations in (4.0.4). Here $\theta : \mathbb{R}_+ \to \mathbb{R}_+$, $\kappa : \mathbb{N} \to \mathbb{R}_+$, and $b, \sigma, \alpha, c, \gamma$, and β are suitable maps. Note that $\mu^n(t)$ in this set of equations can also be represented as on the right side of (4.0.5). In addition to Condition 4.1.2 on the coefficients, we will assume the following condition.

Condition 4.1.4. *The maps c, γ, β are bounded and Lipschitz from $\mathbb{R}^d \times \mathcal{M}_+(\mathbb{R}^d)$ to $\mathbb{R}^d$, $\mathbb{R}^m$, and $\mathbb{R}^k$ respectively. Namely, there is a $K \in (0,\infty)$ such that for each $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^d)$,*

(a) $\|c(x,\mu)\|^2 + \|\gamma(x,\mu)\|^2 + \|\beta(x,\mu)\|^2 \le K^2$, *and*

(b) $\|c(x,\mu) - c(y,\nu)\| + \|\gamma(x,\mu) - \gamma(y,\nu)\| + \|\beta(x,\mu) - \beta(y,\nu)\| \le K\left(\|x-y\| + d_{BL}(\mu,\nu)\right)$.

The weights in the random measure $\mu^n(t)$ are determined through the map θ on which we make the following assumption.

Condition 4.1.5. *Either one of the following hold:*

(a) $\theta \in \mathcal{C}^2(\mathbb{R}_+)$ *and*

$$\sup_{x\in\mathbb{R}_+} |\theta'(x)x| + \sup_{x\in\mathbb{R}_+} |\theta''(x)x^2| < \infty. \tag{4.1.4}$$

(b) There is a $L \in (0,\infty)$ such that $|\theta(x) - \theta(y)| \le L|x-y|$ for all $x, y \in \mathbb{R}_+$.

Condition 4.1.5(b) simply says that θ is a Lipschitz function. It is easily checked that under Condition 4.1.5(a), θ is Lipschitz as well. The latter condition, in addition, implies an (at most) logarithmic growth on θ.

Under Conditions 4.1.2, 4.1.4, and 4.1.5, there is a unique pathwise solution to the system of equations in (4.0.4). Although the proof is standard, we provide a sketch in Appendix 4.6. The object of interest is the stochastic process $\{\mu^n(t)\}_{t\in[0,T]}$ which is regarded as a random variable with values in $\mathcal{C}([0,T] : \mathcal{M}_+(\mathbb{R}^d))$. Our second main result gives a large deviation principle for $\mu^n(\cdot)$ in this path space. We introduce two additional conditions that will be needed for this result. For the initial values $\{(a_i^n, x_i^n)\}$ in (4.0.4) we will assume the following in addition to Condition 4.1.1:

Condition 4.1.6. *There exists $\eta_0 \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}_+)$ such that for all η_0-integrable $g : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n g(x_i^n, a_i^n) = \langle g, \eta_0 \rangle.$$

Furthermore,

$$\sup_{n\ge1} \frac{1}{n} \sum_{i=1}^n (a_i^n)^2 < \infty \quad \textit{and} \quad \sup_{n\ge1} \frac{1}{n} \sum_{i=1}^n (\log a_i^n)^- < \infty.$$

Note that when both Conditions 4.1.1 and 4.1.6 hold, we have $[\eta_0]_1 = \xi_0$, where $[\eta_0]_1$ is the marginal distribution of η_0 on $\mathbb{R}^d$.

Finally, for the large deviations result, in addition to Condition 4.1.3, we will assume that the diffusion coefficient γ depends on the state of the system only through the empirical measure, namely:

Condition 4.1.7. *For $x \in \mathbb{R}^d$ and $\mu \in \mathcal{M}_+(\mathbb{R}^d)$, $\gamma(x,\mu) = \gamma(\mu)$.*

We now present the rate function that will govern the LDP for $\{\mu^n(\cdot)\}$. Given $\varphi \in L^2([0,T] : \mathbb{R}^k)$ and $\nu \in \mathcal{C}([0,T] : \mathcal{M}_+(\mathbb{R}^d))$ as in Section 4.1.1, consider the controlled nonlinear SDE $\mathcal{S}_2[\varphi,\nu]$ given on some filtered probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}, \{\bar{\mathcal{F}}_t\})$, equipped with an m-dimensional $\bar{\mathcal{F}}_t$-Brownian motion W:

$$\mathcal{S}_2[\varphi,\nu] \doteq \begin{cases} d\bar{X}(t) = b(\bar{X}(t),\nu(t))\,dt + \left(\int_{\mathbb{R}^m} \sigma(\bar{X}(t),\nu(t)) y\, \rho_t(dy)\right) dt + \sigma(\bar{X}(t),\nu(t))\,dW(t) \\ \qquad\qquad + \alpha(\bar{X}(t),\nu(t))\varphi(t)\,dt, \\ d\bar{A}(t) = \bar{A}(t) c(\bar{X}(t),\nu(t))\,dt + \left(\int_{\mathbb{R}^m} \bar{A}(t)\gamma^T(\bar{X}(t),\nu(t)) y\, \rho_t(dy)\right) dt \\ \qquad\qquad + \bar{A}(t)\gamma^T(\bar{X}(t),\nu(t))\,dW(t) + \bar{A}(t)\beta^T(\bar{X}(t),\nu(t))\varphi(t)\,dt, \\ \langle f, \nu(t)\rangle = \bar{E}[\theta(\bar{A}(t)) f(\bar{X}(t))] \text{ for every } f \in \mathcal{C}_b(\mathbb{R}^d), \quad t \in [0,T], \quad (\bar{X}(0),\bar{A}(0)) \sim \eta_0, \end{cases} \tag{4.1.5}$$

where $\bar{E}$ denotes expectation with respect to $\bar{P}$. Here ρ is as in Section 4.1.1, and $\bar{X}$ and $\bar{A}$ are $\bar{\mathcal{F}}_t$-adapted stochastic processes with sample paths in $\mathcal{C}([0,T] : \mathbb{R}^d)$ and $\mathcal{C}([0,T] : \mathbb{R}_+)$, respectively, such that

$$\bar{E}\left[\sup_{0 \le t \le T} \theta(\bar{A}(t))\right] < \infty.$$

A quadruple $(\bar{X}, \bar{A}, \rho, W)$ that solves $\mathcal{S}_2[\varphi,\nu]$ is a $\mathcal{Z}_2$-valued random variable, where

$$\mathcal{Z}_2 \doteq \mathcal{X} \times \mathcal{Y} \times \mathcal{R}_1 \times \mathcal{W}, \quad \mathcal{Y} \doteq \mathcal{C}([0,T] : \mathbb{R}_+),$$

and $\mathcal{X}, \mathcal{W}, \mathcal{R}_1$ are as before. The distribution of $(\bar{X}, \bar{A}, \rho, W)$ on $\mathcal{Z}_2$ is an element of $\mathcal{P}(\mathcal{Z}_2)$ and is called a weak solution of $\mathcal{S}_2[\varphi,\nu]$. Let

$$\mathcal{P}_2(\mathcal{Z}_2) \doteq \left\{\Theta \in \mathcal{P}(\mathcal{Z}_2) : E_\Theta\left[\int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \rho(dy\,dt)\right] < \infty,\ E_\Theta\left[\sup_{0 \le t \le T} \theta(\bar{A}(t))\right] < \infty\right\}.$$

Note that if $\Theta \in \mathcal{P}_2(\mathcal{Z}_2)$ then $\nu_\Theta \in \mathcal{C}([0,T] : \mathcal{M}_+(\mathbb{R}^d))$, where ν_Θ is defined as

$$\langle f, \nu_\Theta(t)\rangle \doteq E_\Theta\left[\theta(\bar{A}(t)) f(\bar{X}(t))\right] \quad \text{for} \quad f \in \mathcal{C}_b(\mathbb{R}^d), \quad t \in [0,T], \tag{4.1.6}$$

and if such a Θ is a weak solution of $\mathcal{S}_2[\varphi,\nu]$, then, for every $t \in [0,T]$, $\nu(t) = \nu_\Theta(t)$. Given $\varphi \in L^2([0,T]:\mathbb{R}^d)$, let

$$\mathcal{E}_2[\varphi] \doteq \{\Theta \in \mathcal{P}_2(\mathcal{Z}_2) : \Theta \text{ is a weak solution to } \mathcal{S}_2[\varphi,\nu_\Theta]\}.$$

The candidate rate function is given as

$$I_2(\nu) \doteq \inf_{\varphi \in L^2([0,T]:\mathbb{R}^k)} \left\{ \inf_{\Theta \in \mathcal{E}_2[\varphi]:\nu_\Theta = \nu} E_\Theta \left[\frac{1}{2} \int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \, \rho(dy\,dt) \right] + \frac{1}{2\lambda^2} \int_0^T \|\varphi(t)\|^2 \, dt \right\}, \tag{4.1.7}$$

for $\nu \in \mathcal{C}([0,T]:\mathcal{M}_+(\mathbb{R}^d))$. The following is the second main result of this work. As in Section 4.1.1 here we only consider the case where $\kappa(n)$ is of order $n^{-1/2}$. Values of $\kappa(n)$ of higher or lower order than $n^{-1/2}$ will be considered in Section 4.1.3.

Once more, the first part of the theorem below gives a law of large numbers (LLN) and the second part establishes an LDP. The proof is given in Section 4.3.

Theorem 4.1.2. *Suppose that Conditions 4.1.1, 4.1.2, 4.1.4, 4.1.5, and 4.1.6 hold and that $\kappa(n) \to 0$ as $n \to \infty$.*

(i) There is a $\mu^* \in \mathcal{C}([0,T]:\mathcal{M}_+(\mathbb{R}^d))$ such that $\mu^n \to \mu^*$ in probability. Furthermore, μ^* can be characterized as the map $t \mapsto \nu_\Theta(t)$, where Θ is the unique element in $\mathcal{P}(\mathcal{Z}_2)$ that is a weak solution of $\mathcal{S}_2[0,\nu_\Theta]$ and satisfies $[\Theta]_3 = \delta_{r^o}$.

(ii) Suppose that σ and γ satisfy Conditions 4.1.3 and 4.1.7, and either (i) θ satisfies Condition 4.1.5(a), or (ii) θ satisfies Condition 4.1.5(b) and $\gamma \equiv 0$. Also suppose that $\sqrt{n}\kappa(n) \to \lambda \in (0,\infty)$. Then $\{\mu^n\}_{n\in\mathbb{N}}$ satisfies a large deviation principle on $\mathcal{C}([0,T]:\mathcal{M}_+(\mathbb{R}^d))$ with speed n and rate function I_2.

Remark 4.1.2. *(a) Consider the empirical measure of $\{X_i^n(s), A_i^n(s)\}$ on $\mathbb{R}^d \times \mathbb{R}_+$, given as*

$$\hat{\mu}^n(s) \doteq \frac{1}{n} \sum_{i=1}^n \delta_{\left(X_i^n(s), A_i^n(s)\right)}.$$

Then the system in equation (4.0.4) can be written in form of a system as in (4.0.1) in which X_i^n is replaced by the pair (X_i^n, A_i^n). With such a rewriting, one may attempt to deduce Theorem 4.1.2 as a corollary of Theorem 4.1.1. However, with this reformulation, the conditions needed for Theorem 4.1.1 are too restrictive. In particular, conditions assumed in the statement of Theorem 4.1.2 will, in general, not imply the conditions of Theorem 4.1.1 (with the new coefficients obtained through the reformulation). Specifically, requiring Conditions 4.1.2 and 4.1.3 for the reformulated system will say that θ is bounded and $\gamma \equiv 0$.

(b) A minor modification of the proof of Theorem 4.1.2 shows in fact that the joint empirical measure $\frac{1}{n}\sum_{i=1}^n \delta_{(X_i^n, A_i^n)}$ satisfies an LDP on $\mathcal{P}(\mathcal{C}([0,T]:\mathbb{R}^d \times \mathbb{R}_+))$. Note that since θ may be unbounded, the map $\Theta \mapsto \nu_\Theta$ is not

continuous (in fact in general not even well defined) on all of $\mathcal{P}(\mathcal{C}([0,T]:\mathbb{R}^d \times \mathbb{R}_+))$ and so one cannot deduce an LDP for μ^n from that of the joint empirical measure by a direct application of the contraction principle. In any case, the amount of work needed to establish the LDP for μ^n is about the same as that needed for the LDP for the joint empirical measure.

(c) In Theorem 4.1.2, for the case where θ satisfies Condition 4.1.5(b), we require that $\gamma \equiv 0$. The reason for this restrictive requirement on γ can be seen as follows. Consider the simplest example of a θ satisfying Condition 4.1.5(b), namely $\theta(x) = x$. Consider also the simplest form of a non-zero γ in (4.0.4)*, namely $\gamma(x,\mu) = \gamma \in \mathbb{R}^m \setminus \{0\}$. Also suppose that $\beta \equiv 0$, $c(x,\mu) \equiv c \in \mathbb{R}$, and that $a_i^n = 1$ for all i,n. Then the second set of equations in* (4.0.4) *reduces to*

$$dA_i^n(t) = cA_i^n(t)\,dt + \gamma A_i^n(t)\,dW_i(t), \quad A_i^n(0) = 1, \quad 1 \le i \le n.$$

Namely,

$$A_i^n(t) = \exp\left\{\left(c - \frac{\gamma^2}{2}\right)t + \gamma W_i(t)\right\}.$$

In this case, an LDP for $\mu^n(\cdot)$ will in particular say (by the contraction principle) that the sequence $\{\mu^n(1)(\mathbb{R}^d)\}$ satisfies an LDP. However the latter is just an LDP for the empirical mean of iid random variables, $\{A_i^n(1)\}$, namely $\frac{1}{n}\sum_{i=1}^n A_i^n(1)$, which is the subject of Cramér's theorem. However the key condition for this theorem, namely the finiteness of the moment generating function in a neighborhood of the origin, fails to hold in this case.

4.1.3 Intensity of the Common Noise

The LDP in Theorems 4.1.1 and 4.1.2 are established under the condition that the common noise intensity $\kappa(n)$ is $O(1/\sqrt{n})$. If this intensity approaches 0 at a different rate, the form of the rate function is expected to be different. In this section we discuss such results. We will consider two cases: Case I: $\sqrt{n}\kappa(n) \to 0$, and Case II: $\sqrt{n}\kappa(n) \to \infty$. Let $\mathcal{E}_1[\varphi]$ for a given $\varphi \in L^2([0,T]:\mathbb{R}^k)$ be as in Section 4.1.1. In order to define the rate function in the second case, we consider, for a φ as above and a $\nu \in \mathcal{C}([0,T]:\mathcal{P}(\mathbb{R}^d))$, the controlled nonlinear SDE $\tilde{\mathcal{S}}_1[\varphi,\nu]$, on some filtered probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}, \{\bar{\mathcal{F}}_t\})$, equipped with a m-dimensional $\bar{\mathcal{F}}_t$-Brownian motion W:

$$\tilde{\mathcal{S}}_1[\varphi,\nu] \doteq \begin{cases} d\bar{X}(t) = b(\bar{X}(t),\nu(t))\,dt + \sigma(\bar{X}(t),\nu(t))\,dW(t) + \alpha(\bar{X}(t),\nu(t))\varphi(t)\,dt, \\ \bar{X}(t) \sim \nu(t), \quad t \in [0,T], \quad \nu(0) = \xi_0. \end{cases} \tag{4.1.8}$$

The difference between the above equation and the equation in (4.1.1) is the absence of the control term ρ_t on the right side of (4.1.8). The distribution, on $\mathcal{X} \times \mathcal{W}$, of a pair $(\bar{X}, W)$ that solves (4.1.8) for a given φ and ν will be called a weak solution of $\tilde{\mathcal{S}}_1[\varphi,\nu]$.

For a $\varphi \in L^2([0,T] : \mathbb{R}^k)$, let

$$\tilde{\mathcal{E}}_1[\varphi] \doteq \left\{ \Theta \in \mathcal{P}(\mathcal{X} \times \mathcal{W}) : \Theta \text{ is a weak solution to } \tilde{\mathcal{S}}_1[\varphi, \nu_\Theta] \right\}. \tag{4.1.9}$$

For $\nu \in \mathcal{P}(\mathcal{C}([0,T] : \mathbb{R}^d))$, we denote the map $t \mapsto \nu(t)$, once more as ν. The following result gives an LDP when $\kappa(n)$ is different from $O(1/\sqrt{n})$. Recall that we assume $\kappa(n) \to 0$ as $n \to \infty$. Also recall the space $\mathcal{X} = \mathcal{C}([0,T] : \mathbb{R}^d)$.

Theorem 4.1.3. *Let $\{\mu^n\}_{n\in\mathbb{N}}$ be as in Section 4.1.1. Suppose that Conditions 4.1.1 and 4.1.2 hold.*

(i) Suppose in addition that Condition 4.1.3 is satisfied. If $\sqrt{n}\kappa(n) \to 0$ as $n \to \infty$, then $\{\mu^n\}$ satisfies an LDP on $\mathcal{P}(\mathcal{X})$ with speed n and rate function $\tilde{I}_{1,0}$ given as

$$\tilde{I}_{1,0}(\nu) \doteq \inf_{\Theta \in \mathcal{E}_1[0] : [\Theta]_1 = \nu} E_\Theta \left[\frac{1}{2} \int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \, \rho(dy\, dt) \right], \quad \nu \in \mathcal{P}(\mathcal{X}). \tag{4.1.10}$$

(ii) If $\sqrt{n}\kappa(n) \to \infty$ as $n \to \infty$, then $\{\mu^n\}$ satisfies an LDP on $\mathcal{P}(\mathcal{X})$ with speed $\kappa(n)^{-2}$ and rate function $\tilde{I}_{1,\infty}$ given as

$$\tilde{I}_{1,\infty}(\nu) \doteq \inf_{\varphi \in L^2([0,T]:\mathbb{R}^k)} \left\{ \inf_{\Theta \in \tilde{\mathcal{E}}_1[\varphi] : [\Theta]_1 = \nu} \frac{1}{2} \int_0^T \|\varphi(t)\|^2 \, dt \right\}, \quad \nu \in \mathcal{P}(\mathcal{X}). \tag{4.1.11}$$

The proof of Theorem 4.1.3 is very similar to that of Theorem 4.1.1 and therefore we will only provide a sketch and leave the details to the reader. This sketch is given in Section 4.4.

Remark 4.1.3. *Consider the special case discussed in the Introduction (see* (1.2.14)*) in which the interaction only comes through the common Brownian motion. For this special case the results in Theorems 4.1.1 and 4.1.3 (by some minor proof modifications) say the following. Suppose that the coefficients b, σ and α in* (1.2.14) *are Lipschitz. Also, suppose first that $\sqrt{n}\kappa(n) \to \lambda \in (0,\infty)$. Then $\{\mu^n\}$ as introduced in* (1.2.15) *satisfies an LDP in $\mathcal{P}(\mathcal{X})$ with speed n and rate function I defined in* (1.2.16)*. If $\sqrt{n}\kappa(n) \to 0$, then $\{\mu^n\}$ satisfies an LDP with speed n and rate function $\tilde{I}_{1,0}$ as in* (4.1.10) *and where $\mathcal{E}_1[\cdot]$ is as introduced below* (1.2.15)*. Finally, when $\sqrt{n}\kappa(n) \to \infty$, then $\{\mu^n\}$ satisfies an LDP with speed $\kappa(n)^{-2}$ and rate function $\tilde{I}_{1,\infty}$ given simply as*

$$\tilde{I}_{1,\infty}(\nu) = \inf_{\varphi} \left\{ \frac{1}{2} \int_0^T \|\varphi(t)\|^2 \, dt \right\},$$

where the infimum is taken over all $\varphi \in L^2([0,T] : \mathbb{R}^k)$ such that the solution $\{X\}$ of the controlled SDE

$$X(t) = x + \int_0^t b(X(s))\, ds + \int_0^t \sigma(X(s))\, dW(s) + \int_0^t \alpha(X(s))\varphi(s)\, ds, \quad t \in [0,T],$$

has probability law ν.

One can also give an analogue of Theorem 4.1.3 for Feynman-Kac weighted measures of the form in Section 4.1.2. We state such a result and leave proof details to the reader.

Consider, for a $\varphi \in L^2([0,T]:\mathbb{R}^k)$ and a $\nu \in \mathcal{C}([0,T]:\mathcal{M}_+(\mathbb{R}^d))$, the controlled nonlinear SDE $\tilde{\mathcal{S}}_2[\varphi,\nu]$, on some filtered probability space $(\bar{\Omega},\bar{\mathcal{F}},\bar{P},\{\bar{\mathcal{F}}_t\})$, equipped with a m-dimensional $\bar{\mathcal{F}}_t$-Brownian motion W:

$$\tilde{\mathcal{S}}_2[\varphi,\nu] \doteq \begin{cases} d\bar{X}(t) = b(\bar{X}(t),\nu(t))\,dt + \sigma(\bar{X}(t),\nu(t))\,dW(t) + \alpha(\bar{X}(t),\nu(t))\varphi(t)\,dt, \\ d\bar{A}(t) = \bar{A}(t)c(\bar{X}(t),\nu(t))\,dt + \bar{A}(t)\gamma^T(\bar{X}(t),\nu(t))\,dW(t) + \bar{A}(t)\beta^T(\bar{X}(t),\nu(t))\varphi(t)\,dt, \\ \langle f,\nu(t)\rangle = \bar{E}[\theta(\bar{A}(t))f(\bar{X}(t))] \text{ for every } f\in\mathcal{C}_b(\mathbb{R}^d), \quad t\in[0,T], \quad (\bar{X}(0),\bar{A}(0))\sim\eta_0, \end{cases} \tag{4.1.12}$$

where $\bar{E}$ denotes expectation with respect to $\bar{P}$. The distribution, on $\mathcal{X}\times\mathcal{Y}\times\mathcal{W}$, of $(\bar{X},\bar{A},W)$ that solves (4.1.12) for a given φ and ν will be called a weak solution of $\tilde{\mathcal{S}}_2[\varphi,\nu]$. For a $\varphi\in L^2([0,T]:\mathbb{R}^k)$, let

$$\tilde{\mathcal{E}}_2[\varphi] \doteq \left\{\Theta\in\mathcal{P}(\mathcal{X}\times\mathcal{Y}\times\mathcal{W}) : \Theta \text{ is a weak solution to } \tilde{\mathcal{S}}_2[\varphi,\nu_\Theta]\right\}.$$

Theorem 4.1.4. *Let $\{\mu^n\}_{n\in\mathbb{N}}$ be as in section 4.1.2. Suppose that Conditions 4.1.1, 4.1.2, 4.1.3, 4.1.4, 4.1.6, and 4.1.7 hold. Also suppose that, either (i) θ satisfies Condition 4.1.5(a), or (ii) θ satisfies Condition 4.1.5(b) and $\gamma\equiv 0$.*

(i) If $\sqrt{n}\kappa(n)\to 0$ as $n\to\infty$, then $\{\mu^n\}$ satisfies an LDP on $\mathcal{C}([0,T]:\mathcal{M}_+(\mathbb{R}^d))$ with speed n and rate function $\tilde{I}_{2,0}$ given as

$$\tilde{I}_{2,0}(\nu) \doteq \inf_{\Theta\in\mathcal{E}_2[0]:\nu_\Theta=\nu} E_\Theta\left[\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]}\|y\|^2\,\rho(dy\,dt)\right], \quad \nu\in\mathcal{C}([0,T]:\mathcal{M}_+(\mathbb{R}^d)). \tag{4.1.13}$$

(ii) If $\sqrt{n}\kappa(n)\to\infty$ as $n\to\infty$, then $\{\mu^n\}$ satisfies an LDP on $\mathcal{C}([0,T]:\mathcal{M}_+(\mathbb{R}^d))$ with speed $\kappa(n)^{-2}$ and rate function $\tilde{I}_{2,\infty}$ given as

$$\tilde{I}_{2,\infty}(\nu) \doteq \inf_{\varphi\in L^2([0,T]:\mathbb{R}^k)}\left\{\inf_{\Theta\in\tilde{\mathcal{E}}_2[\varphi]:\nu_\Theta=\nu}\frac{1}{2}\int_0^T\|\varphi(t)\|^2\,dt\right\}, \quad \nu\in\mathcal{C}([0,T]:\mathcal{M}_+(\mathbb{R}^d)). \tag{4.1.14}$$

4.2 Proof of Theorem 4.1.1.

Part 1 follows by a standard argument (cf. [95]), however for completeness we give a sketch in the Appendix. We now consider part 2.

From the well known equivalence between an LDP and a Laplace principle (cf. [37]) it suffices to show that the function I_1 introduced in (4.1.3) is a rate function and for every $F\in\mathcal{C}_b(\mathcal{P}(\mathcal{X}))$ the following upper and lower bounds are satisfied: the Laplace upper bound

$$\liminf_{n\to\infty} -\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] \geq \inf_{\nu\in\mathcal{P}(\mathcal{X})}\left[F(\nu)+I_1(\nu)\right], \tag{4.2.1}$$

and the Laplace lower bound

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] \leq \inf_{\nu\in\mathcal{P}(\mathcal{X})}\left[F(\nu)+I_1(\nu)\right]. \tag{4.2.2}$$

The upper bound is shown in Section 4.2.1 and the lower bound is treated in Section 4.2.2. The upper bound proof does not require Condition 4.1.3 and we present an argument assuming only Conditions 4.1.1 and 4.1.2. The proof of the statement that I_1 is a rate function is very similar to that of the upper bound and thus we only give a brief sketch which appears in Section 4.2.3. Proofs rely on a certain stochastic control representation for the Laplace functional on the left side of (4.2.1) and (4.2.2) which we now present.

Given some filtered probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}, \{\bar{\mathcal{F}}_t\})$ that supports iid m-dimensional Brownian motions $\{W_i\}_{i=1}^{\infty}$ and a k-dimensional Brownian motion B that is independent of the collection $\{W_i\}_{i=1}^{\infty}$ and such that for every s, $\{W_i(t)-W_i(s), B(t)-B(s), i\geq 1, t\geq s\}$ is independent of $\bar{\mathcal{F}}_s$, denote by $\mathcal{A}^{1,n}$ the class of $\bar{\mathcal{F}}_t$-progressively measurable processes $u : [0,T]\times\Omega\to\mathbb{R}^{nm}$ such that

$$\bar{E}\left[\int_0^T \|u(s)\|^2\, ds\right] < \infty.$$

For $u\in\mathcal{A}^{1,n}$, we will write $u=(u_1,\ldots,u_n)$, where u_i is the ith component of u and is m-dimensional. For $M\in(0,\infty)$, let

$$S_M \doteq \left\{v\in L^2([0,T]:\mathbb{R}^k) : \int_0^T \|v(s)\|^2\, ds \leq M\right\}.$$

This space will be equipped with the weak topology under which it is a compact space. Note that

$$\bigcup_{M\in\mathbb{N}} S_M = L^2([0,T]:\mathbb{R}^k).$$

Also let

$$\mathcal{A}^2_M \doteq \left\{\text{Progressively measurable } \mathbb{R}^k\text{-valued processes } v \text{ such that } v\in S_M\ \bar{P}\text{-a.s.}\right\},$$

and

$$\mathcal{A}^2 \doteq \left\{\text{Progressively measurable } \mathbb{R}^k\text{-valued processes } v \text{ such that } \bar{E}\left[\int_0^T \|v(s)\|^2\, ds\right] < \infty\right\}.$$

For $(u,v) \in \mathcal{A}^{1,n} \times \mathcal{A}^2$, consider the controlled analogue of the system in (4.0.1), driven by controls (u,v):

$$\begin{aligned} d\bar{X}_i^n(t) &= b(\bar{X}_i^n(t), \bar{\mu}^n(t))\,dt + \sigma(\bar{X}_i^n(t), \bar{\mu}^n(t))u_i(t)\,dt + \alpha(\bar{X}_i^n(t), \bar{\mu}^n(t))v(t)\,dt \\ &\qquad + \sigma(\bar{X}_i^n(t), \bar{\mu}^n(t))\,dW_i(t) + \kappa(n)\alpha(\bar{X}_i^n(t), \bar{\mu}^n(t))\,dB(t), \\ \bar{X}_i^n(0) &= x_i^n, \quad 1 \le i \le n, \end{aligned} \tag{4.2.3}$$

where $\bar{\mu}^n(t) = \frac{1}{n}\sum_{i=1}^n \delta_{\bar{X}_i^n(t)}$. Using the Lipschitz and boundedness conditions on the coefficients it is easy to check that the above system of equations has a unique solution. We also consider the empirical measure $\bar{\mu}^n = \frac{1}{n}\sum_{i=1}^n \delta_{\bar{X}_i^n}$ which is a $\mathcal{P}(\mathcal{X})$-valued random variable. A form of the following representation was first shown in [17]. The representation given below, that allows for an arbitrary filtered probability space on the right side was given in [20] (see also [22]). All expectations will be denoted by E unless specified otherwise.

Theorem 4.2.1. *For any $F \in \mathcal{C}_b(\mathcal{P}(\mathcal{X}))$ and for each $n \in \mathbb{N}$,*

$$-\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] = \inf_{(u,v)\in\mathcal{A}^{1,n}\times\mathcal{A}^2} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v(t)\|^2\,dt + F(\bar{\mu}^n)\right]. \tag{4.2.4}$$

Furthermore, for every $\delta > 0$, there is an $M < \infty$ such that for each $n \in \mathbb{N}$,

$$-\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] \ge \inf_{(u,v)\in\mathcal{A}^{1,n}\times\mathcal{A}^2_M} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v(t)\|^2\,dt + F(\bar{\mu}^n)\right] - \delta. \tag{4.2.5}$$

We now use the above result to complete the proof of (4.2.1) and (4.2.2).

4.2.1 Laplace Upper Bound

Throughout this section we assume that Conditions 4.1.1 and 4.1.2 are satisfied. As noted previously, the upper bound proof does not require Condition 4.1.3 and so this condition will not be used in this section.

Fix $F \in \mathcal{C}_b(\mathcal{P}(\mathcal{X}))$ and $\delta \in (0,1)$. From Theorem 4.2.1 there is an $M < \infty$ such that for each $n \in \mathbb{N}$, one can find $(u^n, v^n) \in \mathcal{A}^{1,n} \times \mathcal{A}^2_M$ such that

$$-\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] \ge E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\,dt + F(\bar{\mu}^n)\right] - \delta, \tag{4.2.6}$$

where $\bar{\mu}^n = \frac{1}{n}\sum_{i=1}^n \delta_{\bar{X}_i^n}$ and $\bar{X}_i^n$ are given by (4.2.3) (repalcing (u, v) with (u^n, v^n)). We will next show that

$$\begin{aligned}
&\liminf_{n\to\infty} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\,dt + F(\bar{\mu}^n)\right] \\
&\geq \inf_{\varphi\in L^2([0,T]:\mathbb{R}^k)} \inf_{\Theta\in\mathcal{E}_1[\varphi]} \left(E_\Theta\left[\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\,\rho(dy\,dt)\right] + \frac{1}{2\lambda^2}\int_0^T \|\varphi(t)\|^2\,dt + F([\Theta]_1)\right).
\end{aligned} \tag{4.2.7}$$

Since $\delta \in (0, 1)$ is arbitrary, the inequality in (4.2.1) is immediate from (4.2.7) on using the definition of I_1 in (4.1.3).

We now prove (4.2.7). From (4.2.6) it follows that

$$\sup_{n\in\mathbb{N}} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\,dt\right] \leq 2\|F\|_\infty + 1. \tag{4.2.8}$$

The following lemma shows that under such a uniform boundedness property, one has the tightness of certain key occupation measures.

Lemma 4.2.2. *Suppose for some $M \in (0, \infty)$, $\{(u^n, v^n)\}_{n\in\mathbb{N}}$ is a sequence with $(u^n, v^n) \in \mathcal{A}^{1,n} \times \mathcal{A}_M^2$ for each n, and suppose $\{u^n\}_{n\in\mathbb{N}}$ satisfies, for some $L \in (0, \infty)$,*

$$\sup_{n\in\mathbb{N}} E\left[\frac{1}{n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\,dt\right] \leq L. \tag{4.2.9}$$

Define $\mathcal{P}(\mathcal{Z}_1)$-valued random variables

$$Q^n(A \times R \times C) \doteq \frac{1}{n}\sum_{i=1}^n \delta_{\bar{X}_i^n}(A)\delta_{\rho_i^n}(R)\delta_{W_i}(C),\ A \times R \times C \in \mathcal{B}(\mathcal{Z}_1), \tag{4.2.10}$$

where $\bar{X}_i^n$ is defined as in (4.2.3) *(repalcing (u, v) with (u^n, v^n)), and*

$$\rho_i^n(E \times B) \doteq \int_B \delta_{u_i^n(t)}(E)\,dt,\ E \in \mathcal{B}(\mathbb{R}^m), B \in \mathcal{B}([0, T]). \tag{4.2.11}$$

Then $\{(Q^n, v^n)\}_{n\in\mathbb{N}}$ is tight as a sequence of $\mathcal{P}(\mathcal{Z}_1) \times S_M$-valued random variables.

Proof. Since S_M is compact, tightness of $\{v^n\}$ is immediate. The third marginals of Q^n are clearly tight since W_i are iid. The first marginal of Q^n, namely $[Q^n]_1$, equals $\bar{\mu}^n$. For each n let $\gamma^n = E[\bar{\mu}^n]$. For tightness of $\{\bar{\mu}^n\}_{n\in\mathbb{N}}$, it suffices to prove that the family $\{\gamma^n\}_{n\in\mathbb{N}}$ of measures on $\mathcal{X}$ is relatively compact.

By using the growth properties on the coefficients it follows that, for some $c_1 \in (0, \infty)$ and all $n \in \mathbb{N}$,

$$E\left[\sup_{0\leq s\leq T} \|\bar{X}_i^n(s)\|^2\right] \leq c_1\left(1 + \|x_i^n\|^2 + E\left[\int_0^T \|u_i^n(s)\|^2\,ds\right]\right). \tag{4.2.12}$$

Thus,

$$\sup_{n\in\mathbb{N}}\int_{\mathcal{X}}\sup_{0\le t\le T}\|\psi(t)\|^2\,d\gamma^n(\psi)=\sup_{n\in\mathbb{N}}\frac{1}{n}\sum_{i=1}^n E\left[\sup_{0\le t\le T}\|\bar X_i^n(t)\|^2\right]$$
$$\le c_1\sup_{n\in\mathbb{N}}\left(1+\frac{1}{n}\sum_{i=1}^n\|x_i^n\|^2+E\left[\frac{1}{n}\sum_{i=1}^n\int_0^T\|u_i^n(s)\|^2\,ds\right]\right)<\infty, \qquad (4.2.13)$$

where the last inequality is from (4.2.9) and Condition 4.1.1.

Next note that for any $\varepsilon\in(0,1)$ and $t\in[0,T-\varepsilon]$,

$$\|\bar X_i^n(t+\varepsilon)-\bar X_i^n(t)\|^2\le c_2\left(\left\|\int_t^{t+\varepsilon}b(\bar X_i^n(s),\bar\mu^n(s))\,ds\right\|^2+\left\|\int_t^{t+\varepsilon}\sigma(\bar X_i^n(s),\bar\mu^n(s))u_i^n(s)\,ds\right\|^2\right.$$
$$+\left\|\int_t^{t+\varepsilon}\alpha(\bar X_i^n(s),\bar\mu^n(s))v^n(s)\,ds\right\|^2+\left\|\int_t^{t+\varepsilon}\sigma(\bar X_i^n(s),\bar\mu^n(s))\,dW_i(s)\right\|^2$$
$$\left.+\kappa(n)^2\left\|\int_t^{t+\varepsilon}\alpha(\bar X_i^n(s),\bar\mu^n(s))\,dB(s)\right\|^2\right).$$

Thus for any stopping time τ taking values in $[0,T-\varepsilon]$, using the Cauchy-Schwarz inequality, the linear growth of b, and the boundedness of α and σ,

$$E\left[\left\|\bar X_i^n(\tau+\varepsilon)-\bar X_i^n(\tau)\right\|^2\right]\le c_3\varepsilon\left(1+E\left[\sup_{0\le s\le T}\|\bar X_i^n(s)\|^2\right]+E\left[\int_0^T\|u_i^n(s)\|^2\,ds\right]\right),$$

where the constant c_3 does not depend on n, ε, or the stopping time τ. Denoting by $\mathcal{T}_\varepsilon$ the collection of all stopping times τ, with respect to the canonical filtration generated by the coordinate process on $\mathcal{X}$, taking values in $[0,T-\varepsilon]$, we now have

$$\sup_{\tau\in\mathcal{T}_\varepsilon}\int_{\mathcal{X}}\|\varphi(\tau+\varepsilon)-\varphi(\tau)\|^2\,d\gamma^n(\varphi)$$
$$\le c_3\varepsilon\left(1+\frac{1}{n}\sum_{i=1}^n E\left[\sup_{0\le s\le T}\|\bar X_i^n(s)\|^2\right]+E\left[\frac{1}{n}\sum_{i=1}^n\int_0^T\|u_i^n(s)\|^2\,ds\right]\right)$$
$$\le c_3\varepsilon\left(1+\int_{\mathcal{X}}\sup_{0\le t\le T}\|\psi(t)\|^2\,d\gamma^n(\psi)+L\right).$$

Using (4.2.13) in the above display,

$$\limsup_{\varepsilon\to0}\sup_{n\in\mathbb{N}}\sup_{\tau\in\mathcal{T}_\varepsilon}\int_{\mathcal{X}}\|\varphi(\tau+\varepsilon)-\varphi(\tau)\|^2\,d\gamma^n(\varphi)=0.$$

Thus from the Aldous-Kurtz tightness criterion, we have that the collection γ^n is relatively compact which, as noted previously, gives the tightness of the collection $\{\bar\mu^n\}=\{[Q^n]_1\}$.

Finally we consider the second marginals of Q^n. Define

$$g(r) \doteq \int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \, r(dy\, dt), \; r \in \mathcal{R}_1.$$

We note that g has compact level sets. Indeed, for $c \in \mathbb{R}_+$, let $L_c = \{r \in \mathcal{R}_1 : g(r) \leq c\}$ denote the corresponding level set. By Chebyshev's inequality,

$$\sup_{r \in L_c} r\left(\left\{y \in \mathbb{R}^d : \|y\| > M\right\} \times [0,T]\right) \leq \sup_{r \in L_c} \frac{g(r)}{M^2} \leq \frac{c}{M^2} \to 0$$

as $M \to \infty$. This shows that L_c is relatively compact in $\mathcal{R}$. Let $\{r_n\} \subset L_c$ be a sequence that converges in $\mathcal{R}$ to some r^*. By Fatou's lemma, $g(r^*) \leq c$, and so $r^* \in L_c$. Also, by the uniform integrability that follows from

$$\sup_{n \geq 1} \int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \, r_n(dy\, dt) = \sup_{n \geq 1} g(r_n) \leq c,$$

the moments of r_n also converge to the moments of r^*. Thus $r_n \to r^*$ in $\mathcal{R}_1$, establishing compactness of L_c in $\mathcal{R}_1$. Let $G : \mathcal{P}(\mathcal{R}_1) \to [0, \infty]$ be given as

$$G(\theta) \doteq \int_{\mathcal{R}_1} g(r)\, \theta(dr).$$

Then G is a tightness function on $\mathcal{P}(\mathcal{R}_1)$ (namely it has relatively compact level sets), and thus to establish the tightness of the second marginals $\{[Q^n]_2\}$, it suffices to show that

$$\sup_{n \geq 1} E[G([Q^n]_2)] < \infty. \tag{4.2.14}$$

For each $n \in \mathbb{N}$,

$$\begin{aligned} E[G([Q^n]_2)] &= E\left[\int_{\mathcal{R}_1} g(r)\, [Q^n]_2(dr)\right] = E\left[\frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \, \rho_i^n(dy\, dt)\right] \\ &= E\left[\frac{1}{n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2 \, dt\right] \leq L. \end{aligned}$$

This proves (4.2.14) and completes the proof of the tightness of $\{[Q^n]_2\}$. The result follows. $\square$

The next lemma characterizes the weak limit points of the sequence (Q^n, v^n). Recall the collection $\mathcal{E}_1[\varphi]$ from (4.1.2).

Lemma 4.2.3. *Suppose, for some $M \in (0,\infty)$, $\{(u^n, v^n)\}_{n\in\mathbb{N}}$ is a sequence with $(u^n, v^n) \in \mathcal{A}^{1,n} \times \mathcal{A}^2_M$ for each n, and such that $\{u^n\}_{n\in\mathbb{N}}$ satisfies* (4.2.9) *with some $L \in (0,\infty)$. Let Q^n be defined as in Lemma 4.2.2. If (Q^n, v^n) converges in distribution, along some subsequence, to (Q, v), then $Q \in \mathcal{E}_1[v]$ a.s.*

Proof. Let (Q, v) be a weak limit point of (Q^n, v^n) given on some probability space $(\Omega^*, \mathcal{F}^*, P^*)$. Note that by Fatou's lemma,

$$E^*\left[\int_{\mathcal{R}_1}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\, r(dy\, dt)\, [Q]_2(dr)\right] \le \liminf_{n\to\infty} E\left[\frac{1}{n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\, dt\right] \le L, \tag{4.2.15}$$

Thus $Q \in \mathcal{P}_2(\mathcal{Z}_1)$, P^*-a.s. Also, since $\int_0^T \|v(s)\|^2 ds \le M$, $v \in L^2([0,T] : \mathbb{R}^k)$ P^*-a.s. To complete the proof, we need to argue that for P^*-a.e. $\omega \in \Omega^*$, $Q(\omega)$ is a weak solution to $\mathcal{S}_1[v(\omega), \nu_{Q(\omega)}]$.

Denote the canonical coordinate variables on $\mathcal{Z}_1$ by (z, r, w). By Condition 4.1.1, $[Q^n]_1 \circ (z(0))^{-1} \to \xi_0$ weakly, which shows that, for P^*-a.e. ω, under $Q(\omega)$, $z(0)$ has distribution ξ_0. Denote by $\{\mathcal{H}_t\}_{0\le t\le T}$ the canonical filtration on $(\mathcal{Z}_1, \mathcal{B}(\mathcal{Z}_1))$, namely

$$\mathcal{H}_t \doteq \sigma\{z(s), w(s), r(A\times[0,s]),\ A \in \mathcal{B}(\mathbb{R}^m),\ s \le t\}. \tag{4.2.16}$$

For $f \in \mathcal{C}_c^2(\mathbb{R}^d \times \mathbb{R}^m)$, $\varphi \in L^2([0,T] : \mathbb{R}^k)$, and $\Theta \in \mathcal{P}(\mathcal{Z}_1)$, consider the process $\{M^\Theta_{f,\varphi}(t)\}_{0\le t\le T}$ defined on the probability space $(\mathcal{Z}_1, \mathcal{B}(\mathcal{Z}_1), \Theta)$ by

$$\begin{aligned} M^\Theta_{f,\varphi}(t,(z,r,w)) \doteq{}& f(z(t),w(t)) - f(z(0),0) - \int_0^t\int_{\mathbb{R}^m} \mathcal{L}_s^\Theta(f)(z(s),y,w(s))\, r_s(dy)\, ds \\ & - \int_0^t \langle \alpha(z(s),\nu_\Theta(s))\varphi(s), \nabla_x f(z(s),w(s))\rangle\, ds, \end{aligned} \tag{4.2.17}$$

where

$$\begin{aligned} \mathcal{L}_s^\Theta(f)(x,y,w) \doteq{}& \langle b(x,\nu_\Theta(s)) + \sigma(x,\nu_\Theta(s))y, \nabla_x f(x,w)\rangle \\ & + \frac{1}{2}\sum_{j,j'=1}^d (\sigma\sigma^T)_{jj'}(x,\nu_\Theta(s)) \frac{\partial^2 f}{\partial x_j \partial x_{j'}}(x,w) + \frac{1}{2}\sum_{j=1}^m \frac{\partial^2 f}{\partial w_j^2}(x,w) \\ & + \sum_{j=1}^d\sum_{j'=1}^m \sigma_{jj'}(x,\nu_\Theta(s)) \frac{\partial^2 f}{\partial x_j \partial w_{j'}}(x,w) \end{aligned} \tag{4.2.18}$$

for $x \in \mathbb{R}^d$ and $y, w \in \mathbb{R}^m$. Let, for $B \in (0,\infty)$, $\zeta_B : \mathbb{R}^m \to \mathbb{R}^m$ be such that ζ_B is a continuous function with compact support satisfying $\zeta_B(y) = y$ for $\|y\| \le B$ and $\|\zeta_B(y)\| \le \|y\| + 1$ for every $y \in \mathbb{R}^m$. It will be convenient to also consider, along with $\mathcal{L}_s^\Theta$, the operator $\mathcal{L}_s^{\Theta,B}$ which is defined by replacing y on the right side of (4.2.18) with $\zeta_B(y)$. Similarly, define $M^{\Theta,B}_{f,\varphi}$ by replacing $\mathcal{L}_s^\Theta$ in (4.2.17) with $\mathcal{L}_s^{\Theta,B}$.

It suffices to show that for each $f \in \mathcal{C}_c^2(\mathbb{R}^d \times \mathbb{R}^m)$, any time instants $0 \le t_0 < t_1 \le T$, and any $\Psi \in \mathcal{C}_b(\mathcal{Z}_1)$ that is measurable with respect to the sigma field $\mathcal{H}_{t_0}$, we have,

$$E_{Q(\omega)}\left[\Psi\left(M^{Q(\omega)}_{f,v(\omega)}(t_1) - M^{Q(\omega)}_{f,v(\omega)}(t_0)\right)\right] = 0 \text{ for } P^*\text{-a.e. } \omega \in \Omega^*. \tag{4.2.19}$$

In the rest of the proof we suppress ω from the notation. Fix a choice of (t_0, t_1, Ψ, f) and define $\Phi : \mathcal{P}(\mathcal{Z}_1) \times S_M \to \mathbb{R}$ by

$$\Phi(\Theta, \varphi) = E_\Theta\left[\Psi\left(M^\Theta_{f,\varphi}(t_1) - M^\Theta_{f,\varphi}(t_0)\right)\right]. \tag{4.2.20}$$

Also, for every $B \in (0,\infty)$, define Φ_B by replacing $M^\Theta_{f,\varphi}$ with $M^{\Theta,B}_{f,\varphi}$ in the definition of Φ. We will now show that (a) for every $B \in (0,\infty)$, Φ_B is a bounded and continuous map on $\mathcal{P}(\mathcal{Z}_1)\times S_M$, (b) $\sup_n E^*|\Phi_B(Q^n, v^n) - \Phi(Q^n, v^n)| \to 0$ and $E^*|\Phi_B(Q, v) - \Phi(Q, v)| \to 0$ as $B \to \infty$, and (c) $\Phi(Q^n, v^n) \to 0$ in probability as $n \to \infty$. The statement in (4.2.19) is an immediate consequence of (a)-(c).

We first show (a). Let $(\Theta_n, \varphi_n) \to (\Theta, \varphi)$ in $\mathcal{P}(\mathcal{Z}_1) \times S_M$ as $n \to \infty$. Note that this means $\int_0^T \langle \varphi_n(s) - \varphi(s), h(s)\rangle\, ds \to 0$ for all $h \in L^2([0,T] : \mathbb{R}^k)$. Thus,

$$\begin{aligned}
|\Phi_B(\Theta,\varphi_n) - \Phi_B(\Theta,\varphi)| &\le \|\Psi\|_\infty E_\Theta\left|\int_{t_0}^{t_1} \langle \alpha(z(s),\nu_\Theta(s))(\varphi_n(s)-\varphi(s)), \nabla_x f(z(s),w(s))\rangle\; ds\right| \\
&= \|\Psi\|_\infty E_\Theta\left|\int_0^T 1_{[t_0,t_1]}(s)\left\langle(\varphi_n(s)-\varphi(s)), \alpha^T(z(s),\nu_\Theta(s))\nabla_x f(z(s),w(s))\right\rangle\, ds\right| \\
&\to 0
\end{aligned} \tag{4.2.21}$$

as $n \to \infty$, where the last convergence follows from the dominated convergence theorem upon observing that $h(\cdot) = \alpha^T(z(\cdot),\nu_\Theta(\cdot))\nabla_x f(z(\cdot),w(\cdot))1_{[t_0,t_1]}(\cdot)$ is in $L^2([0,T] : \mathbb{R}^k)$. Next note that

$$\sup_{\bar\varphi \in S_M}\left|E_{\Theta_n}\left[\Psi\cdot\left(M^{\Theta,B}_{f,\bar\varphi}(t_1) - M^{\Theta,B}_{f,\bar\varphi}(t_0)\right)\right] - E_\Theta\left[\Psi\cdot\left(M^{\Theta,B}_{f,\bar\varphi}(t_1) - M^{\Theta,B}_{f,\bar\varphi}(t_0)\right)\right]\right| \to 0 \tag{4.2.22}$$

as $n \to \infty$. This convergence is a consequence of the following facts: (i) Continuity and boundedness of the map $(z,r,w) \mapsto f(z(t),w(t)) - \int_0^t\int_{\mathbb{R}^m} \mathcal{L}^{\Theta,B}_s(f)(z(s),y,w(s))\, r_s(dy)\, ds$, (ii) the continuity and boundedness of the map $(z,w) \mapsto \alpha^T(z(s),\nu_\Theta(s))\nabla_x f(z(s),w(s))$, (iii) the property that $\sup_{\bar\varphi\in S_M}\int_0^T \|\bar\varphi(s)\|^2 ds \le M$, and Cauchy-Schwarz inequality. Next, for some $c_1 \in (0,\infty)$ (possibly depending on B), and all $t \in [0,T]$, $\bar\varphi \in S_M$

$$\begin{aligned}
\left|M^{\Theta_n,B}_{f,\bar\varphi}(t) - M^{\Theta,B}_{f,\bar\varphi}(t)\right| &\le \int_0^T\int_{\mathbb{R}^m}\left|\mathcal{L}^{\Theta_n,B}_s(f)(z(s),y,w(s)) - \mathcal{L}^{\Theta,B}_s(f)(z(s),y,w(s))\right|\, r_s(dy)\, ds \\
&\quad + \int_0^T |\langle(\alpha(z(s),\nu_{\Theta_n}(s)) - \alpha(z(s),\nu_\Theta(s))\bar\varphi(s), \nabla_x f(z(s),w(s))\rangle|\; ds
\end{aligned}$$

$$\leq c_1 \left(\int_0^T d_{BL}(\nu_{\Theta_n}(s), \nu_\Theta(s))^2 \, ds \right)^{1/2}.$$

Since, for every $s \in [0,T]$, $\nu_{\Theta_n}(s) \to \nu_\Theta(s)$, we now have

$$\sup_{\bar{\varphi} \in S_M} \left| E_{\Theta_n} \left[\Psi \cdot \left(M_{f,\bar{\varphi}}^{\Theta_n,B}(t_1) - M_{f,\bar{\varphi}}^{\Theta_n,B}(t_0) \right) \right] - E_{\Theta_n} \left[\Psi \cdot \left(M_{f,\bar{\varphi}}^{\Theta,B}(t_1) - M_{f,\bar{\varphi}}^{\Theta,B}(t_0) \right) \right] \right| \to 0 \qquad (4.2.23)$$

as $n \to \infty$. Combining (4.2.22) and (4.2.23)

$$\sup_{\bar{\varphi} \in S_M} |\Phi_B(\Theta_n, \bar{\varphi}) - \Phi_B(\Theta, \bar{\varphi})| \to 0$$

as $n \to \infty$. Together with (4.2.21), the above display completes the proof of (a).

In order to see (b), note that, for some $c_2 \in (0, \infty)$, and every $n \in \mathbb{N}$,

$$\begin{aligned} E|\Phi_B(Q^n, v^n) - \Phi(Q^n, v^n)| &\leq c_2 E \left[E_{Q^n} \left[\int_0^T \left\| \int_{\mathbb{R}^m} (y - \zeta_B(y)) \, r_s(dy) \right\| ds \right] \right] \\ &= c_2 E \left[\frac{1}{n} \sum_{i=1}^n \int_0^T \|u_i^n(s) - \zeta_B(u_i^n(s))\| ds \right] \\ &\leq \frac{c_2}{B} E \left[\frac{1}{n} \sum_{i=1}^n \int_0^T 2(\|u_i^n(s)\| + 1) \|u_i^n(s)\| ds \right] \leq \frac{4c_2(L+T)}{B}. \qquad (4.2.24) \end{aligned}$$

The first statement in (b) is now immediate. The second statement in (b) is shown similarly by using (4.2.15).

Finally we consider (c). By the definition of Q^n and since $\nu_{Q^n}(s) = \bar{\mu}^n(s)$,

$$\begin{aligned} &\Phi(Q^n, v^n) \\ &= E_{Q^n} \left[\Psi \left(M_{f,v^n}^{Q^n}(t_1) - M_{f,v^n}^{Q^n}(t_0) \right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \Psi(\bar{X}_i^n, \rho_i^n, W_i) \cdot \left(M_{f,v^n}^{Q^n}(t_1, (\bar{X}_i^n, \rho_i^n, W_i)) - M_{f,v^n}^{Q^n}(t_0, (\bar{X}_i^n, \rho_i^n, W_i)) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \Psi(\bar{X}_i^n, \rho_i^n, W_i) \cdot \Bigg(f(\bar{X}_i^n(t_1), W_i(t_1)) - f(\bar{X}_i^n(t_0), W_i(t_0)) \\ &\quad - \int_{t_0}^{t_1} \mathcal{L}_s^{Q^n}(f)(\bar{X}_i^n(s), u_i^n(s), W_i(s)) \, ds - \int_{t_0}^{t_1} \left\langle \alpha(\bar{X}_i^n(s), \bar{\mu}^n(s)) v^n(s), \nabla_x f(\bar{X}_i^n(s), W_i(s)) \right\rangle \, ds \Bigg), \end{aligned}$$

By Itô's formula, for each i, a.s.

$$\begin{aligned} &f(\bar{X}_i^n(t_1), W_i(t_1)) - f(\bar{X}_i^n(t_0), W_i(t_0)) \\ &= \int_{t_0}^{t_1} \mathcal{L}_s^{Q^n}(f)(\bar{X}_i^n(s), u_i^n(s), W_i(s)) \, ds + \int_{t_0}^{t_1} \left\langle \alpha(\bar{X}_i^n(s), \bar{\mu}^n(s)) v^n(s), \nabla_x f(\bar{X}_i^n(s), W_i(s)) \right\rangle \, ds \end{aligned}$$

$$
\begin{aligned}
&+\int_{t_0}^{t_1}\left[\nabla_x f(\bar{X}_i^n(s), W_i(s))\right]^T \sigma(\bar{X}_i^n(s), \bar{\mu}^n(s))\, dW_i(s) + \int_{t_0}^{t_1}\left[\nabla_w f(\bar{X}_i^n(s), W_i(s))\right]^T\, dW_i(s)\\
&+\kappa(n)\int_{t_0}^{t_1}\left[\nabla_x f(\bar{X}_i^n(s), W_i(s))\right]^T \alpha(\bar{X}_i^n(s), \bar{\mu}^n(s))\, dB(s)\\
&+\frac{\kappa(n)^2}{2}\int_{t_0}^{t_1} \mathrm{tr}\, \left((\alpha\alpha^T)(\bar{X}_i^n(s), \bar{\mu}^n(s)) D_x^2 f(\bar{X}_i^n(s), W_i(s))\right)\, ds.
\end{aligned}
$$

Writing $\Psi_i^n = \Psi(\bar{X}_i^n, \rho_i^n, W_i)$, we then have

$$
\begin{aligned}
\Phi(Q^n, v^n) = &\frac{1}{n}\sum_{i=1}^n \Psi_i^n \int_{t_0}^{t_1}\left[\nabla_x f(\bar{X}_i^n(s), W_i(s))\right]^T \sigma(\bar{X}_i^n(s), \bar{\mu}^n(s))\, dW_i(s)\\
&+\frac{1}{n}\sum_{i=1}^n \Psi_i^n \int_{t_0}^{t_1}\left[\nabla_w f(\bar{X}_i^n(s), W_i(s))\right]^T\, dW_i(s) + \mathcal{T}_1^n,
\end{aligned}
$$

where using the fact that $\kappa(n) \to 0$ as $n \to \infty$, we have that $\mathcal{T}_1^n \to 0$ in probability as $n \to \infty$.

Denote the first two terms on the right side of above display as J_n^1 and J_n^2 respectively. Using the boundedness of Ψ_i^n, σ, $\nabla_x f$, the independence of the W_i, the fact that Ψ_i^n are $\mathcal{H}_{t_0}$ measurable, and Itô's isometry, $E[(J_n^1)^2] \le c_3/n$ for some $c_3 \in (0, \infty)$ and all $n \in \mathbb{N}$. Thus $J_n^1 \to 0$ in probability as $n \to \infty$. Similarly, $J_n^2 \to 0$ in probability as $n \to \infty$. Combining the above observations we have that $\Phi(Q^n, v^n) \to 0$ in probability, completing the proof of (c) and therefore of the lemma. □

Finally we complete the proof of the Laplace upper bound (4.2.1) by proving (4.2.7). By the definition of Q^n,

$$
\begin{aligned}
&E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\, dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\, dt + F(\bar{\mu}^n)\right]\\
&= E\left[\int_{\mathcal{R}_1}\left(\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\, r(dy\, dt)\right)\, [Q^n]_2(dr) + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\, dt + F([Q^n]_1)\right].
\end{aligned}
$$

Recall the uniform bound (4.2.8). Then from Lemmas 4.2.2 and 4.2.3, (Q^n, v^n) is tight and if (Q, v) is a weak limit point then $Q \in \mathcal{E}_1[v]$ a.s. Assume without loss of generality that $(Q^n, v^n) \to (Q, v)$ along the full sequence. Then by Fatou's lemma and since $\sqrt{n}\kappa(n) \to \lambda$,

$$
\begin{aligned}
&\liminf_{n\to\infty} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\, dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\, dt + F(\bar{\mu}^n)\right]\\
&\ge E\left[\int_{\mathcal{R}_1}\left(\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\, r(dy\, dt)\right)\, [Q]_2(dr) + \frac{1}{2\lambda^2}\int_0^T \|v(t)\|^2\, dt + F([Q]_1)\right]\\
&\ge \inf_{\varphi\in L^2([0,T]:\mathbb{R}^k)}\ \inf_{\Theta\in\mathcal{E}_1[\varphi]}\left(E_\Theta\left[\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\, \rho(dy\, dt)\right] + \frac{1}{2\lambda^2}\int_0^T \|\varphi(t)\|^2\, dt + F([\Theta]_1)\right),
\end{aligned}
$$

where the last inequality uses the fact that $Q \in \mathcal{E}_1[v]$ a.s. This completes the proof of the Laplace upper bound. □

4.2.2 Laplace Lower Bound

Throughout this section we assume that Conditions 4.1.1, 4.1.2, and 4.1.3 are satisfied. Fix $\varepsilon > 0$ and $F \in \mathcal{C}_b(\mathcal{P}(\mathcal{X}))$. Choose a $\varphi \in L^2([0,T]:\mathbb{R}^k)$ and a $\Theta \in \mathcal{E}_1[\varphi]$ such that

$$\frac{1}{2}E_\Theta\left[\int_{\mathbb{R}^m\times[0,T]}\|y\|^2\rho(dy\,dt)\right]+\frac{1}{2\lambda^2}\int_0^T\|\varphi(t)\|^2\,dt+F([\Theta]_1)\leq\inf_{\nu\in\mathcal{P}(\mathcal{X})}[F(\nu)+I_1(\nu)]+\varepsilon. \tag{4.2.25}$$

We will show that there is an $M \in (0,\infty)$ and a sequence (u^n, v^n) with $u^n \in \mathcal{A}^{1,n}$ and $v^n \in \mathcal{A}^2_M$ constructed on some filtered probability space such that

$$\begin{aligned}&\limsup_{n\to\infty}E\left[\frac{1}{2n}\sum_{i=1}^n\int_0^T\|u_i^n(t)\|^2\,dt+\frac{1}{2n\kappa(n)^2}\int_0^T\|v^n(t)\|^2\,dt+F(\bar\mu^n)\right]\\&\leq\frac{1}{2}E_\Theta\left[\int_{\mathbb{R}^m\times[0,T]}\|y\|^2\rho(dy\,dt)\right]+\frac{1}{2\lambda^2}\int_0^T\|\varphi(t)\|^2\,dt+F([\Theta]_1).\end{aligned} \tag{4.2.26}$$

The Laplace lower bound (4.2.2) is then immediate from Theorem 4.2.1 on noting that $\varepsilon > 0$ is arbitrary. The key ingredient in the proof of (4.2.26) is the following uniqueness result. Define the map $\vartheta : \mathcal{Z}_1 \to \mathcal{Z}_1^\vartheta \doteq \mathbb{R}^d \times \mathcal{R}_1 \times \mathcal{W}$ as $\vartheta(z,r,w) \doteq (z(0),r,w)$. For $\Theta \in \mathcal{P}(\mathcal{Z}_1)$, let $\Theta_\vartheta \doteq \Theta\circ\vartheta^{-1}$ be the probability measure on $\mathcal{Z}_1^\vartheta$ induced by Θ under ϑ.

We will say that *weak uniqueness* holds for (4.1.1) if, for any given $\varphi \in L^2([0,T]:\mathbb{R}^k)$ and $\Theta^{(1)},\Theta^{(2)} \in \mathcal{E}_1[\varphi]$, whenever $\Theta^{(1)}_\vartheta = \Theta^{(2)}_\vartheta$, we have that $\Theta^{(1)} = \Theta^{(2)}$.

Lemma 4.2.4. *Weak uniqueness holds for* (4.1.1).

Proof. Fix $\varphi \in L^2([0,T]:\mathbb{R}^k)$ and $\Theta^{(1)},\Theta^{(2)} \in \mathcal{E}_1[\varphi]$. Suppose that $\Theta^{(1)}_\vartheta = \Theta^{(2)}_\vartheta \doteq \Lambda$. Note that $\Theta^{(i)}$, $i = 1,2$ can be disintegrated as

$$\Theta^{(i)}(dx,\,dr,\,dw)=\tilde\Theta^{(i)}(r,w,x_0,dx)\Lambda(dx_0,\,dr,\,dw).$$

Consider $\hat{\mathcal{Z}}_1 = \mathcal{X}\times\mathcal{X}\times\mathbb{R}^d\times\mathcal{R}_1\times\mathcal{W}$. Define $\hat\Theta \in \mathcal{P}(\hat{\mathcal{Z}}_1)$ as

$$\hat\Theta(dx^{(1)},\,dx^{(2)},\,dx_0,\,dr,\,dw)\doteq\tilde\Theta^{(1)}(r,\,w,\,x_0,\,dx^{(1)})\tilde\Theta^{(2)}(r,\,w,\,x_0,\,dx^{(2)})\Lambda(dx_0,\,dr,\,dw)$$

and denote the coordinate maps on $\hat{\mathcal{Z}}_1$ as $(X^{(1)},X^{(2)},X_0,\rho,W)$. Note that the process W is a Brownian motion with respect to the canonical filtration

$$\hat{\mathcal{H}}_t\doteq\sigma\left\{X^{(1)}(s),X^{(2)}(s),\rho(A\times[0,s]),W(s),\ A\in\mathcal{B}(\mathbb{R}^m),s\in[0,t]\right\},\ t\in[0,T],$$

and for $i=1,2$, $X^{(i)}$ satisfy (4.1.1) with $\bar X$ replaced with $X^{(i)}$ and $\nu(t)$ replaced with $\nu^{(i)}(t)\doteq\nu_{\Theta^{(i)}}(t)$. Also, $X^{(i)}(0) = X_0$ for $i = 1,2$. In order to prove the lemma it suffices to show that $X^{(1)} = X^{(2)}$ a.s. Let $u(t) \doteq$

$\int_{\mathbb{R}^m} y\rho_t(dy)$, $t \in [0,T]$. Then $E_{\hat\Theta} \int_0^T \|u(t)\|^2 dt < \infty$. By the Lipschitz properties of b, α, and σ, the property that $\sigma(x,v) \equiv \sigma(v)$, and since $\varphi \in L^2([0,T] : \mathbb{R}^k)$, we have that, for some $c_1 \in (0,\infty)$ and for any $t \in [0,T]$,

$$\begin{aligned} E_{\hat\Theta}\left[\sup_{0\le s\le t} \|X^{(1)}(s) - X^{(2)}(s)\|^2\right] &\le c_1 \int_0^t \left(E_{\hat\Theta}\|X^{(1)}(s) - X^{(2)}(s)\|^2 + d_{BL}(\nu^{(1)}(s), \nu^{(2)}(s))^2\right) ds \\ &\quad + c_1 E_{\hat\Theta}\left(\int_0^T d_{BL}(\nu^{(1)}(s), \nu^{(2)}(s)) \cdot \|u(s)\|\, ds\right)^2. \end{aligned} \tag{4.2.27}$$

Since

$$E_{\hat\Theta}\left(\int_0^t d_{BL}(\nu^{(1)}(s), \nu^{(2)}(s)) \cdot \|u(s)\|\, ds\right)^2 \le \int_0^t d_{BL}(\nu^{(1)}(s), \nu^{(2)}(s))^2\, ds \cdot E_{\hat\Theta}\left[\int_0^T \|u(s)\|^2\, ds\right]$$

and $E_{\hat\Theta} \int_0^T \|u(s)\|^2\, ds < \infty$, we have, for all $t \in [0,T]$,

$$E_{\hat\Theta}\left[\sup_{0\le s\le t} \|X^{(1)}(s) - X^{(2)}(s)\|^2\right] \le c_2 \int_0^t \left(E_{\hat\Theta}\|X^{(1)}(s) - X^{(2)}(s)\|^2 + d_{BL}(\nu^{(1)}(s), \nu^{(2)}(s))^2\right) ds.$$

Furthermore, for each t,

$$d_{BL}(\nu^{(1)}(t), \nu^{(2)}(t)) = \sup_{f\in BL(\mathbb{R}^d)} \left|\int_{\hat{\mathcal{Z}}_1} f(X^{(1)}(t))\, d\hat\Theta - \int_{\hat{\mathcal{Z}}_1} f(X^{(2)}(t))\, d\hat\Theta\right| \le E_{\hat\Theta}\|X^{(1)}(t) - X^{(2)}(t)\|.$$

Thus, for some $c_3 \in (0,\infty)$, we have, for all $t \in [0,T]$,

$$E_{\hat\Theta}\left[\sup_{0\le s\le t} \|X^{(1)}(s) - X^{(2)}(s)\|^2\right] \le c_3 \int_0^t E_{\hat\Theta}\left[\sup_{0\le s\le \tau} \|X^{(1)}(s) - X^{(2)}(s)\|^2\right] d\tau.$$

By Gronwall's inequality, this shows that $X^{(1)}$ and $X^{(2)}$ are indistinguishable on $[0,T]$ and completes the proof of the lemma. □

Now we return to the construction of (u^n, v^n) that satisfy (4.2.26), where recall that Θ and φ are chosen to satisfy (4.2.25). Let $(\bar X, \rho, W)$ be the coordinate maps on the space $(\mathcal{Z}_1, \mathcal{B}(\mathcal{Z}_1), \Theta)$ equipped with the canonical filtration $\mathcal{H}_t$, defined in (4.2.16), namely

$$\mathcal{H}_t = \sigma\left\{X(s), \rho(A \times [0,s]), W(s) : A \in \mathcal{B}(\mathbb{R}^m), s \le t\right\}.$$

Since $\Theta \in \mathcal{E}_1[\varphi]$, equation (4.1.1) is satisfied with $\nu(t) = \nu_\Theta(t)$ and $\nu_\Theta(0) = \xi_0$.

Disintegrate Θ_ϑ as

$$\Theta_\vartheta(dx, dr, dw) = \xi_0(dx)\, [\Theta]_3(dw)\, \hat\Lambda_0(x, w, dr).$$

Let $\mathcal{V} \doteq \mathcal{C}([0,T]:\mathbb{R}^k)$ and define

$$\Omega' \doteq (\mathcal{R}_1 \times \mathcal{W})^\infty \times \mathcal{V}, \quad \mathcal{F}' \doteq \mathcal{B}(\Omega').$$

Elements of Ω' are of the form (r, w, β), where $\beta \in \mathcal{V}$, $r = (r_1, r_2, \ldots)$, $w = (w_1, w_2, \ldots)$, $r_i \in \mathcal{R}_1$ and $w_i \in \mathcal{W}$ for each $i \in \mathbb{N}$. On the measurable space $(\Omega', \mathcal{F}')$ define the random variables

$$W_i(t,(r,w,\beta)) \doteq w_i(t), \qquad B(t,(r,w,\beta)) \doteq \beta(t), \qquad \rho_i(r,w,\beta) \doteq r_i,$$

for each $t \in [0,T]$ and $i \in \mathbb{N}$. Let Γ be the standard Wiener measure on $\mathcal{V}$. Recall the initial values $\{x_i^n\}$ introduced in Section 4.1.1. For each $n \in \mathbb{N}$, define the probability measure P^n on $(\Omega', \mathcal{F}')$ by

$$dP^n(r,w,\beta) = \left[\bigotimes_{i=1}^n [\Theta]_3(dw_i)\,\hat{\Lambda}_0(x_i^n, w_i, dr_i) \bigotimes_{i=n+1}^\infty [\Theta]_{(2,3)}(dr_i, dw_i)\right] \otimes \Gamma(d\beta).$$

Under P^n, $\{W_i\}_{1\le i\le n}$ and B are mutually independent Brownian motions. Define the sequence $\{\Lambda^n\}_{n\in\mathbb{N}}$ of $\mathcal{P}(\mathbb{R}^d \times \mathcal{R}_1 \times \mathcal{W})$-valued random variables on $(\Omega', \mathcal{F}')$ by

$$\Lambda^n(A \times R \times C) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}(A)\delta_{\rho_i}(R)\delta_{W_i}(C),\; A \times R \times C \in \mathcal{B}(\mathbb{R}^d \times \mathcal{R}_1 \times \mathcal{W}).$$

Then by Condition 4.1.1,

$$P^n \circ (\Lambda^n)^{-1} \to \delta_{\Theta_\vartheta}. \tag{4.2.28}$$

Let, for $n \in \mathbb{N}$, $v^n \doteq \varphi$. Denoting $\int_0^T \|\varphi(s)\|^2 ds \doteq M$, we have that $v^n \in S_M$ for every n. Next, for each $i \in \mathbb{N}$, let

$$u_i(t) \doteq \int_{\mathbb{R}^m} y\,(\rho_i)_t(dy),\; t \in [0,T], \tag{4.2.29}$$

where $(\rho_i)_t(dy)\,dt = \rho_i(dy\,dt)$, and for each $n \in \mathbb{N}$, let $(\bar{X}_1^n, \ldots, \bar{X}_n^n)$ be the solution on $(\Omega', \mathcal{F}', P^n)$ of the system (4.2.3), where $\bar{\mu}^n(t) = \frac{1}{n}\sum_{i=1}^n \delta_{\bar{X}_i^n(t)}$ for each $t \in [0,T]$. Unique solvability of the above equation is a consequence of our assumptions on the coefficients, namely Condition 4.1.2.

For each n, define the occupation measure Q^n by (4.2.10), replacing ρ_i^n with ρ_i. That is,

$$Q^n(B \times R \times D) \doteq \frac{1}{n}\sum_{i=1}^n \delta_{\bar{X}_i^n}(B)\delta_{\rho_i}(R)\delta_{W_i}(D),\; B \times R \times D \in \mathcal{B}(\mathcal{Z}_1).$$

Let E^n denote expectation over the probability measure P^n. Then

$$\begin{aligned}
&\limsup_{n\to\infty} E^n\left[\frac{1}{n}\sum_{i=1}^n \int_0^T \|u_i(t)\|^2\,dt\right]\\
&= \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \int_{\mathcal{R}_1\times\mathcal{W}} \int_0^T \left\|\int_{\mathbb{R}^m} y\, r_t(dy)\right\|^2 dt\, \hat{\Lambda}(x_i^n, w, dr)\, [\Theta]_3(dw)\\
&= E_\Theta\left[\int_0^T \left\|\int_{\mathbb{R}^m} y\,\rho_t(dy)\right\|^2 dt\right] \le E_\Theta\left[\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\,\rho(dy\,dt)\right] < \infty, \qquad (4.2.30)
\end{aligned}$$

where the second equality is from Condition 4.1.1. It follows from Lemma 4.2.2 that $\{(Q^n, v^n)\}_{n\in\mathbb{N}}$ is tight. If (Q, v) is a limit point of this sequence defined on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, then $v = \varphi$ $\tilde{P}$-a.s., and, by Lemma 4.2.3, $Q \in \mathcal{E}_1[v] = \mathcal{E}_1[\varphi]$ $\tilde{P}$-a.s. Recall that $\Theta \in \mathcal{E}_1[\varphi]$ as well. By (4.2.28), for $\tilde{P}$-a.e. $\omega \in \tilde{\Omega}$, $Q_\vartheta(\omega) = \Theta_\vartheta$. Thus by the weak uniqueness established in Lemma 4.2.4, $Q = \Theta$ $\tilde{P}$-a.s., and so $Q^n \to \Theta$ in probability. Finally,

$$\begin{aligned}
&\limsup_{n\to\infty} E^n\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\,dt + F(\bar{\mu}^n)\right]\\
&= \limsup_{n\to\infty} E^n\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\,dt + F([Q^n]_1)\right] \qquad (4.2.31)\\
&\le \frac{1}{2}E_\Theta\left[\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\rho(dy\,dt)\right] + \frac{1}{2\lambda^2}\int_0^T \|\varphi(t)\|^2\,dt + F([\Theta]_1),
\end{aligned}$$

which follows from (4.2.30), the equality $v^n = \varphi$, the weak convergence $Q^n \to \Theta$, and the assumption that $\sqrt{n}\kappa(n) \to \lambda$. This proves (4.2.26) and completes the proof of the lower bound. ☐

4.2.3 Rate Function Property

In this section we sketch the proof of the fact that I_1 defined in (4.1.3) is a rate function. The proof is very similar to the Laplace upper bound and so some details are left to the reader. We will assume Conditions 4.1.1 and 4.1.2 are satisfied. Like with the proof of the upper bound, Condition 4.1.3 is not needed.

Fix $L \in (0,\infty)$, let $\Gamma_L \doteq \{\nu \in \mathcal{P}(\mathcal{X}) : I_1(\nu) \le L\}$, and let $\{\nu_n\}$ be a sequence in Γ_L. We need to show that the sequence has a limit point that lies in Γ_L. From the definition of I_1, we can find, for each n, a $\varphi^n \in L^2([0,T] : \mathbb{R}^k)$ and a $\Theta^n \in \mathcal{E}_1[\varphi^n]$ with $[\Theta^n]_1 = \nu^n$ such that

$$E_{\Theta^n}\left[\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\,\rho(dy\,dt)\right] + \frac{1}{2\lambda^2}\int_0^T \|\varphi^n(t)\|^2\,dt \le L + \frac{1}{n}. \qquad (4.2.32)$$

In particular, $\{\varphi^n\} \subset S_M$ where $M = 2(L+1)\lambda^2$. An argument similar to the proof of Lemma 4.2.2 shows that the sequence (Θ^n, φ^n) is relatively compact in $\mathcal{P}(\mathcal{Z}_1) \times S_M$. Suppose that $(\Theta^n, \varphi^n) \to (\Theta, \varphi)$ along some subsequence.

Then (along the subsequence) $\nu^n \to \nu \doteq [\Theta]_1$. Sending $n \to \infty$ and using lower semicontinuity,

$$E_\Theta \left[\frac{1}{2} \int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \rho(dy\,dt) \right] + \frac{1}{2\lambda^2} \int_0^T \|\varphi(t)\|^2 \, dt \le L.$$

Furthermore, since $\Theta^n \in \mathcal{E}_1[\varphi^n]$, $\Phi(\Theta^n, \varphi^n) = 0$ for each n, where Φ is as in (4.2.20). As shown in Lemma 4.2.3, for each $B < \infty$, $\Phi_B(\Theta^n, \varphi^n) \to \Phi_B(\Theta, \varphi)$. Also a similar argument as in (4.2.24) shows that, as $B \to \infty$,

$$\sup_{n \in \mathbb{N}} |\Phi_B(\Theta^n, \varphi^n) - \Phi(\Theta^n, \varphi^n)| \to 0, \ |\Phi_B(\Theta, \varphi) - \Phi(\Theta, \varphi)| \to 0.$$

It then follows that $\Phi(\Theta, \varphi) = 0$, proving that $\Theta \in \mathcal{E}_1[\varphi]$. Thus, since $\nu = [\Theta]_1$,

$$I_1(\nu) \le E_\Theta \left[\frac{1}{2} \int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \rho(dy\,dt) \right] + \frac{1}{2\lambda^2} \int_0^T \|\varphi(t)\|^2 \, dt \le L.$$

The result follows. □

4.3 Proof of Theorem 4.1.2.

In this section we prove Theorem 4.1.2. Proof of part 1 follows by standard arguments and is therefore left to the Appendix. Proof of part 2 follows similar steps as that for Theorem 4.1.1. Namely, we prove the Laplace upper and lower bounds and show that the function I_2 introduced in (4.1.7) is a rate function. The upper bound is established in Section 4.3.1 while the lower bound is given in Section 4.3.2. The rate function property is verified in Section 4.3.3.

For $(u^n, v^n) \in \mathcal{A}^{1,n} \times \mathcal{A}^2_M$, we consider the following system of controlled SDEs:

$$\begin{aligned} d\bar{X}_i^n(t) &= b(\bar{X}_i^n(t), \bar{\mu}^n(t))\, dt + \sigma(\bar{X}_i^n(t), \bar{\mu}^n(t)) u_i^n(t)\, dt + \alpha(\bar{X}_i^n(t), \bar{\mu}^n(t)) v^n(t)\, dt \\ &\quad + \sigma(\bar{X}_i^n(t), \bar{\mu}^n(t))\, dW_i(t) + \kappa(n) \alpha(\bar{X}_i^n(t), \bar{\mu}^n(t))\, dB(t), \\ d\bar{A}_i^n(t) &= \bar{A}_i^n(t) c(\bar{X}_i^n(t), \bar{\mu}^n(t))\, dt + \bar{A}_i^n(t) \gamma^T(\bar{X}_i^n(t), \bar{\mu}^n(t)) u_i^n(t)\, dt + \bar{A}_i^n(t) \beta^T(\bar{X}_i^n(t), \bar{\mu}^n(t)) v^n(t)\, dt \\ &\quad + \bar{A}_i^n(t) \gamma^T(\bar{X}_i^n(t), \bar{\mu}^n(t))\, dW_i(t) + \kappa(n) \bar{A}_i^n(t) \beta^T(\bar{X}_i^n(t), \bar{\mu}^n(t))\, dB(t), \\ \bar{X}_i^n(0) &= x_i^n, \quad \bar{A}_i^n(0) = a_i^n, \quad 1 \le i \le n, \end{aligned} \tag{4.3.1}$$

where $\bar{\mu}^n(t)$ is the weighted empirical measure

$$\bar{\mu}^n(t) = \frac{1}{n} \sum_{i=1}^n \theta(\bar{A}_i^n(t)) \delta_{\bar{X}_i^n(t)}. \tag{4.3.2}$$

The existence and uniqueness of strong solutions of the above system of equations is argued in the same way as for the uncontrolled system in (4.0.4) (see Appendix 4.6).

The following representation follows along the lines of Theorem 4.2.1. Let $\mathcal{K} \doteq \mathcal{C}([0,T] : \mathcal{M}_+(\mathbb{R}^d))$.

Theorem 4.3.1. *For any $F \in \mathcal{C}_b(\mathcal{K})$ and for each $n \in \mathbb{N}$,*

$$-\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] = \inf_{(u,v)\in\mathcal{A}^{1,n}\times\mathcal{A}^2} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i(t)\|^2\,dt + \frac{1}{2}\int_0^T \|v(t)\|^2\,dt + F(\bar{\mu}^n)\right]. \tag{4.3.3}$$

Furthermore, for every $\delta > 0$, there is an $M < \infty$ such that for each $n \in \mathbb{N}$,

$$-\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] \geq \inf_{(u,v)\in\mathcal{A}^{1,n}\times\mathcal{A}^2_M} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i(t)\|^2\,dt + \frac{1}{2}\int_0^T \|v(t)\|^2\,dt + F(\bar{\mu}^n)\right] - \delta. \tag{4.3.4}$$

4.3.1 Laplace Upper Bound

In this section we show that for every $F \in \mathcal{C}_b(\mathcal{K})$

$$\liminf_{n\to\infty} -\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] \geq \inf_{\nu\in\mathcal{K}} \left[F(\nu) + I_2(\nu)\right], \tag{4.3.5}$$

where I_2 is as in (4.1.7). Throughout the section we assume that Conditions 4.1.1, 4.1.2, 4.1.4, 4.1.5, and 4.1.6 are satisfied. We will not make use of Conditions 4.1.3 and 4.1.7 for the upper bound proof.

Fix $F \in \mathcal{C}_b(\mathcal{K})$ and $\delta \in (0,1)$. From Theorem 4.3.1, there is an $M < \infty$ and, for each $n \in \mathbb{N}$, $(u^n, v^n) \in \mathcal{A}^{1,n} \times \mathcal{A}^2_M$ such that

$$-\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] \geq E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\,dt + F(\bar{\mu}^n)\right] - \delta. \tag{4.3.6}$$

We will next show that

$$\begin{aligned}&\liminf_{n\to\infty} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\,dt + F(\bar{\mu}^n)\right]\\ &\geq \inf_{\varphi\in L^2([0,T]:\mathbb{R}^k)}\inf_{\Theta\in\mathcal{E}_2[\varphi]}\left(E_\Theta\left[\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\,\rho(dy\,dt)\right] + \frac{1}{2\lambda^2}\int_0^T \|\varphi(t)\|^2\,dt + F(\nu_\Theta)\right),\end{aligned} \tag{4.3.7}$$

where ν_Θ is as in (4.1.6). Since $\delta \in (0,1)$ is arbitrary, the desired bound in (4.3.5) is immediate from the above inequality on recalling the definition of I_2 in (4.1.7). In the rest of this section we prove (4.3.7).

We begin by observing that from (4.3.6) we have, as in Section 4.2.1, that (4.2.8) is satisfied. The next two lemmas are analogues of Lemmas 4.2.2 and 4.2.3. In Lemma 4.3.2 below, the result under Condition (ii) in (4.3.10) will be used for the proof of the LLN sketched in the Appendix.

Lemma 4.3.2. *Suppose for some $M \in (0,\infty)$, $\{(u^n, v^n)\}_{n\in\mathbb{N}}$ is a sequence with $(u^n, v^n) \in \mathcal{A}^{1,n} \times \mathcal{A}^2_M$ for each n, and suppose $\{u^n\}_{n\in\mathbb{N}}$ satisfies, for some $L \in (0,\infty)$,*

$$\sup_{n\in\mathbb{N}} E\left[\frac{1}{n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\, dt\right] \le L. \tag{4.3.8}$$

Define the $\mathcal{P}(\mathcal{Z}_2)$-valued random variable Q^n as

$$Q^n(A\times R\times C) = \frac{1}{n}\sum_{i=1}^n \delta_{(\bar{X}_i^n, \bar{A}_i^n)}(A)\delta_{\rho_i^n}(R)\delta_{W_i}(C),\ A\times R\times C \in \mathcal{B}(\mathcal{Z}_2), \tag{4.3.9}$$

where ρ_i^n is as in (4.2.11). *Suppose that*

$$\text{either (i) } \gamma\equiv 0,\ \text{ or (ii) } u_i^n \equiv 0 \text{ for all } i,n,\ \text{ or (iii) Condition 4.1.5(a) holds.} \tag{4.3.10}$$

Then $\{(Q^n, v^n)\}_{n\in\mathbb{N}}$ is tight as a sequence of $\mathcal{P}(\mathcal{Z}_2)\times S_M$-valued random variables.

Proof. Tightness of $\{v^n\}$ is immediate from the compactness of S_M. The tightness of $[Q^n]_3$ and $[Q^n]_4$ follows as in the proof of Lemma 4.2.2. Finally we show the tightness of $[Q^n]_{1,2}$. If (i) or (ii) in (4.3.10) hold, this tightness follows as the proof of the tightness of $[Q^n]_1$ in Lemma 4.2.2 on recalling Condition 4.1.6, the linear growth property of θ, and using the following estimate instead of (4.2.12):

$$E\left[\sup_{0\le s\le T}\left(\|\bar{X}_i^n(s)\|^2 + (\bar{A}_i^n(s))^2\right)\right] \le c_1\left(1 + \|x_i^n\|^2 + (a_i^n)^2 + E\left[\int_0^T \|u_i^n(s)\|^2\, ds\right]\right). \tag{4.3.11}$$

For case (iii) in (4.3.10), we cannot ensure the above square integrability property. However, one can proceed as follows. By Itô's formula,

$$\begin{aligned}\theta(\bar{A}_i^n(t)) &= \theta(a_i^n) + \int_0^t \theta'(\bar{A}_i^n(s))\bar{A}_i^n(s)\, dD_i^n(s)\\ &\quad + \frac{1}{2}\int_0^t \theta''(\bar{A}_i^n(s))\bar{A}_i^n(s)^2\left(\|\gamma(\bar{X}_i^n(s), \bar{\mu}^n(s))\|^2 + \kappa(n)^2\|\beta(\bar{X}_i^n(s), \bar{\mu}^n(s)\|^2\right)\, ds,\end{aligned}$$

where

$$\begin{aligned}D_i^n(t) &= \int_0^t c(\bar{X}_i^n(s), \bar{\mu}^n(s))\, ds + \int_0^t \gamma^T(\bar{X}_i^n(s), \bar{\mu}^n(s))\, dW_i(s) + \kappa(n)\int_0^t \beta^T(\bar{X}_i^n(s), \bar{\mu}^n(s))\, dB(s)\\ &\quad + \int_0^t \gamma^T(\bar{X}_i^n(s), \bar{\mu}^n(s))u_i^n(s)ds + \int_0^t \beta^T(\bar{X}_i^n(s), \bar{\mu}^n(s))v^n(s)ds.\end{aligned}$$

By the boundedness of the coefficients and using (4.1.4), i.e. $\sup_x |\theta'(x)x| + \sup_x |\theta''(x)x^2| < \infty$, we then have, for some $c_2 \in (0,\infty)$,

$$E\left[\sup_{0\le s\le T}\left(\|\bar{X}_i^n(s)\|^2 + (\theta(\bar{A}_i^n(s)))^2\right)\right] \le c_2\left(1 + \|x_i^n\|^2 + (a_i^n)^2 + E\left[\int_0^T \|u_i^n(s)\|^2\, ds\right]\right). \tag{4.3.12}$$

Using the above integrability, the tightness of $[Q^n]_1$ follows as in the proof of Lemma 4.2.2. In order to show the tightness of $[Q^n]_2$ we will use the fact that the map $\phi(\cdot) \mapsto e^{\phi(\cdot)}$ is a continuous map from $\mathcal{C}([0,T]:\mathbb{R})$ to $\mathcal{C}([0,T]:\mathbb{R}_+)$. With this fact, it suffices to show that the collection $\{\frac{1}{n}\sum_{i=1}^n \delta_{\log \bar{A}_i^n(\cdot)}, n \in \mathbb{N}\}$ is tight as a sequence of $\mathcal{P}(\mathcal{C}([0,T]:\mathbb{R}))$-valued random variables. This tightness follows, once again as in the proof of Lemma 4.2.2, from Condition 4.1.6 and the estimates

$$E\left[\sup_{0\le s\le T}\left|\log(\bar{A}_i^n(s))\right|\right] \le c_3\left(1 + |\log a_i^n| + E\left[\int_0^T \|u_i^n(s)\|^2\, ds\right]\right) \tag{4.3.13}$$

and

$$E\left[\left|\log \bar{A}_i^n(\tau+\varepsilon) - \log \bar{A}_i^n(\tau)\right|^2\right] \le c_3\varepsilon\left(1 + E\left[\int_0^T \|u_i^n(s)\|^2\, ds\right]\right),$$

where τ is a stopping time taking values in $[0, T-\varepsilon]$, and the constant c_3 does not depend on n, i, ε, or the stopping time τ. ☐

Lemma 4.3.3. *Let $\{(u^n, v^n)\}_{n\in\mathbb{N}}$ be as in Lemma 4.3.2. Suppose that one of the conditions in* (4.3.10) *is satisfied. Also suppose that (Q^n, v^n) converges, in distribution, along a subsequence to a $\mathcal{P}(\mathcal{Z}_2)\times S_M$-valued random variable (Q, v). Then $Q \in \mathcal{E}_2[v]$ a.s.*

Proof. Suppose that (Q, v) is given on the probability space $(\Omega^*, \mathcal{F}^*, P^*)$. In a similar manner as in the proof of Lemma 4.2.3 (in particular using (4.3.11) and (4.3.12)) we see that $Q \in \mathcal{P}_2(\mathcal{Z}_2)$ P^*-a.s. We need to show that $Q(\omega)$ is a weak solution to $\mathcal{S}_2[v(\omega), \nu_Q(\omega)]$ for P^*-a.e. $\omega \in \Omega^*$. Note that $[Q^n]_{1,2} \circ (z(0), \varsigma(0))^{-1} \to \eta_0$ weakly, which shows that, for P^*-a.e. ω, under $Q(\omega)$, $(z(0), \varsigma(0))$ has distribution η_0, where (z, ς, r, w) denote the canonical coordinate variables on $\mathcal{Z}_2$.

Thus to prove the result it suffices to show that for every $f \in \mathcal{C}_c^2(\mathbb{R}^m \times \mathbb{R}_+ \times \mathbb{R}^d)$, for a.e. ω, $M_{f,v(\omega)}^{Q(\omega)}$ is a martingale under $Q(\omega)$ with respect to the canonical filtration $\tilde{\mathcal{H}}_t \doteq \sigma\{z(s), \varsigma(s), w(s), r(A\times[0,s]), A \in \mathcal{B}(\mathbb{R}^m), s \le t\}$, $t \in [0,T]$, where for each $\varphi \in L^2([0,T]:\mathbb{R}^k)$ and $\Theta \in \mathcal{P}_2(\mathcal{Z}_2)$, the process $\{M_{f,\varphi}^{\Theta}(t), 0 \le t \le T\}$ is defined on

$(\mathcal{Z}_2, \mathcal{B}(\mathcal{Z}_2), \Theta)$ by

$$M_{f,\varphi}^{\Theta}(t,(z,\varsigma,r,w)) = f(z(t),\varsigma(t),w(t)) - f(z(0),\varsigma(0),0) - \int_0^t \int_{\mathbb{R}^m} \mathcal{L}_s^{\Theta}(f)(z(s),\varsigma(s),y,w(s))\, r_s(dy)\, ds$$
$$- \int_0^t \langle \alpha(z(s),\nu_\Theta(s))\varphi(s), \nabla_x f(z(s),\varsigma(s),w(s))\rangle\, ds$$
$$- \int_0^t \varsigma(s)\beta^T(z(s),\nu_\Theta(s))\varphi(s)\frac{\partial f}{\partial a}(z(s),\varsigma(s),w(s))\, ds, \tag{4.3.14}$$

and where

$$\begin{aligned}
&\mathcal{L}_s^{\Theta}(f)(x,a,y,w)\\
&= \langle b(x,\nu_\Theta(s)) + \sigma(x,\nu_\Theta(s))y, \nabla_x f(x,a,w)\rangle + \left(ac(x,\nu_\Theta(s)) + a\gamma^T(x,\nu_\Theta(s))y\right)\frac{\partial f}{\partial a}(x,a,w)\\
&\quad + \frac{1}{2}\sum_{j,j'=1}^{d}(\sigma\sigma^T)_{jj'}(\nu_\Theta(s))\frac{\partial^2 f}{\partial x_j \partial x_{j'}}(x,a,w) + \frac{1}{2}a^2\|\gamma(x,\nu_\Theta(s))\|^2\frac{\partial^2 f}{\partial a^2}(x,a,w)\\
&\quad + \frac{1}{2}\sum_{j=1}^{d} a(\sigma\gamma)_j(x,\nu_\Theta(s))\frac{\partial^2 f}{\partial x_j \partial a}(x,a,w) + \frac{1}{2}\sum_{j=1}^{m}\frac{\partial^2 f}{\partial w_j^2}(x,a,w)\\
&\quad + \sum_{j=1}^{d}\sum_{j'=1}^{m}\sigma_{jj'}(x,\nu_\Theta(s))\frac{\partial^2 f}{\partial x_j \partial w_{j'}}(x,a,w) + \sum_{j=1}^{m} a\gamma_j(x,\nu_\Theta(s))\frac{\partial^2 f}{\partial a \partial w_j}(x,a,w),
\end{aligned} \tag{4.3.15}$$

for $(x,a,y,w) \in \mathbb{R}^d \times \mathbb{R}_+ \times \mathbb{R}^m \times \mathbb{R}^m$.

In order to prove the martingale property, as previously, it suffices to show that for any time instants $0 \le t_0 < t_1 \le T$, and any $\Psi \in \mathcal{C}_b(\mathcal{Z}_2)$ that is measurable with respect to the sigma field $\tilde{\mathcal{H}}_{t_0}$, we have,

$$E_{Q(\omega)}\left[\Psi\left(M_{f,v(\omega)}^{Q(\omega)}(t_1) - M_{f,v(\omega)}^{Q(\omega)}(t_0)\right)\right] = 0, \text{ for } P^*\text{-a.e. } \omega \in \Omega^*. \tag{4.3.16}$$

We suppress ω in the notation of the remaining proof. Fix a choice of (t_0, t_1, Ψ, f) and define $\Phi : \mathcal{P}_2(\mathcal{Z}_2) \times S_M \to \mathbb{R}$ by

$$\Phi(\Theta,\varphi) = E_\Theta\left[\Psi\left(M_{f,\varphi}^{\Theta}(t_1) - M_{f,\varphi}^{\Theta}(t_0)\right)\right]. \tag{4.3.17}$$

Fix $B \in (0,\infty)$. For $\Theta \in \mathcal{P}_2(\mathcal{Z}_2)$, define $\mathcal{L}_s^{\Theta,B}$ by replacing y on the right side of (4.3.15) by $\zeta_B(y)$ and ν_Θ by ν_Θ^B, where ζ_B is as in the proof of Lemma 4.2.3 and $\nu_\Theta^B \in \mathcal{K}$ is defined as

$$\langle f, \nu_\Theta^B(t)\rangle \doteq E_\Theta\left[(\theta(\varsigma(t)) \wedge B) f(z(t))\right], \ t \in [0,T], \ f \in \mathcal{C}_b(\mathbb{R}^d). \tag{4.3.18}$$

Similarly define $M_{f,\varphi}^{\Theta,B}$ by replacing $\mathcal{L}_s^{\Theta}$ with $\mathcal{L}_s^{\Theta,B}$ and ν_Θ with ν_Θ^B in (4.3.14). Finally, define Φ_B by replacing $M_{f,\varphi}^{\Theta}$ with $M_{f,\varphi}^{\Theta,B}$ on the right side of (4.3.17). Then, as before, we will argue (a) for every $B \in (0,\infty)$, Φ_B is a bounded and continuous map on $\mathcal{P}_2(\mathcal{Z}_2)\times S_M$, (b) $\sup_n E[|\Phi_B(Q^n,v^n)-\Phi(Q^n,v^n)|\wedge 1] \to 0$ and $E^*[|\Phi_B(Q,v)-\Phi(Q,v)|\wedge 1] \to 0$ as $B\to\infty$, (c) $\Phi(Q^n,v^n)\to 0$ in probability as $n\to\infty$. The statement in (4.3.16) is immediate from (a)-(c).

Part (a) is shown exactly as in the proof of Lemma 4.2.3. Next consider (b). Using the Lipschitz property of the coefficients, for some $c_1\in(0,\infty)$ and all $n\in\mathbb{N}$,

$$\begin{aligned}|\Phi_B(Q^n,v^n)-\Phi(Q^n,v^n)| \le c_1 \sup_{0\le t\le T} d_{BL}\left(\nu_{Q^n}(t),\nu^B_{Q^n}(t)\right)\left(1+\frac{1}{n}\sum_{i=1}^n\int_0^T \|u_i^n(s)\|\,ds\right)\\ + \frac{c_1}{n}\sum_{i=1}^n\int_0^T \|u_i^n(s)-\zeta_B(u_i^n(s))\|\,ds.\end{aligned} \tag{4.3.19}$$

Also,

$$\sup_{0\le t\le T} d_{BL}\left(\nu_{Q^n}(t),\nu^B_{Q^n}(t)\right) \le \frac{1}{n}\sum_{i=1}^n \sup_{0\le t\le T}\theta(\bar{A}_i^n(t))1_{\left\{\sup_{0\le t\le T}\theta(\bar{A}_i^n(t))>B\right\}} \le \frac{1}{nB}\sum_{i=1}^n\sup_{0\le t\le T}\left[\theta(\bar{A}_i^n(t))\right]^2. \tag{4.3.20}$$

Combining this with the bounds in (4.3.11), (4.3.12), we have, for some $c_2\in(0,\infty)$ and every $B<\infty$,

$$\sup_{n\in\mathbb{N}} E\left[\sup_{0\le t\le T} d_{BL}\left(\nu_{Q^n}(t),\nu^B_{Q^n}(t)\right)\right] \le \frac{c_2}{B}.$$

Fix $\varepsilon\in(0,1)$ and using (4.3.8) choose $m_1\in(0,\infty)$ such that

$$\sup_{n\in\mathbb{N}} P\left(\frac{1}{n}\sum_{i=1}^n\int_0^T \|u_i^n(s)\|ds > m_1\right) < \varepsilon.$$

Then using the inequality $E[(UV)\wedge 1]\le P(V>m_1+1)+(m_1+1)E[U]$ for non-negative random variables U and V, we have

$$E\left[\left\{c_1\sup_{0\le t\le T} d_{BL}\left(\nu_{Q^n}(t),\nu^B_{Q^n}(t)\right)\left(1+\frac{1}{n}\sum_{i=1}^n\int_0^T\|u_i^n(s)\|ds\right)\right\}\wedge 1\right] \le \varepsilon + \frac{(m_1+1)c_1c_2}{B}.$$

Using this estimate in (4.3.19), for some $c_3\in(0,\infty)$,

$$\sup_{n\in\mathbb{N}} E\left[|\Phi_B(Q^n,v^n)-\Phi(Q^n,v^n)|\wedge 1\right] \le \frac{c_3(1+m_1)}{B}+\varepsilon.$$

Sending $B\to\infty$ and since ε is arbitrary, we have the first statement in (b). The second statement in (b) follows in a similar manner on noting the properties

$$E^*\left[E_Q\left[\sup_{0\le t\le T}\theta(\varsigma(t))^2\right]\right]<\infty, \quad E^*\left[E_Q\left[\int_0^T\left\|\int_{\mathbb{R}^m} y\, r_s(dy)\right\|^2 ds\right]\right]<\infty,$$

which follow from analogous (uniform in n) bounds when Q is replaced by Q^n and E^* by E.

Finally we consider (c). For each $n \in \mathbb{N}$,

$$\begin{aligned}
&\Phi(Q^n, v^n) \\
&= E_{Q^n}\left[\Psi\left(M^{Q^n}_{f,v^n}(t_1) - M^{Q^n}_{f,v^n}(t_0)\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\Psi(\bar{X}^n_i, \bar{A}^n_i, \rho^n_i, W_i)\left(M^{Q^n}_{f,v^n}(t_1, (\bar{X}^n_i, \bar{A}^n_i, \rho^n_i, W_i)) - M^{Q^n}_{f,v^n}(t_0, (\bar{X}^n_i, \bar{A}^n_i, \rho^n_i, W_i))\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\Psi(\bar{X}^n_i, \bar{A}^n_i, \rho^n_i, W_i)\left(f(\bar{X}^n_i(t_1), \bar{A}^n_i(t_1), W_i(t_1)) - f(\bar{X}^n_i(t_0), \bar{A}^n_i(t_0), W_i(t_0)) - \int_{t_0}^{t_1}\mathcal{U}^n(s)\,ds\right),
\end{aligned}$$

where, noting that $\nu_{Q^n}(s) = \bar{\mu}^n(s)$,

$$\begin{aligned}
\mathcal{U}^n(s) \doteq{}& \mathcal{L}^{Q^n}_s(f)(\bar{X}^n_i(s), \bar{A}^n_i(s), u^n_i(s), W_i(s)) + [\nabla_x f(\bar{X}^n_i(s), \bar{A}^n_i(s), W_i(s))]^T\alpha(\bar{X}^n_i(s), \bar{\mu}^n(s))v^n(s) \\
&+ \bar{A}^n_i(s)\beta^T(\bar{X}^n_i(s), \bar{\mu}^n(s))v^n(s)\frac{\partial f}{\partial a}(\bar{X}^n_i(s), \bar{A}^n_i(s), W_i(s)).
\end{aligned}$$

By Itô's formula, for each i and n, we have a.s. that

$$\begin{aligned}
&f(\bar{X}^n_i(t_1), \bar{A}^n_i(t_1), W_i(t_1)) - f(\bar{X}^n_i(t_0), \bar{A}^n_i(t_0), W_i(t_0)) \\
&= \int_{t_0}^{t_1}\mathcal{U}^n(s)\,ds + \int_{t_0}^{t_1}[\nabla_x f(\bar{X}^n_i(s), \bar{A}^n_i(s), W_i(s))]^T\sigma(\bar{X}^n_i(s), \bar{\mu}^n(s))\,dW_i(s) \\
&\quad + \kappa(n)\int_{t_0}^{t_1}[\nabla_x f(\bar{X}^n_i(s), \bar{A}^n_i(s), W_i(s))]^T\alpha(\bar{X}^n_i(s), \bar{\mu}^n(s))\,dB(s) \\
&\quad + \int_{t_0}^{t_1}\frac{\partial f}{\partial a}(\bar{X}^n_i(s), \bar{A}^n_i(s), W_i(s))\bar{A}^n_i(s)\gamma^T(\bar{X}^n_i(s), \bar{\mu}^n(s))\,dW_i(s) \\
&\quad + \kappa(n)\int_{t_0}^{t_1}\frac{\partial f}{\partial a}(\bar{X}^n_i(s), \bar{A}^n_i(s), W_i(s))\bar{A}^n_i(s)\beta^T(\bar{X}^n_i(s), \bar{\mu}^n(s)\,dB(s) \\
&\quad + \int_{t_0}^{t_1}[\nabla_w f(\bar{X}^n_i(s), \bar{A}^n_i(s), W_i(s))]^T\,dW_i(s) + \mathcal{T}^n_i,
\end{aligned}$$

where, for some $c_1 \in (0,\infty)$, $|\mathcal{T}^n_i| \le c_1\kappa(n)^2$ for all n, i. Letting $\Psi^n_i = \Psi(\bar{X}^n_i, \bar{A}^n_i, \rho^n_i, W_i)$ and $f^n_i(s) = f(\bar{X}^n_i(s), \bar{A}^n_i(s), W_i(s))$, and using similar notation for derivatives of f, we have a.s. that

$$\begin{aligned}
&\Phi(Q^n, v^n) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\Psi^n_i\left[\int_{t_0}^{t_1}[\nabla_x f^n_i(s)]^T\sigma(\bar{X}^n_i(s), \bar{\mu}^n(s))\,dW_i(s) + \int_{t_0}^{t_1}\frac{\partial f^n_i(s)}{\partial a}\bar{A}^n_i(s)\gamma^T(\bar{X}^n_i(s), \bar{\mu}^n(s))\,dW_i(s)\right. \\
&\qquad \left. + \int_{t_0}^{t_1}[\nabla_w f^n_i(s)]^T\,dW_i(s)\right] + \mathcal{T}^n,
\end{aligned}$$

where, as in the proof of Lemma 4.2.3, $\mathcal{T}^n \to 0$ in probability. Now by the same argument as in Lemma 4.2.3, $\Phi(Q^n, v^n) \to 0$ in probability, proving (c). Thus we have $\Phi(Q, v) = 0$ a.s., which proves (4.3.16) and completes the proof. $\square$

We now complete the proof of (4.3.7). In addition to the standing assumptions of this section (namely Conditions 4.1.1, 4.1.2, 4.1.4, 4.1.5 and 4.1.6) suppose that if Condition 4.1.5(a) is not satisfied then $\gamma = 0$.

Since $\bar{\mu}^n = \nu_{Q^n}$, we have

$$\begin{aligned}
&E\left[\frac{1}{2n}\sum_{i=1}^{n}\int_0^T \|u_i^n(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\,dt + F(\bar{\mu}^n)\right]\\
&= E\left[\int_{\mathcal{R}_1}\left(\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]}\|y\|^2\,r(dy\,dt)\right)[Q^n]_2(dr) + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\,dt + F(\nu_{Q^n})\right],
\end{aligned}$$

where $[Q^n]_2$ denotes the second marginal of Q^n. Recalling the bound (4.2.8), we have from Lemmas 4.3.2 and 4.3.3 that, (Q^n, v^n) is tight and if (Q, v) is a weak limit point then $Q \in \mathcal{E}_2[v]$ a.s. Assume without loss of generality that $(Q^n, v^n) \to (Q, v)$ along the full sequence. We claim that $(Q^n, v^n, \nu_{Q^n}) \to (Q, v, \nu_Q)$, in distribution, in $\mathcal{P}(\mathcal{Z}_2) \times S_M \times \mathcal{K}$. For $\Theta \in \mathcal{P}(\mathcal{Z}_2)$ and $B \in (0, \infty)$, define $\nu_\Theta^B \in \mathcal{K}$ as in (4.3.18), i.e.

$$\nu_\Theta^B(t)(C) \doteq E_\Theta\left[(\theta(\varsigma(t)) \wedge B)\, 1_C(z(t))\right], \quad C \in \mathcal{B}(\mathbb{R}^d).$$

Then it is easy to check that, since $\theta(\cdot) \wedge B$ is a bounded Lipschitz function, $\Theta \mapsto \nu_\Theta^B$ is a continuous map from $\mathcal{P}(\mathcal{Z}_2)$ to $\mathcal{K}$ for every B. Also, from (4.3.20), for some $c_1 \in (0, \infty)$,

$$\sup_{n\in\mathbb{N}} E\left[\sup_{0\le t\le T} d_{BL}\left(\nu_{Q^n}(t), \nu_{Q^n}^B(t)\right)\right] \le \frac{c_1}{B}\sup_{n\in\mathbb{N}} E\left[\frac{1}{n}\sum_{i=1}^{n}\sup_{0\le t\le T}\left(\theta(\bar{A}_i^n(t))\right)^2\right] \to 0$$

as $B \to \infty$, since $\sup_{n\in\mathbb{N}} E[\frac{1}{n}\sum_{i=1}^n \sup_{0\le t\le T}(\theta(\bar{A}_i^n(t)))^2] < \infty$, which follows from (4.3.11) and linear growth of θ when property (i) of (4.3.10) holds and from (4.3.12) when property (iii) in (4.3.10) is satisfied. Combining the above uniform convergence with the fact that $(Q^n, v^n, \nu_{Q^n}^B) \Rightarrow (Q, v, \nu_Q^B)$ for every B proves the claim.

Finally by Fatou's lemma and since $\sqrt{n}\kappa(n) \to \lambda$,

$$\begin{aligned}
&\liminf_{n\to\infty} E\left[\frac{1}{2n}\sum_{i=1}^{n}\int_0^T \|u_i^n(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\,dt + F(\bar{\mu}^n)\right]\\
&\ge E\left[\int_{\mathcal{R}_1}\left(\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]}\|y\|^2\,r(dy\,dt)\right)[Q]_2(dr) + \frac{1}{2\lambda^2}\int_0^T \|v(t)\|^2\,dt + F(\nu_Q)\right]\\
&\ge \inf_{\varphi\in L^2([0,T]:\mathbb{R}^k)}\ \inf_{\Theta\in\mathcal{E}_2[\varphi]}\left(E_\Theta\left[\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]}\|y\|^2\,\rho(dy\,dt)\right] + \frac{1}{2\lambda^2}\int_0^T \|\varphi(t)\|^2\,dt + F(\nu_\Theta)\right).
\end{aligned}$$

This proves (4.3.7) and completes the proof of the Laplace upper bound. $\square$

4.3.2 Laplace Lower Bound

Throughout this section we assume that Conditions 4.1.1-4.1.7 are satisfied. Additionally we assume that if Condition 4.1.5(a) does not hold then $\gamma = 0$. We will proceed as in Section 4.2.2.

Fix $\varepsilon > 0$ and $F \in \mathcal{C}_b(\mathcal{K})$. Choose a $\varphi \in L^2([0,T] : \mathbb{R}^k)$ and a $\Theta \in \mathcal{E}_2[\varphi]$ such that

$$\frac{1}{2}E_\Theta\left[\int_{\mathbb{R}^m\times[0,T]}\|y\|^2\rho(dy\,dt)\right] + \frac{1}{2\lambda^2}\int_0^T\|\varphi(t)\|^2\,dt + F(\nu_\Theta) \le \inf_{\nu\in\mathcal{K}}[F(\nu) + I_2(\nu)] + \varepsilon.$$

We will show that there is a $M \in (0,\infty)$ and a sequence (u^n, v^n) with $u^n \in \mathcal{A}^{1,n}$ and $v^n \in \mathcal{A}^2_M$ constructed on some filtered probability space such that

$$\begin{aligned}&\limsup_{n\to\infty} E\left[\frac{1}{2n}\sum_{i=1}^n\int_0^T\|u_i^n(t)\|^2\,dt + \frac{1}{2n\kappa(n)^2}\int_0^T\|v^n(t)\|^2\,dt + F(\bar\mu^n)\right]\\ &\le \frac{1}{2}E_\Theta\left[\int_{\mathbb{R}^m\times[0,T]}\|y\|^2\rho(dy\,dt)\right] + \frac{1}{2\lambda^2}\int_0^T\|\varphi(t)\|^2\,dt + F(\nu_\Theta).\end{aligned} \tag{4.3.21}$$

The Laplace lower bound

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right] \le \inf_{\nu\in\mathcal{K}}[F(\nu) + I_2(\nu)]$$

is then immediate from Theorem 4.3.1 on noting that $\varepsilon > 0$ is arbitrary. We begin with the following uniqueness result. Analogous to Section 4.2.2, define the map $\vartheta : \mathcal{Z}_2 \to \mathcal{Z}_2^\vartheta \doteq \mathbb{R}^d \times \mathbb{R}_+ \times \mathcal{R}_1 \times \mathcal{W}$ as $\vartheta(z,\varsigma,r,w) \doteq (z(0),\varsigma(0),r,w)$. For $\Theta \in \mathcal{P}(\mathcal{Z}_2)$, let $\Theta_\vartheta \doteq \Theta\circ\vartheta^{-1}$ be the probability measure on $\mathcal{Z}_2^\vartheta$ induced by Θ under ϑ. We will say that *weak uniqueness* holds for (4.1.5) if, for any given $\varphi \in L^2([0,T] : \mathbb{R}^k)$ and $\Theta^{(1)}, \Theta^{(2)} \in \mathcal{E}_2[\varphi]$, whenever $\Theta^{(1)}_\vartheta = \Theta^{(2)}_\vartheta$, we have that $\Theta^{(1)} = \Theta^{(2)}$.

Lemma 4.3.4. *Weak uniqueness holds for* (4.1.5).

Proof. Fix $\varphi \in L^2([0,T] : \mathbb{R}^k)$ and $\Theta^{(1)}, \Theta^{(2)} \in \mathcal{E}_2[\varphi]$. Suppose that $\Theta^{(1)}_\vartheta = \Theta^{(2)}_\vartheta \doteq \Lambda$. Note that $\Theta^{(i)}$, $i = 1,2$ can be disintegrated as

$$\Theta^{(i)}(dx,\,da,\,dr,\,dw) = \tilde\Theta^{(i)}(dx,\,da,\,x_0,\,a_0,\,r,\,w)\Lambda(dx_0,\,da_0,\,dr,\,dw).$$

Consider $\hat{\mathcal{Z}}_2 = \mathcal{X}\times\mathcal{Y}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}_2^\vartheta$, and define $\hat\Theta \in \mathcal{P}(\hat{\mathcal{Z}}_2)$ as

$$\begin{aligned}&\hat\Theta(dx^{(1)},\,da^{(1)},\,dx^{(2)},\,da^{(2)},\,dx_0,\,da_0,\,dr,\,dw)\\ &\doteq \tilde\Theta^{(1)}(dx^{(1)},\,da^{(1)},\,x_0,\,a_0,\,r,\,w)\,\tilde\Theta^{(2)}(dx^{(2)},\,da^{(2)},\,x_0,\,a_0,\,r,\,w)\,\Lambda(dx_0,\,da_0,\,dr,\,dw),\end{aligned}$$

and denote the coordinate maps on $\hat{\mathcal{Z}}_2$ as $(X^{(1)}, A^{(1)}, X^{(2)}, A^{(2)}, X_0, A_0, \rho, W)$. Note that $\{W(t), t \in [0,T]\}$ is a Brownian motion with respect to the canonical filtration

$$\hat{\mathcal{H}}_t \doteq \sigma\left\{X^{(i)}(s), A^{(i)}(s), \rho(A \times [0,s]), W(s),\ i=1,2,\ A \in \mathcal{B}(\mathbb{R}^m),\ s \in [0,t]\right\}, \quad t \in [0,T],$$

and for $i = 1,2$, $(X^{(i)}, A^{(i)})$ satisfy (4.1.5) with $(\bar{X}, \bar{A})$ replaced with $(X^{(i)}, A^{(i)})$ and $\nu(t)$ replaced with $\nu^{(i)}(t) \doteq \nu_{\Theta^{(i)}}(t)$. In order to prove the lemma it suffices to show that $(X^{(1)}, A^{(1)}) = (X^{(2)}, A^{(2)})$ a.s. Let $u(t) \doteq \int_{\mathbb{R}^m} y\, \rho_t(dy)$, $t \in [0,T]$. Then, $E_{\hat{\Theta}} \int_0^T \|u(t)\|^2\, dt < \infty$. By similar estimates as in the proof of Lemma 4.3.2 we see that

$$\text{when Condition 4.1.5(a) is satisfied, } E_{\hat{\Theta}}\left[\sup_{0 \le s \le T}\left(\|X^{(i)}(s)\|^2 + (\theta(A^{(i)}(s)))^2\right)\right] < \infty \text{ for } i = 1,2, \tag{4.3.22}$$

and

$$\text{when } \gamma = 0,\ E_{\hat{\Theta}}\left[\sup_{0 \le s \le T}\left(\|X^{(i)}(s)\|^2 + (A^{(i)}(s))^2\right)\right] < \infty \text{ for } i = 1,2. \tag{4.3.23}$$

Consider first the case $\gamma = 0$. For $t \in [0,T]$, define

$$g(t) = E_{\hat{\Theta}}\left[\sup_{0 \le s \le t} \|X^{(1)}(s) - X^{(2)}(s)\|^2\right], \quad h(t) = \left(E_{\hat{\Theta}}\left[\sup_{0 \le s \le t} |A^{(1)}(s) - A^{(2)}(s)|\right]\right)^2.$$

Since θ is a Lipschitz function under Condition 4.1.5, we have

$$\begin{aligned} d_{BL}\left(\nu^{(1)}(s), \nu^{(2)}(s)\right) &\le \sup_{f \in BL(\mathbb{R}^d)} E_{\hat{\Theta}}\left|\theta(A^{(1)}(s)) f(X^{(1)}(s)) - \theta(A^{(2)}(s)) f(X^{(2)}(s))\right| \\ &\le E_{\hat{\Theta}}\left[\theta(A^{(1)}(s))\|X^{(1)}(s) - X^{(2)}(s)\|\right] + L E_{\hat{\Theta}}\left|A^{(1)}(s) - A^{(2)}(s)\right|, \end{aligned}$$

where L is the Lipschitz constant for θ. Then by the Cauchy-Schwarz inequality and (4.3.23), for some $c_1 \in (0,\infty)$,

$$\sup_{0 \le s \le t} d_{BL}\left(\nu^{(1)}(s), \nu^{(2)}(s)\right)^2 \le c_1(g(t) + h(t)) \text{ for all } t \in [0,T]. \tag{4.3.24}$$

By the Lipschitz properties of b, σ, and α, the property $\sigma(x,\nu) = \sigma(\nu)$, the Burkholder-Davis-Gundy and Cauchy-Schwarz inequalities, and the fact that $\int_0^T \|\varphi(s)\|^2\, ds < \infty$, there are $c_2, c_3 \in (0,\infty)$ such that, for all $t \in [0,T]$,

$$\begin{aligned} g(t) &\le c_2 E_{\hat{\Theta}}\left[\int_0^t \sup_{0 \le \tau \le s}\left(\|X^{(1)}(\tau) - X^{(2)}(\tau)\|^2 + d_{BL}\left(\nu^{(1)}(\tau), \nu^{(2)}(\tau)\right)^2\right) ds\right] \\ &\quad + c_2 E_{\hat{\Theta}}\left[\left(\int_0^t \sup_{0 \le \tau \le s} d_{BL}\left(\nu^{(1)}(\tau), \nu^{(2)}(\tau)\right) \cdot \|u(s)\|\, ds\right)^2\right] \\ &\le c_3 \int_0^t (g(s) + h(s))\, ds. \end{aligned} \tag{4.3.25}$$

Furthermore, since $\gamma = 0$, for $j = 1, 2$, $A^{(j)}(t) = e^{Y^{(j)}(t)}$, where

$$Y^{(j)}(t) = Y^{(j)}(0) + \int_0^t c(X^{(j)}(s), \nu^{(j)}(s))\, ds + \int_0^t \beta^T(X^{(j)}(s), \nu^{(j)}(s))\varphi(s)\, ds.$$

Using the inequality $|e^x - e^y| \le (e^x \vee e^y)|x - y|$, the Lipschitz property of c and β, (4.3.24) and (4.3.23), and the Cauchy-Schwarz inequality, there is $c_4 \in (0, \infty)$ such that

$$\begin{aligned} h(t) &\le \left(E_{\hat{\Theta}} \left[\sup_{0\le s\le t} (A^{(1)}(s) \vee A^{(2)}(s)) |Y^{(1)}(s) - Y^{(2)}(s)| \right] \right)^2 \\ &\le E_{\hat{\Theta}} \left[\sup_{0\le s\le t} (A^{(1)}(s) \vee A^{(2)}(s))^2 \right] E_{\hat{\Theta}} \left[\sup_{0\le s\le t} |Y^{(1)}(s) - Y^{(2)}(s)|^2 \right] \\ &\le c_4 \int_0^t (g(s) + h(s))\, ds. \end{aligned} \tag{4.3.26}$$

Thus,

$$g(t) + h(t) \le (c_3 + c_4) \int_0^t (g(s) + h(s))\, ds \quad \text{for every } t \in [0, T],$$

and hence by Gronwall's inequality, $g(T) + h(T) = 0$, from which it follows that $(X^{(1)}, A^{(1)})$ and $(X^{(2)}, A^{(2)})$ are indistinguishable on $[0, T]$.

Consider now the case where Condition 4.1.5(a) is satisfied. Define

$$\tilde{h}(t) = \left(E_{\hat{\Theta}} \left[\sup_{0\le s\le t} |\log A^{(1)}(s) - \log A^{(2)}(s)| \right] \right)^2.$$

Since $c_5 \doteq \sup_{x\in\mathbb{R}_+} |\theta'(x)x| < \infty$, we have

$$|\theta(e^x) - \theta(e^y)| \le \sup_{z\in\mathbb{R}_+} |\theta'(z)z| \cdot |x - y| = c_5|x - y| \quad \text{for all } x, y \in \mathbb{R}.$$

Thus,

$$\begin{aligned} d_{BL}\left(\nu^{(1)}(s), \nu^{(2)}(s)\right) &\le \sup_{f\in BL(\mathbb{R}^d)} E_{\hat{\Theta}} \left| \theta(A^{(1)}(s)) f(X^{(1)}(s)) - \theta(A^{(2)}(s)) f(X^{(2)}(s)) \right| \\ &\le E_{\hat{\Theta}} \left[\theta(A^{(1)}(s)) \|X^{(1)}(s) - X^{(2)}(s)\| \right] + c_5 E_{\hat{\Theta}} \left| \log A^{(1)}(s) - \log A^{(2)}(s) \right|. \end{aligned}$$

Hence, using (4.3.22), for some $c_6 \in (0, \infty)$,

$$\sup_{0\le s\le t} d_{BL}\left(\nu^{(1)}(s), \nu^{(2)}(s)\right)^2 \le c_6(g(t) + \tilde{h}(t)) \quad \text{for all } t \in [0, T].$$

Now exactly as in (4.3.25), we have that for some $c_7 \in (0, \infty)$,

$$g(t) \le c_7 \int_0^t (g(s) + \tilde{h}(s))\, ds \ \text{ for every } t \in [0, T].$$

Note that, for some $c_8 \in (0, \infty)$,

$$\left(E_{\hat{\Theta}} \left[\int_0^t \left(\gamma(\nu^{(1)}(s)) - \gamma(\nu^{(2)}(s)) \right)^T u(s)\, ds \right] \right)^2 \le c_8 \int_0^t d_{BL} \left(\nu^{(1)}(s), \nu^{(2)}(s) \right)^2 ds.$$

Using this estimate and Lipschitz properties of c, γ, and β, we now have that, for some $c_9 \in (0, \infty)$

$$\tilde{h}(t) \le c_9 \int_0^t (g(s) + \tilde{h}(s))\, ds \text{ for every } t \in [0, T].$$

Thus

$$g(t) + \tilde{h}(t) \le (c_7 + c_9) \int_0^t (g(s) + \tilde{h}(s))\, ds \text{ for every } t \in [0, T],$$

which shows the indistinguishability of $(X^{(1)}, \log A^{(1)})$ and $(X^{(2)}, \log A^{(2)})$ and hence the indistinguishability of $(X^{(1)}, A^{(1)})$ and $(X^{(1)}, A^{(1)})$ on $[0, T]$. $\square$

We now complete the proof of the lower bound by constructing (u^n, v^n) that satisfy (4.3.21). Let $(\bar{X}, \bar{A}, \rho, W)$ be the coordinate maps on the space $(\mathcal{Z}_2, \mathcal{B}(\mathcal{Z}_2), \Theta)$ equipped with the canonical filtration $\tilde{\mathcal{H}}_t$ defined as in the proof of Lemma 4.3.3. Since $\Theta \in \mathcal{E}_2[\varphi]$, equation (4.1.5) is satisfied with $\nu(t) = \nu_\Theta(t)$. Disintegrate Θ_ϑ as

$$\Theta_\vartheta(dx, da, dr, dw) = \eta_0(dx, da)\, [\Theta]_4(dw)\, \hat{\Lambda}_0(x, a, w, dr).$$

Let $\mathcal{V}, \Omega', \mathcal{F}'$ and coordinate processses W_i, B, ρ_i be as introduced in Section 4.2.2. As before, let Γ be the standard Wiener measure on $\mathcal{V}$. Next, for each $n \in \mathbb{N}$, define the probability measure P^n on $(\Omega', \mathcal{F}')$ by

$$P^n(dr, dw, d\beta) = \left[\bigotimes_{i=1}^n [\Theta]_4(dw_i)\, \hat{\Lambda}_0(x_i^n, a_i^n, w_i, dr_i) \bigotimes_{i=n+1}^\infty [\Theta]_{(3,4)}(dr_i, dw_i) \right] \otimes \Gamma(d\beta).$$

Under P^n, $\{W_i\}_{1 \le i \le n}$ and B are mutually independent Brownian motions. Define the sequence $\{\Lambda^n\}_{n \in \mathbb{N}}$ of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}_+ \times \mathcal{R}_1 \times \mathcal{W})$-valued random variables by

$$\Lambda^n(A \times B \times R \times C) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i^n}(A) \delta_{a_i^n}(B) \delta_{\rho_i}(R) \delta_{W_i}(C), \ A \times B \times R \times C \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}_+ \times \mathcal{R}_1 \times \mathcal{W}).$$

Then by Condition 4.1.6,

$$P^n \circ (\Lambda^n)^{-1} \to \delta_{\Theta_\vartheta}. \tag{4.3.27}$$

Let, for $n \in \mathbb{N}$, $v^n \doteq \varphi$. Then $v^n \in S_M$ for every n, where $M \doteq \int_0^T \|\varphi(s)\|^2 ds$. Next, define u_i by (4.2.29) and for each $n \in \mathbb{N}$ let $(\bar{X}_1^n, \bar{A}_1^n \ldots, \bar{X}_n^n, \bar{A}_n^n)$ be the solution on $(\Omega', \mathcal{F}', P^n)$ of the system in (4.3.1) (with u_i^n replaced with u_i).

Define the occupation measure Q^n by the right side of (4.3.9), replacing ρ_i^n with ρ_i. Let E^n denote expectation over the probability measure P^n. Then, as in (4.2.30) (using Condition 4.1.6 instead of 4.1.1), we see that

$$\limsup_{n\to\infty} E^n \left[\frac{1}{n} \sum_{i=1}^{n} \int_0^T \|u_i(t)\|^2 \, dt \right] = E_\Theta \left[\int_0^T \left\| \int_{\mathbb{R}^m} y \, \rho_t(dy) \right\|^2 dt \right] < \infty. \tag{4.3.28}$$

It now follows from Lemma 4.3.2 that $\{(Q^n, v^n)\}_{n\in\mathbb{N}}$ is tight. If (Q, v) is a limit point of this sequence defined on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, then $v = \varphi$ $\tilde{P}$-a.s., and, by Lemma 4.3.3, $Q \in \mathcal{E}_2[v] = \mathcal{E}_2[\varphi]$ $\tilde{P}$-a.s. Also, $\Theta \in \mathcal{E}_2[\varphi]$. By (4.3.27), for $\tilde{P}$-a.e. $\omega \in \tilde{\Omega}$, $Q_\vartheta(\omega) = \Theta_\vartheta$. Thus by the weak uniqueness established in Lemma 4.3.4, $Q = \Theta$ $\tilde{P}$-a.s. Thus we have $Q^n \to \Theta$ in probability. A similar argument as in Lemma 4.3.3 now shows that $(Q^n, \nu_{Q^n}) \to (\Theta, \nu_\Theta)$ in probability. Finally,

$$\begin{aligned}
&\limsup_{n\to\infty} E^n \left[\frac{1}{2n} \sum_{i=1}^{n} \int_0^T \|u_i(t)\|^2 \, dt + \frac{1}{2n\kappa(n)^2} \int_0^T \|v^n(t)\|^2 \, dt + F(\bar{\mu}^n) \right] \\
&= \limsup_{n\to\infty} E^n \left[\frac{1}{2n} \sum_{i=1}^{n} \int_0^T \|u_i(t)\|^2 \, dt + \frac{1}{2n\kappa(n)^2} \int_0^T \|v^n(t)\|^2 \, dt + F(\nu_{Q^n}) \right] \\
&\le \frac{1}{2} E_\Theta \left[\int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \rho(dy\, dt) \right] + \frac{1}{2\lambda^2} \int_0^T \|\varphi(t)\|^2 \, dt + F(\nu_\Theta),
\end{aligned}$$

which follows from (4.3.28), the equality $v^n = \varphi$, the convergence $(Q^n, \nu_{Q^n}) \to (\Theta, \nu_\Theta)$, and the assumption that $\sqrt{n}\kappa(n) \to \lambda$. This proves (4.3.21) and completes the proof of the lower bound. □

4.3.3 Rate Function Property of I_2

The proof is very similar to the argument in Section 4.2.3 and so we omit the details and note only that we use the argument in Lemma 4.3.3 to show that if for $\Theta^n, \Theta \in \mathcal{P}_2(\mathcal{Z}_2)$, $\Theta^n \to \Theta$, and a bound as in (4.2.32) is satisfied for every n, then under the conditions of Theorem 4.1.2, $\nu_{\Theta^n} \to \nu_\Theta$ in $\mathcal{K}$. □

4.4 Proof Sketch of Theorem 4.1.3.

In Section 4.4.1 we sketch the proof of part (i) of the theorem while part (ii) is sketched in Section 4.4.2.

4.4.1 The Case $\sqrt{n}\kappa(n) \to 0$

Recall that we assume Conditions 4.1.1, 4.1.2, and 4.1.3 hold. For the Laplace upper bound we start with the inequality in (4.2.6) for some $(u^n, v^n) \in \mathcal{A}^{1,n} \times \mathcal{A}^2_M$. This inequality gives the uniform bound in (4.2.8). With this uniform bound, the tightness of the sequence of $\mathcal{P}(\mathcal{Z}_1)$-valued random variables Q^n defined in (4.2.10) is shown as in Lemma 4.2.2.

Furthermore, the inequality in (4.2.8) also shows that

$$E\left[\int_0^T \|v^n(t)\|^2 dt\right] \leq 2n\kappa(n)^2(2\|F\|_\infty + 1) \to 0 \text{ as } n \to \infty,$$

since $n\kappa(n)^2 \to 0$. Thus $v^n \to 0$ in $L^2([0,T] : \mathbb{R}^k)$, in probability.

Now a similar argument as in Lemma 4.2.3 shows that if Q is a weak limit point of Q^n, then $Q \in \mathcal{E}_1[0]$ a.s. Finally, with (u^n, v^n) as above and $\bar{\mu}^n$ defined as below (4.2.3), taking the limit as $n \to \infty$ along any convergent subsequence of $\{(Q^n, v^n)\}$,

$$\begin{aligned}
&\liminf_{n\to\infty} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\, dt + \frac{1}{2n\kappa(n)^2}\int_0^T \|v^n(t)\|^2\, dt + F(\bar{\mu}^n)\right] \\
&\geq \liminf_{n\to\infty} E\left[\frac{1}{2n}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\, dt + F(\bar{\mu}^n)\right] \\
&\geq E\left[\int_{\mathcal{R}_1}\left(\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\, r(dy\, dt)\right) [Q]_2(dr) + F([Q]_1)\right] \\
&\geq \inf_{\Theta\in\mathcal{E}_1[0]}\left(E_\Theta\left[\frac{1}{2}\int_{\mathbb{R}^m\times[0,T]} \|y\|^2\, \rho(dy\, dt)\right] + F([\Theta]_1)\right),
\end{aligned}$$

where the last inequality uses the fact that $Q \in \mathcal{E}_1[0]$ a.s. Since $\delta \in (0,1)$ in (4.2.6) is arbitrary, recalling the definition of $\tilde{I}_{1,0}$ in (4.1.10), the above inequality completes the proof of the Laplace upper bound.

For the proof of the lower bound we proceed as follows. Fix $\varepsilon > 0$ and $F \in \mathcal{C}_b(\mathcal{P}(\mathcal{X}))$. Choose $\Theta \in \mathcal{E}_1[0]$ such that

$$\frac{1}{2}E_\Theta\left[\int_{\mathbb{R}^m\times[0,T]} \|y\|^2 \rho(dy\, dt)\right] + F([\Theta]_1) \leq \inf_{\nu\in\mathcal{P}(\mathcal{X})}\left[F(\nu) + \tilde{I}_{1,0}(\nu)\right] + \varepsilon. \tag{4.4.1}$$

Using this Θ, define $(\Omega', \mathcal{F}', P^n)$, as in Section 4.2.2. Also, take $v^n = 0$ for every n. Then with u_i defined as in (4.2.29) and $\bar{\mu}^n$ and Q^n constructed as below (4.2.29), we have exactly as in (4.2.31) that

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\left[e^{-nF(\mu^n)}\right]$$

$$\leq \limsup_{n\to\infty} E^n \left[\frac{1}{2n} \sum_{i=1}^n \int_0^T \|u_i(t)\|^2 \, dt + \frac{1}{2n\kappa(n)^2} \int_0^T \|v^n(t)\|^2 \, dt + F(\bar{\mu}^n) \right]$$
$$= \limsup_{n\to\infty} E^n \left[\frac{1}{2n} \sum_{i=1}^n \int_0^T \|u_i(t)\|^2 \, dt + F([Q^n]_1) \right]$$
$$\leq \frac{1}{2} E_\Theta \left[\int_{\mathbb{R}^m \times [0,T]} \|y\|^2 \rho(dy \, dt) \right] + F([\Theta]_1).$$

In particular, in obtaining the last equality we have used the uniqueness result in Lemma 4.2.4 (applied to the case where $\varphi = 0$). Combining the above inequality with (4.4.1) and since $\varepsilon > 0$ is arbitrary, we have the desired lower bound.

Finally, the proof that $\tilde{I}_{1,0}$ is a rate function can be carried out as in Section 4.2.3. We omit the details. □

4.4.2 The Case $\sqrt{n}\kappa(n) \to \infty$

For this case we assume Conditions 4.1.1 and 4.1.2. Condition 4.1.3 is not needed. In a similar manner to Theorem 4.2.1 it can be shown that for any $F \in \mathcal{C}_b(\mathcal{P}(\mathcal{X}))$ and for each $n \in \mathbb{N}$,

$$-\kappa(n)^2 \log E\left[e^{-\frac{1}{\kappa(n)^2}F(\mu^n)}\right] = \inf_{(u,v)\in\mathcal{A}^{1,n}\times\mathcal{A}^2} E\left[\frac{\kappa(n)^2}{2}\sum_{i=1}^n \int_0^T \|u_i(t)\|^2\,dt + \frac{1}{2}\int_0^T \|v(t)\|^2\,dt + F(\bar{\mu}^n)\right], \tag{4.4.2}$$

where $\bar{\mu}^n$ is as introduced below (4.2.3). Furthermore, for every $\delta > 0$, there is an $M < \infty$ such that for each $n \in \mathbb{N}$,

$$\begin{aligned} &-\kappa(n)^2 \log E\left[e^{-\frac{1}{\kappa(n)^2}F(\mu^n)}\right] \\ &\quad \geq \inf_{(u,v)\in\mathcal{A}^{1,n}\times\mathcal{A}_M^2} E\left[\frac{\kappa(n)^2}{2}\sum_{i=1}^n \int_0^T \|u_i(t)\|^2\,dt + \frac{1}{2}\int_0^T \|v(t)\|^2\,dt + F(\bar{\mu}^n)\right] - \delta. \end{aligned} \tag{4.4.3}$$

Fix $F \in \mathcal{C}_b(\mathcal{P}(\mathcal{X}))$ and $\delta \in (0,1)$. Select, for each $n \in \mathbb{N}$, $(u^n, v^n) \in \mathcal{A}^{1,n} \times \mathcal{A}_M^2$ such that

$$-\kappa(n)^2 \log E\left[e^{-\frac{1}{\kappa(n)^2}F(\mu^n)}\right] \geq E\left[\frac{\kappa(n)^2}{2}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\,dt + \frac{1}{2}\int_0^T \|v^n(t)\|^2\,dt + F(\bar{\mu}^n)\right] - 2\delta, \tag{4.4.4}$$

where $\bar{\mu}^n = \frac{1}{n}\sum_{i=1}^n \delta_{\bar{X}_i^n}$ and $\bar{X}_i^n$ is given by (4.2.3) (repalcing (u,v) with (u^n, v^n)). The uniform bound in (4.2.8) now gets replaced by

$$\sup_{n\in\mathbb{N}} E\left[\frac{\kappa(n)^2}{2}\sum_{i=1}^n \int_0^T \|u_i^n(t)\|^2\,dt + \frac{1}{2}\int_0^T \|v^n(t)\|^2\,dt\right] \leq 2(\|F\|_\infty + 1). \tag{4.4.5}$$

This in particular says that

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{T}\|u_i^n(t)\|^2\,dt\right]\leq\frac{4(\|F\|_\infty+1)}{n\kappa(n)^2}\to 0 \text{ as } n\to\infty, \tag{4.4.6}$$

since $n\kappa(n)^2\to\infty$. Define Q^n by (4.2.10), where ρ_i^n are as in (4.2.11). The tightness of (Q^n, v^n) is shown as in Lemma 4.2.2. Let (Q, v) be a weak limit point (along some subsequence) given on some probability space $(\Omega^*, \mathcal{F}^*, P^*)$. Then using (4.4.6) we see that

$$E^*\left[\int_{\mathcal{R}_1}\int_{\mathbb{R}^m\times[0,T]}\|y\|^2\,r(dy\,dt)\,[Q]_2(dr)\right]\leq\liminf_{n\to\infty}E\left[\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{T}\|u_i^n(t)\|^2\,dt\right]=0.$$

Thus we have that $[Q]_2=\delta_{r^o}$, where we recall that $r^o(dy\,dt)=\delta_{\{0\}}(dy)\,dt$. Also, as in Lemma 4.2.3, it can be seen that $Q\in\mathcal{E}_1[v]$ a.s. Combining this fact with $[Q]_2=\delta_{r^o}$ and recalling the definition of $\tilde{\mathcal{E}}_1$ given in Section 4.1.3, we now see that $[Q]_{(1,3)}\in\tilde{\mathcal{E}}_1[v]$ P^*-a.s. Taking the limit as $n\to\infty$ along a convergent subsequence

$$\begin{aligned}&\liminf_{n\to\infty}E\left[\frac{\kappa(n)^2}{2}\sum_{i=1}^{n}\int_0^T\|u_i^n(t)\|^2\,dt+\frac{1}{2}\int_0^T\|v^n(t)\|^2\,dt+F(\bar{\mu}^n)\right]\\&\geq E^*\left[\frac{1}{2}\int_0^T\|v(t)\|^2\,dt+F([Q]_1)\right]\\&\geq\inf_{\varphi\in L^2([0,T]:\mathbb{R}^k)}\inf_{\Theta\in\tilde{\mathcal{E}}_1[\varphi]}\left(\frac{1}{2}\int_0^T\|\varphi(t)\|^2\,dt+F([\Theta]_1)\right),\end{aligned}$$

where the last inequality uses the fact that $[Q]_{(1,3)}\in\tilde{\mathcal{E}}_1[v]$ P^*-a.s. Combining this with (4.4.4) and recalling that $\delta\in(0,1)$ is arbitrary completes the proof of the Laplace upper bound.

Now we consider the lower bound. Fix $\varepsilon>0$ and $F\in\mathcal{C}_b(\mathcal{P}(\mathcal{X}))$. Choose a $\varphi\in L^2([0,T]:\mathbb{R}^k)$ and a $\Theta^o\in\tilde{\mathcal{E}}_1[\varphi]$ such that

$$\frac{1}{2}\int_0^T\|\varphi(t)\|^2\,dt+F([\Theta^o]_1)\leq\inf_{\nu\in\mathcal{P}(\mathcal{X})}\left[F(\nu)+\tilde{I}_{1,\infty}(\nu)\right]+\varepsilon, \tag{4.4.7}$$

where $\tilde{I}_{1,\infty}$ is as in (4.1.11). Define Θ on $(\mathcal{Z}_1,\mathcal{B}(\mathcal{Z}_1))$ as $\Theta(dz,\,dr,\,dw)=\Theta^o(dz,\,dw)\,\delta_{r^o}(dr)$. Using this Θ, define $(\Omega',\mathcal{F}',P^n)$ as in Section 4.2.2. Also, let $v^n=\varphi$ for every n. Note that u_i defined through (4.2.29) satisfies $u_i=0$ P^n-a.s. Now with $\bar{\mu}^n$ and Q^n constructed as below (4.2.29), we have as in (4.2.31) that

$$\begin{aligned}&\limsup_{n\to\infty}-\kappa(n)^2\log E\left[e^{-\frac{1}{\kappa(n)^2}F(\mu^n)}\right]\\&\leq\limsup_{n\to\infty}E^n\left[\frac{\kappa(n)^2}{2}\sum_{i=1}^{n}\int_0^T\|u_i(t)\|^2\,dt+\frac{1}{2}\int_0^T\|v^n(t)\|^2\,dt+F(\bar{\mu}^n)\right]\end{aligned}$$

$$= \limsup_{n\to\infty} \left(\frac{1}{2} \int_0^T \|\varphi(t)\|^2 \, dt + E^n \left[F([Q^n]_1) \right] \right)$$
$$= \frac{1}{2} \int_0^T \|\varphi(t)\|^2 \, dt + F([\Theta]_1).$$

The last equality uses a uniqueness result of the type in Lemma 4.2.4 which is shown in the same manner. In particular, since $[\Theta]_2 = \delta_{r^o}$, the proof does not require Condition 4.1.3 since the analogue of the last term on the right side of (4.2.27), namely

$$E_{\hat{\Theta}} \left[\left(\int_0^T \left(\|X^{(1)}(s) - X^{(2)}(s)\| + d_{BL}(\nu^{(1)}(s), \nu^{(2)}(s)) \right) \|u(s)\| \, ds \right)^2 \right],$$

is simply zero. Combining the above inequality with (4.4.7) and since $\varepsilon > 0$ is arbitrary, we have the desired lower bound.

Finally, the proof that $\tilde{I}_{1,\infty}$ in (4.1.11) is a rate function is carried out as before and is omitted. □

4.5 Proof Sketch of Theorem 4.1.1(i)

Let $\rho_i^n = r^o$ for all $i = 1, \ldots n$ and $n \in \mathbb{N}$. With this choice of ρ_i^n, define Q^n by (4.2.10) by replacing $\bar{X}_i^n$ with X_i^n. By Lemmas 4.2.2 and 4.2.3, $\{Q^n\}$ is tight and any weak limit point Q satisfies $Q \in \mathcal{E}_1[0]$. This in particular shows that the nonlinear SDE

$$\begin{aligned} d\tilde{X}(t) &= b(\tilde{X}(t), \tilde{\mu}(t)) \, dt + \sigma(\tilde{X}(t), \tilde{\mu}(t)) \, dW(t), \\ \tilde{X}(t) &\sim \tilde{\mu}(t), \quad t \geq 0, \quad \tilde{\mu}(0) = \xi_0, \end{aligned} \tag{4.5.1}$$

has a weak solution, namely on some filtered probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}, \{\bar{\mathcal{F}}_t\})$ equipped with an m-dimensional $\bar{\mathcal{F}}_t$-Brownian motion W, there is an $\bar{\mathcal{F}}_t$-adapted process $\tilde{X}$ with sample paths in $\mathcal{C}([0,T] : \mathbb{R}^d)$ which satisfies the above equation. Furthermore, using standard Lipschitz estimates, martingale inequalities, and Gronwall's lemma, we see that pathwise uniqueness holds for (4.5.1). Thus, by the Yamada-Watanabe results (cf. [59, Chapter IV]) there is a unique weak solution to (4.5.1). Denote this weak solution (namely the probability law of $(\tilde{X}, W)$)) as $\Theta^*_{(1,3)}$. Let $\Theta^* \in \mathcal{P}(\mathcal{Z}_1)$ be defined as $\Theta^*(dx, dr, dw) = \Theta^*_{(1,3)}(dx, dw) \, \delta_{r^o}(dr)$. Then any weak limit point Q of Q^n must equal Θ^* a.s. As argued above, Θ^* is the unique element in $\mathcal{P}(\mathcal{Z}_1)$ that is a weak solution of $\mathcal{S}_1[0, \nu_\Theta]$ and satisfies $[\Theta^*]_2 = \delta_{r^o}$. The result follows. □

4.6 Proof Sketch of Theorem 4.1.2(i)

It was noted in Section 4.1.2, that the system of equations in (4.0.4) has a unique strong solution under Conditions 4.1.1, 4.1.2, 4.1.4, 4.1.5, and 4.1.6. This can be seen as follows. Note that, with $\zeta = b, \sigma, \alpha$, the maps

$$(x, a) = (x_1, \dots, x_n, a_1, \dots, a_n) \mapsto (\zeta(x_1, \mu(x, a)), \dots, \zeta(x_n, \mu(x, a))),$$

and with $\varsigma = c, \gamma^T, \beta^T$, the maps

$$(x, a) \mapsto (a_1 \varsigma(x_1, \mu(x, a)), \dots, a_n \varsigma(x_n, \mu(x, a))),$$

where $\mu(x, a) = \frac{1}{n} \sum_{i=1}^n \theta(a_i) \delta_{x_i}$, are locally Lipschitz functions with (at most) linear growth from $\mathbb{R}^{nd} \times \mathbb{R}_+^n$ to appropriate Euclidean spaces. For example, for $(x, a), (\tilde{x}, \tilde{a}) \in \mathbb{R}^{nd} \times \mathbb{R}_+^n$,

$$\begin{aligned}
|b(x_i, \mu(x, a)) - b(\tilde{x}_i, \mu(\tilde{x}, \tilde{a}))| &\le K(\|x_i - \tilde{x}_i\| + d_{BL}(\mu(x, a), \mu(\tilde{x}, \tilde{a}))) \\
&\le K \left(\|x_i - \tilde{x}_i\| + \frac{1}{n} \sum_{i=1}^n (|\theta(a_i) - \theta(\tilde{a}_i)| + \theta(a_i)\|x_i - \tilde{x}_i\|) \right).
\end{aligned}$$

The local Lipschitz property of $(x, a) \mapsto b(x_1, \mu(x, a))$ is immediate from the above estimate on recalling that under Conditon 4.1.5, θ is a Lipschitz function. Properties on other coefficients can be verified in a similar manner. Existence and uniqueness of strong solutions of (4.0.4) follows from this.

We are interested in the asymptotic behavior of $t \mapsto \mu^n(t)$ regarded as a sequence of $\mathcal{C}([0, T] : \mathcal{M}_+(\mathbb{R}^d))$-valued random variables, where $\mu^n(t)$ is defined as in (4.0.4). In order to characterize the limit of μ^n, we consider the nonlinear SDE $\mathcal{S}_2[0, \tilde{\mu}]$ in (4.1.5) with $\rho(dy\, dt) = r^o(dy)\, dt$, namely the following equation given on some filtered probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}, \{\bar{\mathcal{F}}_t\})$, equipped with an m-dimensional $\bar{\mathcal{F}}_t$-Brownian motion W:

$$\begin{aligned}
d\tilde{X}(t) &= b(\tilde{X}(t), \tilde{\mu}(t))\, dt + \sigma(\tilde{X}(t), \tilde{\mu}(t))\, dW(t) \\
d\tilde{A}(t) &= \tilde{A}(t) c(\tilde{X}(t), \tilde{\mu}(t))\, dt + \tilde{A}(t)\gamma^T(\tilde{X}(t), \tilde{\mu}(t))\, dW(t), \\
\langle f, \tilde{\mu}(t) \rangle &= \bar{E}[\theta(\tilde{A}(t)) f(\tilde{X}(t))], \quad f \in \mathcal{C}_b(\mathbb{R}^d), \quad t \ge 0, \quad (\tilde{X}(0), \tilde{A}(0)) \sim \eta_0.
\end{aligned} \tag{4.6.1}$$

Let $\tilde{\mathcal{Z}}_2 \doteq \mathcal{X} \times \mathcal{Y} \times \mathcal{W}$, and denote the canonical coordinate maps on this space as $(\tilde{X}, \tilde{A}, W)$. Let $\tilde{\mathcal{H}}_t \doteq \sigma\{\tilde{X}(s), \tilde{A}(s), W(s), s \le t\}$ be the canonical filtration on this space. By a *weak solution* of (4.6.1) we mean a probability measure Θ on $\tilde{\mathcal{Z}}_2$ such that, under Θ, W is a standard $\tilde{\mathcal{H}}_t$-Brownian motion and the system of equations (4.6.1) are satisfied a.s.

As before, let $\rho_i^n = r^o$ for all $i = 1, \ldots, n$ and $n \in \mathbb{N}$. Define Q^n by (4.3.9) by replacing $(\bar{X}_i^n, \bar{A}_i^n)$ with (X_i^n, A_i^n). By Lemmas 4.3.2 and 4.3.3, $\{Q^n\}$ is tight and any weak limit point Q satisfies $Q \in \mathcal{E}_2[0]$ (we use (4.3.10)(ii) here). In particular, this shows that $[Q]_{1,2,4}$ is a weak solution of (4.6.1). The following result shows the equation in fact has a unique weak solution.

Lemma 4.6.1. *Under Conditions 4.1.1, 4.1.2, 4.1.4, 4.1.5, and 4.1.6, equation* (4.6.1) *has a unique weak solution.*

Proof. It sufices to show that the equation has a unique pathwise solution, namely that if $(\tilde{X}^{(i)}, \tilde{A}^{(i)}, \tilde{\mu}^{(i)})$, $i = 1, 2$ are two solutions of (4.6.1) given on some filtered probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}, \{\bar{\mathcal{F}}_t\})$ equipped with an m-dimensional $\bar{\mathcal{F}}_t$-Brownian motion W (namely, $(\tilde{X}^{(i)}, \tilde{A}^{(i)})$ are continuous $\{\bar{\mathcal{F}}_t\}$ adapted processes and (4.6.1) is satisfied with $(\tilde{X}, \tilde{A}, \tilde{\mu})$ replaced with $(\tilde{X}^{(i)}, \tilde{A}^{(i)}, \tilde{\mu}^{(i)})$, $i = 1, 2$), and such that $(\tilde{X}^{(1)}(0), \tilde{A}^{(1)}(0)) = (\tilde{X}^{(2)}(0), \tilde{A}^{(2)}(0))$ $\bar{P}$-a.s., then

$$(\tilde{X}^{(1)}, \tilde{A}^{(1)}, \tilde{\mu}^{(1)}) = (\tilde{X}^{(2)}, \tilde{A}^{(2)}, \tilde{\mu}^{(2)}) \quad \bar{P}\text{-a.s.} \tag{4.6.2}$$

Using Conditions 4.1.1 and 4.1.6 on the initial random variables and Conditions 4.1.2 and 4.1.4 on the coefficients, it is easy to check by Gronwall's inequality that

$$\bar{E}\left[\sup_{0 \le t \le T} \left(\|\tilde{X}^{(i)}(t)\|^2 + \tilde{A}^{(i)}(t)^2\right)\right] < \infty \ \text{ for } i = 1, 2. \tag{4.6.3}$$

Let

$$g(t) = \bar{E}\left[\sup_{0 \le s \le t} \|\tilde{X}^{(1)}(s) - \tilde{X}^{(2)}(s)\|^2\right] \text{ and } h(t) = \left(\bar{E}\left[\sup_{0 \le s \le t} |\tilde{A}^{(1)}(s) - \tilde{A}^{(2)}(s)|\right]\right)^2.$$

Then, exactly as for (4.3.24), there is a $c_1 \in (0, \infty)$ such that for all $0 \le s \le t \le T$,

$$d_{BL}\left(\tilde{\mu}^{(1)}(s), \tilde{\mu}^{(2)}(s)\right)^2 \le c_1(g(t) + h(t)).$$

By the Lipschitz property of b and σ, we then have that for some $c_2 \in (0, \infty)$ and all $0 \le t \le T$,

$$g(t) \le c_2 \int_0^t (g(s) + h(s)) \, ds.$$

Writing $\tilde{A}^{(i)}(t) = e^{\tilde{Y}^{(i)}(t)}$ for $i = 1, 2$ and using the bounded Lipschitz properties of c and γ, we see as in (4.3.26) that for some $c_3 \in (0, \infty)$ and all $0 \le t \le T$,

$$h(t) \le c_3 \int_0^t (g(s) + h(s)) \, ds.$$

Thus, $g(t)+h(t) \le (c_2+c_3)\int_0^t (g(s)+h(s))\, ds$ for all $t \in [0,T]$ which, by Gronwall's inequality, then shows that $g(T)+h(T)=0$. Thus $(\tilde{X}^{(1)}, \tilde{A}^{(1)})$ and $(\tilde{X}^{(2)}, \tilde{A}^{(2)})$ are indistinguishable on $[0,T]$ which proves (4.6.2). $\square$

We now complete the proof of Theorem 4.1.2(1). Denoting the unique weak solution of (4.6.1) as $Q^*_{(1,2,4)}$ we now have that $[Q^n]_{(1,2,4)} \to Q^*_{(1,2,4)}$ in probability as $n \to \infty$. Let $Q^*(dx,\, da,\, dr, dw) \doteq Q^*_{(1,2,4)}(dx,\, da,\, dw)\delta_{r^o}(dr)$. Then $Q^n \to Q^*$ in probability. Note that Q^* is the unique element Θ in $\mathcal{P}(\mathcal{Z}_2)$ that is a weak solution of $\mathcal{S}_2[0,\nu_\Theta]$ and satisfies $[\Theta]_3 = \delta_{r^o}$. Using the estimate

$$\sup_{n\in\mathbb{N}} E\left[\frac{1}{n}\sum_{i=1}^{n} \sup_{0\le t\le T} A_i^n(t)^2\right] < \infty,$$

which follows by the argument in (4.3.11), it now follows exactly as in the proof of the Laplace upper bound (see arguments below the proof of Lemma 4.3.3) that $\nu_{Q^n} \to \nu_{Q^*}$ in $\mathcal{K}$, in probability, where for $\Theta \in \mathcal{P}_2(\mathcal{Z}_2)$, ν_Θ is defined as in (4.1.6). $\square$

CHAPTER 5

Large Deviations For Stochastic Currents with Mean Field Interaction and Vanishing Noise

In this chapter we consider the interacting particle system described through a collection of SDEs on $\mathbb{R}^d$ given as

$$dX_j^N(t) = b\left(X_j^N(t), V^N(t)\right)\, dt + \varepsilon_N \sigma\left(X_j^N(t), V^N(t)\right)\, dW_j(t), \qquad 1 \le j \le N, \qquad N \in \mathbb{N}, \tag{5.0.1}$$

on some finite time horizon $0 \le t \le T$, where $\varepsilon_N \downarrow 0$ as $N \to \infty$ and $\{W_j, j \in \mathbb{N}\}$ are independent m-dimentional Brownian motions on $[0, T]$. Here $V^N(t)$ is the empirical measure of the particle states at time t, namely

$$V^N(t) = \frac{1}{N}\sum_{j=1}^{N} \delta_{X_j^N(t)}, \qquad 0 \le t \le T,$$

and thus the interaction among the particles is of the mean-field type and influences both the drift and diffusion coefficients of each particle. The law of large numbers and fluctuation results for such mean-field systems have been widely studied, see for instance [18, 32, 71, 82, 83, 85, 91]. In particular, when $N \to \infty$, under conditions on the coefficients and the initial data, $\{V^N(t), 0 \le t \le T\}$ converges to the solution of the Vlasov equation

$$\frac{\partial}{\partial t} V + \nabla \cdot b(\cdot, V) V = 0,$$

which can be formally written as

$$\frac{\partial}{\partial t} V + \nabla \cdot \mathcal{J} = 0, \tag{5.0.2}$$

where $\mathcal{J} \doteq b(\cdot, V)V$ is the nonlinear current given as the limit of the *stochastic currents*

$$J^N(\varphi) = \frac{1}{N}\sum_{j=1}^{N} \int_0^T \varphi\left(t, X_j^N(t)\right) \circ dX_j^N(t), \tag{5.0.3}$$

defined for arbitrary smooth and compactly supported $\varphi : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d$, where $\circ$ denotes the Stratonovich integral. Currents and their stochastic counterparts are key objects in geometric measure theory and play an important role in the theory of rough paths (cf. [44, 80, 54, 78]). In the current context they provide a convenient way to describe the asymptotics of the empirical measure process V^N.

In this chapter we are interested in studying the asymptotics of probabilities of significant deviations of the empirical measure V^N, for the N-particle microscopic stochastic evolution described by (5.0.1), from its macroscopic hydrodynamic limit described by the first order Vlasov equation in (5.0.2). A common approach to such a study is by establishing a general LDP on an appropriate abstract space from which the information on probabilities of deviations for specific events involving the N-particle system (5.0.1) can be obtained by a suitable application of the contraction principle. In view of the representation of the hydrodynamic limit of V^N in terms of the nonlinear current functional $\mathcal{J}$, a natural candidate for an LDP are the pairs (V^N, J^N) regarded as random elements of an appropriate space. Under the conditions on the coefficients considered in this work (see Condition 5.1.1), V^N will take values in $\mathcal{V} \doteq \mathcal{C}([0,T], \mathcal{P}_1(\mathbb{R}^d))$, namely, the space of continuous functions from $[0,T]$ to the space $\mathcal{P}_1(\mathbb{R}^d)$ of probability measures on $\mathbb{R}^d$ with finite first moment, equipped with the Wasserstein-1 distance (see Section 5.1 for precise definitions). The identification of an appropriate space for J^N requires a bit more work (cf. [42, 88]). In particular, note that (5.0.3) describes an uncountably infinite collection of identities in which the right side is defined in an almost sure sense for each fixed φ. Thus a basic problem is to provide a pathwise representation for the collection

$$\left\{ \varphi \mapsto \frac{1}{N} \sum_{j=1}^{N} \int_0^T \varphi\left(t, X_j^N(t)\right) \circ dX_j^N(t) \right\}, \tag{5.0.4}$$

which defines a continuous, linear map on a suitable function space. This problem was studied in [42] (see also [88]) where it was shown that there is a random variable $\mathcal{J}^N$ with values in a certain negative Sobolev space $\mathbf{H}^{-\mathbf{s}}$ of distributions (see Section 5.1.2), which gives a pathwise representation for the collection in (5.0.4) in the sense that

$$\langle \mathcal{J}^N, \varphi \rangle = \frac{1}{N} \sum_{j=1}^{N} \int_0^T \varphi\left(t, X_j^N(t)\right) \circ dX_j^N(t) \qquad \text{a.s.,}$$

for every smooth φ with compact support. Thus the stochastic currents $\mathcal{J}^N$ can be viewed as random elements of the Hilbert space $\mathbf{H}^{-\mathbf{s}}$, and the basic problem of interest is then to establish a large deviation principle for $(V^N, \mathcal{J}^N)$ in $\mathcal{V} \times \mathbf{H}^{-\mathbf{s}}$.

This large deviation problem in the setting where $m = d$ and $\sigma = \mathrm{Id}$ was studied in [88] by direct change of measure arguments. Specifically, [88] treats the large deviation upper bound by first establishing an estimate for compact sets by considering an explicit tilt of the measure and then extends the estimate to all closed sets by establishing certain exponential tightness estimates. The lower bound is proved by exploiting connections between large deviations and Γ-convergence from [81], in particular the key idea is to construct a suitable 'recovery sequence' using results from [79]. One important aspect of the results and proof methods in [88] is that the LDP is established with the weak topology on the Hilbert space $\mathbf{H}^{-\mathbf{s}}$. Indeed, both the proofs of the upper and lower bounds rely on the use of the weak topology in important ways, e.g. since bounded sets are relatively compact under the weak topology in $\mathbf{H}^{-\mathbf{s}}$, in proving exponential tightness it suffices to estimate the probability that $\mathcal{J}^N$ takes values in the complement of a bounded ball.

In the current work we take a different approach to the study of the large deviation principle that is based on methods from stochastic control, the theory of weak convergence of probability measures, and Laplace asymptotics. This approach allows us to avoid establishing exponential tightness estimates of the form in [88] and enables us to treat diffusion coefficients that are state dependent and possibly degenerate (see Section 5.1.1). In addition, since in this approach one needs to establish ordinary tightness rather than exponential tightness, by appealing to certain compact embedding results for Sobolev spaces, we are able to establish an LDP with the norm topology on $\mathbf{H}^{-\mathbf{s}}$ instead of the weak topology considered in [88]. In fact, we establish a somewhat more general large deviation principle than the one considered in [88] from which the LDP for $(V^N, \mathcal{J}^N)$ can be deduced by the contraction principle. Specifically, we consider path empirical measures μ^N associated with the interacting particle system in (5.0.1) defined as

$$\mu^N = \frac{1}{N}\sum_{j=1}^{N} \delta_{X_j^N}.$$

Under the conditions of this work it follows that μ^N is a random variable with values in $\mathcal{P}_1(\mathcal{C}([0,T],\mathbb{R}^d))$, namely the space of probability measures, on the Banach space of $\mathbb{R}^d$-valued continuous trajectories on $[0,T]$, with integrable norm (equipped with the Wasserstein-1 metric). Our main result, Theorem 5.1.3, gives an LDP for $(\mu^N, \mathcal{J}^N)$ in $\mathcal{P}_1(\mathcal{C}([0,T],\mathbb{R}^d)) \times \mathbf{H}^{-\mathbf{s}}$. Using the continuity of the map $\nu \mapsto \{t \mapsto \nu \circ \pi_t^{-1}\}$ from $\mathcal{P}_1(\mathcal{C}([0,T],\mathbb{R}^d))$ into $\mathcal{V}$, where π_t is the projection map on $\mathcal{C}([0,T],\mathbb{R}^d)$ giving the evaluation at time t, we then deduce an LDP for the sequence $(V^N, \mathcal{J}^N)$ in $\mathcal{V} \times \mathbf{H}^{-\mathbf{s}}$ in Corollary 5.1.4. The rate function, in the general setting of a state dependent diffusion coefficient, is given as a value function of a certain deterministic mean field control problem with a quadratic cost (see (5.1.11) and (5.1.16)). In Proposition 5.1.5 we show that in the special case where $\sigma = \mathrm{Id}$, this representation of the rate function simplifies to a more explicit form given in terms of certain controlled Vlasov equations (see (5.1.17)) which was obtained in [88].

As noted previously, proof techniques here are quite different from [88]. The starting point of our analysis is a certain variational representation for exponential functionals of finite dimensional Brownian motions (see [17, 20]), using which the proof of the large deviation principle reduces to a study of tightness and convergence properties of certain controls and controlled analogues of the state processes $\{X_j^N, 1 \le j \le N\}$, state empirical measures V^N, path occupation measures μ^N, and stochastic currents $\mathcal{J}^N$, denoted as $\{\bar{X}_j^N, 1 \le j \le N\}$, $\bar{V}^N$, $\bar{\mu}^N$, and $\bar{\mathcal{J}}^N$, respectively. For the upper bound proof we introduce certain joint empirical measures, denoted as Q^N (see (5.2.7)), of particle trajectories and associated control processes. The main step in the proof of the upper bound is to establish the tightness of the sequence $\{(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N), N \in \mathbb{N}\}$ and to provide a suitable characterization of the weak limit points of this sequence. In particular, the tightness of the controlled stochastic currents $\{\bar{\mathcal{J}}^N\}$ is established with the norm topology on $\mathbf{H}^{-\mathbf{s}}$ and relies on approximations of $\{\bar{\mathcal{J}}^N\}$ by distributions with compact support as well as certain compact embedding results for Sobolev spaces (see Lemma 5.3.4). The lower bound proof is constructive in that, given a near optimal measure μ on $\mathcal{C}([0,T],\mathbb{R}^d)$ and a near optimal current $\mathcal{J}$ in a certain variational problem associated with the rate function, we construct a sequence of controls and controlled variables $(\bar{\mu}^N, \bar{\mathcal{J}}^N)$ that converge to $(\mu, \mathcal{J})$ in a suitable manner. The key ingredients in the proof here are a weak uniqueness (i.e. uniqueness in probability laws) property of certain equations associated with the controlled versions of the Vlasov equation (5.0.2) (see Lemma 5.2.4) and certain infinite product space constructions.

Large deviation principles for weakly interacting diffusions as in (5.0.1) with non-vanishing noise (i.e. $\varepsilon_N = 1$) have been studied in [33]. A different approach, based on weak convergence methods of the form used in the current work, was taken in [22]. The latter paper, in contrast to [33], allowed for degenerate diffusion coefficients and for a mean field interaction in the diffusion coefficient. There have also been several works (in addition to the paper [88] discussed above) that have studied large deviation problems for weakly interacting diffusions with small noise. In particular, see [56], [36], and references therein, for large deviations results for McKean-Vlasov equations in the small noise limit; and see [57] for an analysis of interchanging of mean-field limit with the small noise limit at the level of rate function convergence.

This chapter is organized as follows. In Section 5.1, we specify our model, describe the space on which the large deviation principle will hold, define the rate function, and present our main large deviation result. Section 5.2 provides the proof of this result, with the proofs of its key lemmas given in Section 5.3. The proofs of some auxiliary results are given at the end.

The following notation will be used throughout. We use $\mathcal{C}(R,S)$, $\mathcal{C}_c(R,S)$, and $\mathcal{C}^k(R,S)$, $k \in \mathbb{N} \cup \{\infty\}$, to denote the spaces of continuous, continuous and compactly supported, and k-times continuously differentiable functions from R into S, respectively. Also, $\mathcal{C}_c^k(R,S) = \mathcal{C}_c(R,S) \cap \mathcal{C}^k(R,S)$ for $k \in \mathbb{N} \cup \{\infty\}$. We denote by $L^2(\mu, R, S)$ the space of μ-square integrable functions from R into S. When μ is the Lebesgue measure, we will occasionally suppress it in the notation and write $L^2(\mu, R, S)$ as $L^2(R,S)$. The evaluation of a distribution F on a test function φ will be denoted by $\langle F, \varphi \rangle$, and integration of a function f with respect to a measure μ will be denoted by $\langle \mu, f \rangle$. $\mathcal{B}(S)$ denotes the collection of all Borel sets on S. For a Polish space (S, d_S), $\mathcal{P}(S)$ denotes the space of probability measures on S, endowed with the topology of weak convergence. A convenient metric on this space is the bounded Lipschitz metric given as

$$d_{\text{bl}}(\mu,\nu) \doteq \sup_{f \in \mathcal{L}_{\text{b}}(S)} |\langle \mu, f\rangle - \langle \nu, f \rangle|, \qquad \mu, \nu \in \mathcal{P}(S), \qquad \text{where}$$

$$\mathcal{L}_{\text{b}}(S) \doteq \left\{ f \in \mathcal{C}(S,\mathbb{R}) : \sup_{x \neq y} \frac{|f(x) - f(y)|}{d_S(x,y)} \leq 1, \ \sup_x |f(x)| \leq 1 \right\}.$$

When $\theta \in \mathcal{P}(S)$, the notation E_θ will be used to denote expectation on the probability space $(S, \mathcal{B}(S), \theta)$. For two spaces S_1 and S_2 and $\theta \in \mathcal{P}(S_1 \times S_2)$, $\theta_{(1)}$ and $\theta_{(2)}$ will denote the marginal distributions on S_1 and S_2, respectively. Similar notation will be used when more than two spaces are involved. Euclidean norms will be denoted by $|\cdot|$. For a Polish space (S, d_S), the space $\mathcal{C}([0,T], S)$ will be equipped with the metric

$$d(x,y) = \sup_{0 \le t \le T} d_S(x(t), y(t)),$$

under which it is a Polish space as well. On $\mathcal{C}([0,T], \mathbb{R}^d)$, we define the norm $\|x\|_\infty \doteq \sup_{0 \le t \le T} |x(t)|$, and the metric above becomes $d(x,y) = \|x - y\|_\infty$. We will use $\Rightarrow$ to denote convergence in distribution, and $\xrightarrow{P}$ to denote convergence in P-probability. Infimum over an empty set, by convention, is taken to be $+\infty$. For a metric space S, a function $I : S \to [0, \infty]$ is called a rate function if $\{x \in S : I(x) \le l\}$ is a compact set for every $l < \infty$.

5.1 Preliminaries and Main Result

Let $(\Omega, \mathcal{F}, P, \{\mathcal{F}(t), 0 \le t \le T\})$ be a filtered probability space where the filtration satisfies the *usual conditions* (see [69, Definition 21.22]). Fix $m \in \mathbb{N}$, and let $\{W_j, j \in \mathbb{N}\}$ be a sequence of independent m-dimensional $\{\mathcal{F}(t)\}$-Brownian motions on the time horizon $0 \le t \le T$. For each $N \in \mathbb{N}$, we consider the following system of stochastic differential equations in $\mathbb{R}^d$:

$$X_j^N(t) = X_j^N(0) + \int_0^t b\left(X_j^N(s), V^N(s)\right)\, ds + \varepsilon_N \int_0^t \sigma\left(X_j^N(s), V^N(s)\right)\, dW_j(s), \qquad 1 \le j \le N, \tag{5.1.1}$$

where $V^N(t)$ denotes the $\mathcal{P}(\mathbb{R}^d)$-valued empirical measure

$$V^N(t) \doteq \frac{1}{N} \sum_{j=1}^N \delta_{X_j^N(t)}, \qquad 0 \le t \le T, \tag{5.1.2}$$

and $\{\varepsilon_N, N \in \mathbb{N}\}$ is some sequence in $\mathbb{R}_+$ such that $\varepsilon_N \downarrow 0$ as $N \to \infty$. Without loss of generality, we will assume that $\sup_N \varepsilon_N \le 1$ throughout. Denote $\mathcal{X} \doteq \mathcal{C}([0,T], \mathbb{R}^d)$, and define $\mathcal{P}(\mathcal{X})$-valued random variables, given as the empirical measure of $(X_1^N, \ldots, X_N^N)$, by

$$\mu^N \doteq \frac{1}{N} \sum_{j=1}^N \delta_{X_j^N}. \tag{5.1.3}$$

Note that the marginal of μ^N at time t is $V^N(t)$, that is, defining $\pi_t : \mathcal{C}([0,T], \mathbb{R}^d) \to \mathbb{R}^d$ as the projection map $\pi_t(x) = x(t)$, we have

$$\mu^N \circ \pi_t^{-1} = V^N(t), \qquad 0 \le t \le T.$$

We will view each μ^N as a random variable taking values in the Wasserstein-1 space which is defined as follows. For a Polish space (S, d_S), define the space $\mathcal{P}_1(S)$ by

$$\mathcal{P}_1(S) \doteq \left\{ \mu \in \mathcal{P}(S) : \int_S d_S(x, x_0)\, \mu(dx) < \infty \right\},$$

for some choice of $x_0 \in S$ (the space does not depend on the choice of x_0). Then $\mathcal{P}_1(S)$ is a Polish space under the Wassertstein-1 distance given by

$$d_1(\mu, \nu) \doteq \sup_{f \in \mathcal{L}(S)} |\langle \mu, f \rangle - \langle \nu, f \rangle|, \qquad \mathcal{L}(S) \doteq \left\{ f \in \mathcal{C}(S, \mathbb{R}) : \sup_{x \neq y} \frac{|f(x) - f(y)|}{d_S(x, y)} \leq 1 \right\}. \tag{5.1.4}$$

For further details on Wassertstein spaces, we refer to [99]. The particular cases of interest here are the spaces $\mathcal{P}_1(\mathbb{R}^d)$ and $\mathcal{P}_1(\mathcal{X})$, and the notation d_1 will be used for the metric on both spaces, with the distinction being clear from context. Noting that (under Condition 5.1.1 given below)

$$\int_{\mathcal{X}} d_{\mathcal{X}}(x, 0)\, \mu^N(dx) = \int_{\mathcal{X}} \|x\|_\infty\, \mu^N(dx) = \frac{1}{N} \sum_{j=1}^N \left\| X_j^N \right\|_\infty < \infty \qquad \text{a.s.},$$

we see that indeed μ^N is a $\mathcal{P}_1(\mathcal{X})$-valued random variable. Similarly, it can be checked that V^N is a $\mathcal{C}([0,T], \mathcal{P}_1(\mathbb{R}^d))$-valued random variable. Throughout, we will denote $\mathcal{V} \doteq \mathcal{C}([0,T], \mathcal{P}_1(\mathbb{R}^d))$.

5.1.1 Main Conditions.

The following is our main assumption on the coefficients.

Condition 5.1.1. *There is some $L < \infty$ such that for all $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$,*

$$|b(x, \mu) - b(y, \nu)| + |\sigma(x, \mu) - \sigma(y, \nu)| \leq L \left(|x - y| + d_1(\mu, \nu) \right),$$

and $|\sigma(x, \mu)| \leq L$.

Note that the above condition implies in particular that for all $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$,

$$|b(x, \mu)| \leq L \left(1 + |x| + \int_{\mathbb{R}^d} |y|\, \mu(dy) \right). \tag{5.1.5}$$

with possibly a larger choice of L than in Condition 5.1.1. By standard arguments, Condition 5.1.1 implies that there exists a unique pathwise solution to (5.1.1) for each $N \in \mathbb{N}$.

We assume the following on the initial conditions of (5.1.1).

Condition 5.1.2. *For each $N \in \mathbb{N}$ and $1 \leq j \leq N$, $X_j^N(0) = x_j^N \in \mathbb{R}^d$ is deterministic. The collection of initial conditions satisfies the following.*

(i) There exists some $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$ such that, $d_{bl}\left(V^N(0), \mu_0\right) \to 0$.

(ii) $\sup_{N \geq 1} \frac{1}{N} \sum_{j=1}^{N} \left|x_j^N\right|^2 < \infty.$

Note that (i) and (ii) above imply that $\int_{\mathbb{R}^d} |x|^2 \mu_0(dx) < \infty$ from the observation

$$\int_{\mathbb{R}^d} \left(|x|^2 \wedge K\right) \mu_0(dx) = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \left(\left|x_j^N\right|^2 \wedge K\right) \leq \sup_{N \geq 1} \frac{1}{N} \sum_{j=1}^{N} \left|x_j^N\right|^2$$

for any $K \in (0, \infty)$, and applying Fatou's lemma. The above condition also gives that, as $N \to \infty$,

$$d_1\left(V^N(0), \mu_0\right) \to 0.$$

In order to prove the Laplace lower bound, we will make a stronger assumption given below on the diffusion coefficient σ which says that it depends on the state of the system only through the empirical measure. We will also require the convergence of the initial data in a somewhat stronger sense.

Condition 5.1.3. *(i) For each $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, $\sigma(x, \mu) = \sigma(\mu)$.*

(ii) For all μ_0-integrable $f : \mathbb{R}^d \to \mathbb{R}$,

$$\left\langle V^N(0), f\right\rangle \to \langle \mu_0, f\rangle \qquad \textit{as } N \to \infty$$

We are interested in the large deviations behavior of μ^N and V^N as well as a collection of random linear functionals, referred to as *stochastic currents*, associated with the sequence of processes $\{X_j^N(t)\}$. We now introduce these objects. For each N and $\varphi \in \mathcal{C}_c^\infty([0, T] \times \mathbb{R}^d, \mathbb{R}^d)$ define

$$J^N(\varphi) \doteq \frac{1}{N} \sum_{j=1}^{N} \int_0^T \varphi\left(t, X_j^N(t)\right) \circ dX_j^N(t), \tag{5.1.6}$$

where the above is a Stratanovich stochastic integral. The relationship between Stratanovich and Itô integrals gives the following formula for $J^N(\varphi)$:

$$J^N(\varphi) = \frac{1}{N} \sum_{j=1}^{N} \left(\int_0^T \varphi\left(t, X_j^N(t)\right) \cdot dX_j^N(t) + \frac{1}{2}\left\langle \varphi\left(\cdot, X_j^N(\cdot)\right), X_j^N(\cdot)\right\rangle_T\right),$$

where $\langle Y, Z\rangle_t$ denotes the quadratic variation at time t of two continuous semimartingales Y and Z. From results in [42], J^N can be viewed as a random linear functional on a suitable Sobolev space. We now briefly describe these results and make precise the space in which these random linear functionals take values.

5.1.2 Stochastic Currents

Recall that for $k \in \mathbb{N}$, $H^k(\mathbb{R}^d, \mathbb{R}^d)$ is the Hilbert space of functions $f \in L^2(\mathbb{R}^d, \mathbb{R}^d)$ such that the distributional derivatives $D^\alpha f$ are also L^2 functions for all $|\alpha| \le k$, where $\alpha = (\alpha_1, \dots, \alpha_d)$ denotes a multi-index. More generally, for any $s \in \mathbb{R}_+$, $H^s(\mathbb{R}^d, \mathbb{R}^d)$ is defined as the space of functions $f \in L^2(\mathbb{R}^d, \mathbb{R}^d)$ such that

$$\|f\|_s^2 \doteq \int_{\mathbb{R}^d} |\hat{f}(\xi)|^2 (1 + |\xi|^2)^s \, d\xi < \infty, \tag{5.1.7}$$

where $\hat{f}(\xi) = \int e^{-2\pi i \xi \cdot x} f(x)\, dx$ is the Fourier transform on $\mathbb{R}^d$. We refer the reader to [1, 45, 39] for details on these spaces.

In order to describe the linear space associated with the map $\varphi \mapsto J^N(\varphi)$, we will need to consider a suitable Sobolev space of functions of time and space. Following [73, 42, 88], a natural choice in this regard is the space

$$H^{s_1}\left((0,T), H^{s_2}\left(\mathbb{R}^d, \mathbb{R}^d\right)\right),$$

where $\mathbf{s} = (s_1, s_2) \in \left(\frac{1}{2}, 1\right) \times \left(\frac{d}{2} + 1, \infty\right)$ (see [88] for a precise description of the space). However in order to apply certain compact embedding results (see e.g. the proof of Lemma 5.2.3) we will consider a slight modification of these spaces defined as follows.

Fix $a, b \in \mathbb{R}$ such that $a < 0 < T < b$ and define $U \doteq (a,b)$ and $\mathcal{O}_d \doteq \left(\frac{1}{2}, 1\right) \times \left(\frac{d}{2} + 1, \infty\right)$. Then define

$$\mathbf{H}^{\mathbf{s}} \doteq H^{s_1}\left(U, H^{s_2}\left(\mathbb{R}^d, \mathbb{R}^d\right)\right), \qquad \mathbf{s} \in \mathcal{O}_d,$$

as the space of functions $f : U \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying

$$\begin{aligned} \|f\|_{\mathbf{s}}^2 &\doteq \|f\|^2_{L^2(U, H^{s_2}(\mathbb{R}^d, \mathbb{R}^d))} + [f]_{\mathbf{s}}^2 \\ &\doteq \int_U \|f(u,\cdot)\|_{s_2}^2 \, du + \int_U \int_U \frac{\|f(u,\cdot) - f(v,\cdot)\|_{s_2}^2}{|u-v|^{1+2s_1}} \, du \, dv < \infty, \end{aligned} \tag{5.1.8}$$

where $\|\cdot\|_{s_2}$ is as in (5.1.7). The norm $\|\cdot\|_{\mathbf{s}}$ is usually referred to as a *Gagliardo norm*, and in fact corresponds to an inner product which makes $\mathbf{H}^{\mathbf{s}}$ a separable Hilbert space (see [39, Section 3]). The topological dual of the Hilbert space $\mathbf{H}^{\mathbf{s}}$ will be denoted as $\mathbf{H}^{-\mathbf{s}}$, namely

$$\mathbf{H}^{-\mathbf{s}} \doteq \left(\mathbf{H}^{\mathbf{s}}\right)'.$$

The norm on this space is given as

$$\|F\|_{-\mathbf{s}} \doteq \sup_{\varphi \in \mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)} \frac{|\langle F, \varphi\rangle|}{\|\varphi\|_{\mathbf{s}}}.$$

For $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$, abusing notation, we let

$$J^N(\varphi) \doteq \frac{1}{N}\sum_{j=1}^N \int_0^T \varphi\left(t, X_j^N(t)\right) \circ dX_j^N(t).$$

Note that if φ_{res} denotes the restriction of φ to $[0,T]\times\mathbb{R}^d$, then $J^N(\varphi) = J^N(\varphi_{\text{res}})$. Also, any $\varphi \in \mathcal{C}_c^\infty([0,T]\times\mathbb{R}^d,\mathbb{R}^d)$ can be extended to a $\varphi_{\text{ext}} \in \mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$ where once more $J^N(\varphi) = J^N(\varphi_{\text{ext}})$. By a *pathwise realization* of the collection $\{\varphi \mapsto J^N(\varphi)\}$ on $\mathcal{C}_c^\infty([0,T]\times\mathbb{R}^d,\mathbb{R}^d)\}$, we mean a random variable $\mathcal{J}^N$ with values in $\mathbf{H}^{-\mathbf{s}}$ such that for any $\varphi \in \mathcal{C}_c^\infty([0,T]\times\mathbb{R}^d,\mathbb{R}^d)$ and any extension φ_{ext} of φ in $\mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$, $\langle \mathcal{J}^N, \varphi_{\text{ext}}\rangle = J^N(\varphi)$ a.s.

The following result, giving the existence of a pathwise realization, follows along the lines of [88] . The proof is an immediate consequence of Lemma 5.2.1 below (on taking $u_j^N = 0$ in the lemma), the proof of which is given in the Appendix.

Theorem 5.1.1. *Suppose Conditions 5.1.1 and 5.1.2 hold. Then for each $N \in \mathbb{N}$ and $\mathbf{s} \in \mathcal{O}_d$, there is an $\mathbf{H}^{-\mathbf{s}}$-valued random variable $\mathcal{J}^N$ on $(\Omega, \mathcal{F}, P)$ such that for every $\varphi \in \mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$, $\langle \mathcal{J}^N(\omega), \varphi\rangle = [J^N(\varphi)](\omega)$ for a.e. $\omega \in \Omega$. Namely, $\mathcal{J}^N$ is a pathwise realization of $\{\varphi \mapsto J^N(\varphi)\}$.*

Note that the pathwise realizations $\{\mathcal{J}^N\}$ are a.s. compactly supported in the first coordinate. Namely, if $U_0 \subset U$ is an open set such that $U_0 \cap [0,T] = \varnothing$, then for all φ with compact support in $U_0 \times \mathbb{R}^d$, $\langle \mathcal{J}^N, \varphi\rangle = 0$ a.s. In particular, $\mathcal{J}^N$ is a distribution a.s. supported in $[0,T]\times\mathbb{R}^d$.

In this work we will prove a large deviation principle for the pair $(\mu^N, \mathcal{J}^N)$ in the space $\mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}}$ for each $\mathbf{s} \in \mathcal{O}_d$, from which a LDP describing the asymptotics of V^N will follow by the contraction principle. We begin by introducing the rate function that will govern the large deviation behavior.

5.1.3 Rate Function

Let $\mathcal{R}$ denote the set of positive measures r on $\mathcal{B}([0,T]\times\mathbb{R}^m)$ such that $r([0,t]\times\mathbb{R}^m) = t$ for all $0 \le t \le T$, and define

$$\mathcal{R}_1 \doteq \left\{ r \in \mathcal{R} : \int_{[0,T]\times\mathbb{R}^m} |y|\, r(dt, dy) < \infty \right\}.$$

The space $\mathcal{R}_1$ is a Polish space under the Wasserstein-1 metric (defined as in (5.1.4) with $S = [0,T]\times\mathbb{R}^m$). Each $r \in \mathcal{R}_1$ can be decomposed as $r(dt,dy) = r_t(dy)\,dt$, where $r_t \in \mathcal{P}(\mathbb{R}^m)$. For an $\mathcal{R}_1$-valued random variable ρ, consider the McKean-Vlasov equation

$$\begin{aligned} dX(t) &= b(X(t),V(t))\,dt + \int_{\mathbb{R}^m} \sigma(X(t),V(t))y\,\rho_t(dy)\,dt, \\ V(t) &= P\circ X(t)^{-1}, \qquad V(0)=\mu_0, \end{aligned} \tag{5.1.9}$$

where X is stochastic process with sample paths in $\mathcal{X}$, $\rho(dt,dy)=\rho_t(dy)\,dt$ is the disintegration of ρ, and μ_0 is the measure in Condition 5.1.2(i). The distribution of a pair (X,ρ) that solves (5.1.9), which is a probability measure on $\mathcal{Z}\doteq\mathcal{X}\times\mathcal{R}_1$, is called a weak solution of (5.1.9). Let $\mathcal{S}(\mathcal{Z})\subset\mathcal{P}(\mathcal{Z})$ denote the set of all such weak solutions. With an abuse of notation, we will denote the canonical coordinate maps on $(\mathcal{Z},\mathcal{B}(\mathcal{Z}))$ by (X,ρ) once more. That is,

$$X(\xi,r)=\xi, \qquad \rho(\xi,r)=r, \qquad (\xi,r)\in\mathcal{Z}.$$

Note that if $\Theta\in\mathcal{S}(\mathcal{Z})$, then (X,ρ) satisfy (5.1.9) Θ-a.s. For each $\Theta\in\mathcal{P}(\mathcal{Z})$ and $0\le t\le T$, define the measure

$$\nu_\Theta(t)\doteq\Theta\circ X(t)^{-1},$$

which is an element of $\mathcal{P}(\mathbb{R}^d)$. When $\Theta\in\mathcal{S}(\mathcal{Z})$, it is easy to check that Condition 5.1.1 and Gronwall's lemma imply that $E_\Theta\left[|X(t)|\right]<\infty$, and hence $\nu_\Theta(t)\in\mathcal{P}_1(\mathbb{R}^d)$ for each $0\le t\le T$. Letting ν_Θ denote the map $t\mapsto\nu_\Theta(t)$, in fact we have that $\nu_\Theta\in\mathcal{V}$. For each $\varphi\in\mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$, define the map $G_\varphi:\mathcal{S}(\mathcal{Z})\to\mathbb{R}$ by

$$\begin{aligned} G_\varphi(\Theta) &\doteq E_\Theta\left[\int_0^T \varphi\left(t,X(t)\right)\cdot dX(t)\right] \\ &= E_\Theta\left[\int_0^T \varphi\left(t,X(t)\right)\cdot b(X(t),\nu_\Theta(t))dt\right] \\ &\quad + E_\Theta\left[\int_{[0,T]\times\mathbb{R}^m} \varphi\left(t,X(t)\right)\cdot\sigma(X(t),\nu_\Theta(t))y\,\rho(dt,dy)\right]. \end{aligned} \tag{5.1.10}$$

Now let

$$\mathcal{P}_2(\mathcal{Z})\doteq\left\{\Theta\in\mathcal{P}(\mathcal{Z}): E_\Theta\left[\int_{[0,T]\times\mathbb{R}^m}|y|^2\,\rho(dt,dy)\right]<\infty\right\},$$

and for $\mathcal{J}\in\mathbf{H}^{-\mathbf{s}}$, define

$$\mathcal{P}^*(\mathcal{J})\doteq\left\{\Theta\in\mathcal{S}(\mathcal{Z})\cap\mathcal{P}_2(\mathcal{Z}):\langle\mathcal{J},\varphi\rangle=G_\varphi(\Theta)\text{ for all }\varphi\in\mathcal{C}_c^\infty\left(U\times\mathbb{R}^d,\mathbb{R}^d\right)\right\}.$$

Define $I : \mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}} \to [0, \infty]$ as

$$I(\mu, \mathcal{J}) \doteq \inf \left\{ E_\Theta \left[\frac{1}{2} \int_{[0,T]\times\mathbb{R}^m} |y|^2 \, \rho(dt, dy) \right] : \Theta_{(1)} = \mu, \Theta \in \mathcal{P}^*(\mathcal{J}) \right\}, \tag{5.1.11}$$

where we recall that $\Theta_{(1)}$ denotes the marginal of Θ on $\mathcal{X}$.

Remark 5.1.1. *Note that the domain of the function I depends on $\mathbf{s} \in \mathcal{O}_d$. However, it turns out (see Lemma 5.3.7) that if $I(\mu, \mathcal{J}) < \infty$ for some $\mathbf{s} \in \mathcal{O}_d$ and $(\mu, \mathcal{J}) \in \mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}}$, then $\mathcal{J} \in \mathbf{H}^{-\mathbf{s}'}$ for all $\mathbf{s}' \in \mathcal{O}_d$, and the value of $I(\mu, \mathcal{J})$ is independent of $\mathbf{s}$.*

5.1.4 Main Results

In this section we present the main results. For each $N \in \mathbb{N}$, let μ^N, V^N and $\mathcal{J}^N$ be as in (5.1.3) , (5.1.2), and Theorem 5.1.1 respectively. Our first main result is a law of large numbers for $(\mu^N, V^N, \mathcal{J}^N)$.

By using the Lipschitz property of b it can be checked that for μ_0 as in Condition 5.1.2 and any $\mathbb{R}^d$ valued random variable ξ_0 on $(\Omega, \mathcal{F}, P)$ with distribution μ_0, there is an a.s. unique solution ξ, with sample paths in $\mathcal{X}$, to the equation

$$\xi(t) = \xi_0 + \int_0^t b\left(\xi(s), V^*(s)\right) \, ds, \qquad V^*(t) = P \circ \xi(t)^{-1}, \; 0 \le t \le T. \tag{5.1.12}$$

Let

$$\mu^* = P \circ \xi^{-1}. \tag{5.1.13}$$

Using the linear growth of b and Condition 5.1.2(ii) it can be checked that $\mu^* \in \mathcal{P}_1(\mathcal{X})$.

The following theorem gives the law of large numbers. Its proof is given in Section 5.2.6.

Theorem 5.1.2 (LLN)**.** *Assume Conditions 5.1.1 and 5.1.2 hold and let $\mathbf{s} \in \mathcal{O}_d$. Then,*

$$\left(\mu^N, V^N, \mathcal{J}^N\right) \xrightarrow{P} (\mu^*, V^*, \mathcal{J}^*) \qquad \text{as } N \to \infty,$$

in $\mathcal{P}_1(\mathcal{X}) \times \mathcal{V} \times \mathbf{H}^{-\mathbf{s}}$, where V^* and μ^* are as in (5.1.12) *and* (5.1.13) *and $\mathcal{J}^*$ is characterized as*

$$\langle \mathcal{J}^*, \varphi \rangle = \int_0^T \langle V^*(t), \varphi(t, \cdot) \cdot b\left(\cdot, V^*(t)\right) \rangle \, dt, \tag{5.1.14}$$

for $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$.

Remark 5.1.2. *The pair $(V^*, \mathcal{J}^*)$ can alternatively be characterized as the unique solution of the equation*

$$\frac{\partial}{\partial t} V + \nabla \cdot b(\cdot, V) V = 0, \qquad \mathcal{J} = b(\cdot, V) V, \qquad V(0) = \mu_0, \tag{5.1.15}$$

in the distributional sense on $(0,T)\times\mathbb{R}^d$*, by which we mean that for all* $\varphi\in\mathcal{C}_c^\infty((0,T)\times\mathbb{R}^d,\mathbb{R})$*,*

$$\int_0^T \left\langle V(t), \frac{\partial}{\partial t}\varphi(t,\cdot)\right\rangle dt + \int_0^T \langle V(t), \nabla\varphi(t,\cdot)\cdot b(\cdot, V(t))\rangle\, dt = 0,$$

and for all $\varphi\in\mathcal{C}_c^\infty((0,T)\times\mathbb{R}^d,\mathbb{R}^d)$*,*

$$\langle \mathcal{J},\varphi\rangle = \int_0^T \langle V(t), \varphi(t,\cdot)\cdot b(\cdot, V(t))\rangle\, dt.$$

Recall the function I defined in (5.1.11), and for each $N\in\mathbb{N}$ let $a_N \doteq N/\varepsilon_N^2$. Our main large deviation result is as follows.

Theorem 5.1.3 (LDP)**.** *Assume Conditions 5.1.1 and 5.1.2 hold. For each* $\mathbf{s}\in\mathcal{O}_d$*,* I *is a rate function on* $\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}$*. Furthermore,*

(i) The sequence $\{(\mu^N,\mathcal{J}^N), N\in\mathbb{N}\}$ *satisfies the large deviation upper bound on* $\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}$ *with speed* a_N *and rate function* I*. Namely, for all closed sets* F *in* $\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}$*,*

$$\limsup_{N\to\infty}\frac{1}{a_N}\log P\left((\mu^N,\mathcal{J}^N)\in F\right) \le -\inf_{(\mu,\mathcal{J})\in F} I(\mu,\mathcal{J}).$$

(ii) If in addition Condition 5.1.3 holds, then $\{(\mu^N,\mathcal{J}^N), N\in\mathbb{N}\}$ *satisfies the large deviation lower bound on* $\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}$ *with speed* a_N *and rate function* I*. Namely, for all open sets* G *in* $\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}$*,*

$$\liminf_{N\to\infty}\frac{1}{a_N}\log P\left((\mu^N,\mathcal{J}^N)\in G\right) \ge -\inf_{(\mu,\mathcal{J})\in G} I(\mu,\mathcal{J}).$$

The proof of Theorem 5.1.3(i) is in Section 5.2.3, and the proof of Theorem 5.1.3(ii) is in Section 5.2.4. The rate function property of I is proved in Section 5.2.5. The proof of Theorem 5.1.2 is saved for Section 5.2.6, since it follows along the lines of the proof of the large deviation upper bound.

It is easy to verify that the map $\nu\mapsto\{t\mapsto\nu\circ\pi_t^{-1}\}$ is a continuous map from $\mathcal{P}_1(\mathcal{X})$ into $\mathcal{V}$, and recall from above that each $\Theta\in\mathcal{S}(\mathcal{Z})$ induces $\nu_\Theta\in\mathcal{V}$. From this and the contraction principle we immediately have a large deviation principle for $\{(\mu^N,V^N,\mathcal{J}^N)\}$. In particular, we have the following corollary. Define $\tilde{I}:\mathcal{V}\times\mathbf{H}^{-\mathbf{s}}\to[0,\infty]$ as

$$\tilde{I}(V,\mathcal{J}) \doteq \inf\left\{E_\Theta\left[\frac{1}{2}\int_{[0,T]\times\mathbb{R}^m}|y|^2\,\rho(dt,dy)\right] : \nu_\Theta = V, \Theta\in\mathcal{P}^*(\mathcal{J})\right\}. \tag{5.1.16}$$

Corollary 5.1.4. *Assume Conditions 5.1.1 and 5.1.2 hold. For each* $\mathbf{s}\in\mathcal{O}_d$*,* $\tilde{I}$ *is a rate function on* $\mathcal{V}\times\mathbf{H}^{-\mathbf{s}}$*. Furthermore,* *(i) The sequence $\{(V^N, \mathcal{J}^N), N \in \mathbb{N}\}$ satisfies the large deviation upper bound on $\mathcal{V} \times \mathbf{H}^{-\mathbf{s}}$ with speed a_N and rate function $\tilde{I}$.*

(ii) If in addition Condition 5.1.3 holds, then $\{(V^N, \mathcal{J}^N), N \in \mathbb{N}\}$ satisfies the large deviation lower bound on $\mathcal{V} \times \mathbf{H}^{-\mathbf{s}}$ with speed a_N and rate function $\tilde{I}$.

When $m = d$ and $\sigma = \mathrm{Id} \in \mathbb{R}^{d\times d}$, one can give a more explicit representation for the rate function $\tilde{I}$ as follows. (A similar representation can be found in [88].) For $V \in \mathcal{V}$, define $\mathbf{V} \in \mathcal{P}([0,T] \times \mathbb{R}^d)$ as $\mathbf{V}(dt, dx) \doteq V(t, dx)\, dt$. Define $\tilde{I}_0 : \mathcal{V} \times \mathbf{H}^{-\mathbf{s}} \to [0, \infty]$ as

$$\tilde{I}_0(V, \mathcal{J}) \doteq \inf \left\{ \frac{1}{2} \int_0^T \left| \langle V(t), h(t, \cdot) - b(\cdot, V(t)) \rangle \right|^2 dt \right\}, \tag{5.1.17}$$

where the infimum is taken over all $(h, \Theta) \in L^2(\mathbf{V}, [0,T] \times \mathbb{R}^d, \mathbb{R}^d) \times (\mathcal{S}(\mathcal{Z}) \cap \mathcal{P}_2(\mathcal{Z}))$ such that $V = \nu_\Theta$ and $(V, \mathcal{J})$ is a distributional-sense solution of the equation

$$\frac{\partial}{\partial t} V + \nabla \cdot hV = 0, \qquad \mathcal{J} = hV, \qquad V(0) = \mu_0, \tag{5.1.18}$$

on $(0,T) \times \mathbb{R}^d$. Namely, for all $\varphi \in \mathcal{C}_c^\infty((0,T) \times \mathbb{R}^d, \mathbb{R})$,

$$\int_0^T \left\langle V(t), \frac{\partial}{\partial t} \varphi(t, \cdot) \right\rangle dt + \int_0^T \langle V(t), \nabla \varphi(t, \cdot) \cdot h(t, \cdot) \rangle \; dt = 0,$$

and for all $\varphi \in \mathcal{C}_c^\infty((0,T) \times \mathbb{R}^d, \mathbb{R}^d)$,

$$\langle \mathcal{J}, \varphi \rangle = \int_0^T \langle V(t), \varphi(t, \cdot) \cdot h(t, \cdot) \rangle \; dt.$$

The following result shows that $\tilde{I} = \tilde{I}_0$. The proof is given in Section 5.2.7.

Proposition 5.1.5. *Suppose that $m = d$, $\sigma = Id \in \mathbb{R}^{d\times d}$, and Condition 5.1.1 is satisfied. Then $\tilde{I} = \tilde{I}_0$.*

5.2 Laplace Asymptotics and Variational Representation

Using the well-known equivalence (cf. [21, 37]) between the large deviation upper bound (resp. lower bound) and the Laplace upper bound (resp. lower bound), we will prove Theorem 5.1.3 by establishing a Laplace principle on the space $\mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}}$. Specifically, Theorem 5.1.3(i) will follow from the upper bound

$$\liminf_{N\to\infty} -\frac{1}{a_N} \log E\left[e^{-a_N F\left(\mu^N, \mathcal{J}^N\right)} \right] \geq \inf_{(\mu, \mathcal{J}) \in \mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}}} \left(F(\mu, \mathcal{J}) + I(\mu, \mathcal{J}) \right), \tag{5.2.1}$$

and Theorem 5.1.3(ii) will follow from the lower bound

$$\limsup_{N\to\infty} -\frac{1}{a_N}\log E\left[e^{-a_N F\left(\mu^N,\mathcal{J}^N\right)}\right] \leq \inf_{(\mu,\mathcal{J})\in\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}} \left(F(\mu,\mathcal{J})+I(\mu,\mathcal{J})\right), \tag{5.2.2}$$

where F is any bounded, continuous function on $\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}$.

The inequality (5.2.1) will be proved in Section 5.2.3 (under Conditions 5.1.1 and 5.1.2), and the inequality (5.2.2) will be proved in Section 5.2.4 (under Conditions 5.1.1, 5.1.2, and 5.1.3). The rate function property of I is shown in Section 5.2.5. The starting point for both upper and lower bounds is the following variational representation.

5.2.1 Variational Representation

Let $\mathcal{A}_N$ denote the class of $\mathbb{R}^{Nm}$-valued $\mathcal{F}(t)$-progressively measurable processes u such that $E\left[\int_0^T |u(t)|^2\,dt\right]<\infty$. For $u^N=(u_1^N,\ldots,u_N^N)\in\mathcal{A}_N$, with each $u_j^N(t)$ taking values in $\mathbb{R}^m$, consider the controlled version of (5.1.1) given as

$$d\bar{X}_j^N(t)=b\left(\bar{X}_j^N(t),\bar{V}^N(t)\right)\,dt+\varepsilon_N\sigma\left(\bar{X}_j^N(t),\bar{V}^N(t)\right)\,dW_j(t)+\sigma\left(\bar{X}_j^N(t),\bar{V}^N(t)\right)u_j^N(t)\,dt, \tag{5.2.3}$$

where $\bar{X}_j^N(0)=x_j^N$ and

$$\bar{V}^N(t)\doteq\frac{1}{N}\sum_{j=1}^N\delta_{\bar{X}_j^N(t)},\qquad 0\le t\le T.$$

Analogous to (5.1.3), $\bar{\mu}^N$ will denote the empirical measure of $(\bar{X}_1^N,\ldots,\bar{X}_N^N)$, so that $\bar{\mu}^N\circ\pi_t^{-1}=\bar{V}^N(t)$ for each $0\le t\le T$. We will also need a controlled analogue of the stochastic current in Theorem 5.1.1. For $\varphi\in\mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$, define

$$\bar{J}_j^N(\varphi)\doteq\int_0^T\varphi\left(t,\bar{X}_j^N(t)\right)\circ d\bar{X}_j^N(t),\qquad \bar{J}^N(\varphi)\doteq\frac{1}{N}\sum_{j=1}^N\bar{J}_j^N(\varphi). \tag{5.2.4}$$

The proof of the following result, which is given in the Appendix, is similar to that of Theorem 5.1.1.

Lemma 5.2.1. *Suppose that Conditions 5.1.1 and 5.1.2 hold. Then, for each $N\in\mathbb{N}$, $1\le j\le N$, and $\mathbf{s}\in\mathcal{O}_d$, there is a nonnegative square-integrable random variable $C_{j,\mathbf{s}}^N$ such that for all $\varphi\in\mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$,*

$$\left|\bar{J}_j^N(\varphi)\right|\le C_{j,\mathbf{s}}^N\|\varphi\|_{\mathbf{s}}\qquad \textit{a.s.}$$

In particular, the collection $\{\varphi \mapsto \bar{J}^N(\varphi)\}$ *has a pathwise realization* $\bar{\mathcal{J}}^N$ *on* $(\Omega, \mathcal{F}, P)$*, namely* $\bar{\mathcal{J}}^N$ *is an* $\mathbf{H}^{-\mathbf{s}}$*-valued random variable such that* $\langle \bar{\mathcal{J}}^N(\omega), \varphi \rangle = [\bar{J}^N(\varphi)](\omega)$ *for a.e.* $\omega \in \Omega$ *and all* $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$*. Furthermore, if*

$$\sup_{N \geq 1} E\left[\frac{1}{N}\sum_{j=1}^N \int_0^T \left|u_j^N(t)\right|^2 \, dt\right] < \infty, \tag{5.2.5}$$

then $\sup_{N\geq 1} E\left[\frac{1}{N}\sum_{j=1}^N \left(C_{j,\mathbf{s}}^N\right)^2\right] < \infty$*. In particular, if* $C_{\mathbf{s}}^N \doteq \frac{1}{N}\sum_{j=1}^N C_{j,\mathbf{s}}^N$*, then* $\sup_{N\geq 1} E\left[\left(C_{\mathbf{s}}^N\right)^2\right] < \infty$*.*

The following variational representation follows from [17, 20] (see also [22]). Specifically, the case where $\{\mathcal{F}(t)\}$ is the filtration generated by the m-dimensional Brownian motions $\{W_j\}$ is covered in [17], while the setting of a general filtration is treated in [20]. Recall that $a_N = N/\varepsilon_N^2$.

Theorem 5.2.2 (Variational Representation)**.** *Suppose that Conditions 5.1.1 and 5.1.2 hold. Let* $\mathbf{s} \in \mathcal{O}_d$ *and let* F *be a real-valued, bounded, continuous function on* $\mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}}$*. Then for each* $N \in \mathbb{N}$*,*

$$-\frac{1}{a_N}\log E\left[e^{-a_N F\left(\mu^N, \mathcal{J}^N\right)}\right] = \inf_{u^N \in \mathcal{A}_N} E\left[\frac{1}{2N}\sum_{j=1}^N \int_0^T \left|u_j^N(t)\right|^2 \, dt + F\left(\bar{\mu}^N, \bar{\mathcal{J}}^N\right)\right]. \tag{5.2.6}$$

5.2.2 Tightness Properties.

The following lemma gives a key tightness property that will be needed in the proofs of both upper and lower Laplace bounds. The proof is given in Section 5.3.1.

Lemma 5.2.3. *Suppose Conditions 5.1.1 and 5.1.2 hold. Fix* $\mathbf{s} \in \mathcal{O}_d$*, and let* $\{u^N, N \in \mathbb{N}\}$ *with* $u_N \in \mathcal{A}_N$ *for each* N *be such that*

$$\sup_{N \geq 1} E\left[\frac{1}{N}\sum_{j=1}^N \int_0^T \left|u_j^N(t)\right|^2 \, dt\right] < \infty.$$

Let $\bar{X}_j^N$*,* $\bar{\mu}^N$*, and* $\bar{\mathcal{J}}^N$ *be the controlled sequences corresponding to sequence of controls* $\{u^N\}$ *as defined in Section 5.2.1. For each* j *and* N*, let* ρ_j^N *be the* $\mathcal{R}_1$*-valued random variable given as*

$$\rho_j^N(dt, dy) \doteq \delta_{u_j^N(t)}(dy)\, dt,$$

and consider the sequence of $\mathcal{P}(\mathcal{Z})$*-valued random variables defined as*

$$Q^N \doteq \frac{1}{N}\sum_{j=1}^N \delta_{\left(\bar{X}_j^N, \rho_j^N\right)}, \qquad N \in \mathbb{N}. \tag{5.2.7}$$

Then,

(i) The sequence $\{(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N), N \in \mathbb{N}\}$ *is tight in* $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z}) \times \mathbf{H}^{-\mathbf{s}}$*,*

(ii) If $(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N) \Rightarrow (\bar{\mu}, Q, \bar{\mathcal{J}})$ *as* $N \to \infty$ *in* $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z}) \times \mathbf{H}^{-\mathbf{s}}$*, then* $Q_{(1)} = \bar{\mu}$ *and* $Q \in \mathcal{P}^*(\bar{\mathcal{J}})$ *a.s.*

5.2.3 Proof of the Upper Bound

In this section we prove part (i) of Theorem 5.1.3 by showing that (5.2.1) holds. Assume Conditions 5.1.1 and 5.1.2. Fix $\mathbf{s} = (s_1, s_2) \in \mathcal{O}_d$, and a real-valued, bounded, continuous function F on $\mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}}$. Let $\varepsilon \in (0,1)$, and using Theorem 5.2.2 choose $\{u^N, N \in \mathbb{N}\}$ with $u_N \in \mathcal{A}_N$ for each N such that

$$-\frac{1}{a_N} \log E\left[e^{-a_N F\left(\mu^N, \mathcal{J}^N\right)}\right] \geq E\left[\frac{1}{2N}\sum_{j=1}^N \int_0^T \left|u_j^N(t)\right|^2 dt + F\left(\bar{\mu}^N, \bar{\mathcal{J}}^N\right)\right] - \varepsilon, \tag{5.2.8}$$

where $(\bar{\mu}^N, \bar{\mathcal{J}}^N)$ are controlled variables corresponding to the control u^N as defined in Section 5.2.1. From the boundedness of F it follows that

$$\sup_{N\geq 1} E\left[\frac{1}{2N}\sum_{j=1}^N \int_0^T \left|u_j^N(t)\right|^2 dt\right] \leq 2 \sup_{(\mu,\mathcal{J})\in\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}} |F(\mu,\mathcal{J})| + 1 < \infty.$$

By Lemma 5.2.3, $(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N)$ is tight in $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z}) \times \mathbf{H}^{-\mathbf{s}}$. Thus the sequence $(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N)$ has a weak limit point $(\bar{\mu}, Q, \bar{\mathcal{J}})$ along some subsequence, and once again by Lemma 5.2.3, $Q \in \mathcal{P}^*(\mathcal{J})$ and $Q_{(1)} = \bar{\mu}$ a.s. Assume without loss of generality that $(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N) \Rightarrow (\bar{\mu}, Q, \bar{\mathcal{J}})$ along the full sequence. Noting that $Q^N_{(1)} = \bar{\mu}^N$, we have, by (5.2.8),

$$-\frac{1}{a_N} \log E\left[e^{-a_N F\left(\mu^N, \mathcal{J}^N\right)}\right] \geq E\left[\frac{1}{2}\int_{\mathcal{R}_1}\int_{[0,T]\times\mathbb{R}^m} |y|^2\, r(dt,dy)\, Q^N_{(2)}(dr) + F\left(Q^N_{(1)}, \bar{\mathcal{J}}^N\right)\right] - \varepsilon.$$

By Fatou's lemma and lower semicontinuity of the map $r \mapsto \int_{[0,T]\times\mathbb{R}^m} |y|^2\, r(dt,dy)$ on $\mathcal{R}_1$,

$$\begin{aligned}
&\liminf_{N\to\infty} -\frac{1}{a_N} \log E\left[e^{-a_N F\left(\mu^N, \mathcal{J}^N\right)}\right] \\
&\geq E\left[\frac{1}{2}\int_{\mathcal{R}_1}\int_{[0,T]\times\mathbb{R}^m} |y|^2\, r(dt,dy)\, Q_{(2)}(dr) + F\left(Q_{(1)}, \bar{\mathcal{J}}\right)\right] - \varepsilon \\
&= E\left[E_Q\left[\frac{1}{2}\int_{[0,T]\times\mathbb{R}^m} |y|^2\, \rho(dt,dy)\right] + F\left(\bar{\mu}, \bar{\mathcal{J}}\right)\right] - \varepsilon \\
&\geq \inf_{(\mu,\mathcal{J})\in\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}} \left(I(\mu,\mathcal{J}) + F(\mu,\mathcal{J})\right) - \varepsilon,
\end{aligned}$$

where the last line follows on recalling the definition of I and the facts that $Q \in \mathcal{P}^*(\mathcal{J})$ and $Q_{(1)} = \bar{\mu}$ a.s. Since $\varepsilon \in (0,1)$ is arbitrary, this completes the proof of the upper bound in (5.2.1) and thus that of Theorem 5.1.3(i). $\square$

5.2.4 Proof of the Lower Bound

In this section we prove part (ii) of Theorem 5.1.3 by showing (5.2.2). Fix $\mathbf{s} = (s_1, s_2) \in \mathcal{O}_d$. We assume Conditions 5.1.1, 5.1.2, and 5.1.3 hold. Let $\varepsilon \in (0,1)$ and choose $(\Theta_0, \mathcal{J}_0) \in \mathcal{P}(\mathcal{Z}) \times \mathbf{H}^{-\mathbf{s}}$ such that $\Theta_0 \in \mathcal{P}^*(\mathcal{J}_0)$ and

$$E_{\Theta_0}\left[\frac{1}{2}\int_{[0,T]\times\mathbb{R}^m} |y|^2 \rho(dt,dy)\right] + F\left((\Theta_0)_{(1)}, \mathcal{J}_0\right) \le \inf_{(\mu,\mathcal{J})\in\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}} (I(\mu,\mathcal{J}) + F(\mu,\mathcal{J})) + \varepsilon. \tag{5.2.9}$$

To prove the lower bound we will construct a sequence $\{u^N\}$ of controls on some filtered probability space such that $u^N \in \mathcal{A}_N$ for each N and

$$\begin{aligned}&\limsup_{N\to\infty} E\left[\frac{1}{2N}\sum_{j=1}^N \int_0^T \left|u_j^N(t)\right|^2 dt + F\left(\bar{\mu}^N, \bar{\mathcal{J}}^N\right)\right]\\ &\le E_{\Theta_0}\left[\frac{1}{2}\int_{[0,T]\times\mathbb{R}^m} |y|^2 \rho(dt,dy)\right] + F\left((\Theta_0)_{(1)}, \mathcal{J}_0\right),\end{aligned} \tag{5.2.10}$$

where $\bar{\mu}^N$ and $\bar{\mathcal{J}}^N$ are the controlled processes corresponding to $\{u^N\}$. It will then follow by Theorem 5.2.2 and (5.2.9) that

$$\begin{aligned}\limsup_{N\to\infty} -\frac{1}{a_N}\log E\left[e^{-a_N F\left(\mu^N,\mathcal{J}^N\right)}\right] &\le \limsup_{N\to\infty} E\left[\frac{1}{2N}\sum_{j=1}^N \int_0^T \left|u_j^N(t)\right|^2 dt + F\left(\bar{\mu}^N, \bar{\mathcal{J}}^N\right)\right]\\ &\le \inf_{(\mu,\mathcal{J})\in\mathcal{P}_1(\mathcal{X})\times\mathbf{H}^{-\mathbf{s}}} (I(\mu,\mathcal{J}) + F(\mu,\mathcal{J})) + \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, the lower bound follows.

The construction of a sequence $\{u^N\}$ such that the inequality in (5.2.10) holds will need the following uniqueness property.

Definition 5.2.1. *Let $\theta : \mathcal{Z} \to \mathbb{R}^d \times \mathcal{R}_1$ denote the map $\theta(\xi, r) = (\xi(0), r)$. We say that weak uniqueness of solutions of* (5.1.9) *holds if $\Theta_1, \Theta_2 \in \mathcal{S}(\mathcal{Z}) \cap \mathcal{P}_2(\mathcal{Z})$ and $\Theta_1 \circ \theta^{-1} = \Theta_2 \circ \theta^{-1}$ implies that $\Theta_1 = \Theta_2$.*

The following lemma is key to the proof of the lower bound. The proof is provided in Section 5.3.3. Recall that in this section we assume that Conditions 5.1.1, 5.1.2 and 5.1.3 hold.

Lemma 5.2.4. *Weak uniqueness of solutions holds for* (5.1.9).

We now construct the sequence $\{u^N\}$ that satisfies (5.2.10). Because $\Theta_0 \in \mathcal{S}(\mathcal{Z})$, we can disintegrate

$$\Theta_0 \circ \theta^{-1}(dx\,dr) = \mu_0(dx)\,\Lambda_0(x, dr),$$

for some measurable map $\Lambda_0 : \mathbb{R}^d \to \mathcal{P}(\mathcal{R}_1)$. Let $\mathcal{W} \doteq \mathcal{C}([0,T],\mathbb{R}^m)$, and let γ be the standard Wiener measure on $\mathcal{W}$. Define a measurable map $\Lambda : \mathbb{R}^d \to \mathcal{P}(\mathcal{R}_1 \times \mathcal{W})$ as

$$\Lambda(x, dr, dw) \doteq \Lambda_0(x, dr) \otimes \gamma(dw), \qquad x \in \mathbb{R}^d.$$

Define the measurable space $(\tilde{\Omega}, \tilde{\mathcal{F}})$ by

$$\tilde{\Omega} = (\mathcal{R}_1 \times \mathcal{W})^\infty, \quad \tilde{\mathcal{F}} = \mathcal{B}\big(\tilde{\Omega}\big),$$

where an element $(r, w) \in \tilde{\Omega}$ has the coordinates $r = (r_1, r_2, \ldots)$ and $w = (w_1, w_2, \ldots)$ with $r_j \in \mathcal{R}_1$ and $w_j \in \mathcal{W}$ for each j. Consider the canonical filtration $\{\tilde{\mathcal{F}}(t)\}$ on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ defined as

$$\tilde{\mathcal{F}}(t) \doteq \sigma\left(w_j(s), r_j([0,s] \times A),\, j \in \mathbb{N},\, A \in \mathcal{B}(\mathbb{R}^m),\, s \le t\right), \qquad 0 \le t \le T,$$

and define the sequence $\{P^N, N \in \mathbb{N}\}$ of probability measures on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ by

$$P^N(dr, dw) = \bigotimes_{j \le N} \Lambda\left(x_j^N, dr_j, dw_j\right) \bigotimes_{j > N} \big((\Theta_0)_{(2)} \otimes \gamma\big)\,(dr_j, dw_j),$$

where $\{x_j^N\}$ are as in Condition 5.1.2. Next define the sequence $\{\Lambda^N, N \in \mathbb{N}\}$ of $\mathcal{P}(\mathbb{R}^d \times \mathcal{R}_1)$-valued random variables on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ by

$$\Lambda^N \doteq \frac{1}{N} \sum_{j=1}^N \delta_{(x_j^N, \rho_j)},$$

where for each $j \in \mathbb{N}$, ρ_j is the $\mathcal{R}_1$-valued random variable on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ defined as $\rho_j(r, w) = r_j$. Using Condition 5.1.3(ii), we see by a standard argument that

$$P^N \circ (\Lambda^N)^{-1} \to \delta_{\Theta_0 \circ \theta^{-1}} \qquad \text{as } N \to \infty, \tag{5.2.11}$$

in $\mathcal{P}(\mathcal{P}(\mathbb{R}^d \times \mathcal{R}_1))$.

Now, for each $j \in \mathbb{N}$, disintegrating ρ_j as $\rho_j(dt, dy) = (\rho_j)_t(dy)\, dt$, define

$$u_j(t) \doteq \int_{\mathbb{R}^m} y\, (\rho_j)_t(dy), \qquad 0 \le t \le T,$$

and define $u^N \doteq (u_1, \ldots, u_N)$ for each $N \in \mathbb{N}$. Furthermore, for each j and $(r, w) \in \tilde{\Omega}$, let

$$W_j(t, (r, w)) \doteq w_j(t), \qquad 0 \le t \le T.$$

Then for each N, $W_1, \ldots, W_N$ are mutually independent $\{\tilde{\mathcal{F}}(t)\}$-Brownian motions on $(\tilde{\Omega}, \tilde{\mathcal{F}}, P^N)$. Recall that in this section we are assuming Condition 5.1.3, and so $\sigma(x,\nu) = \sigma(\nu)$ for $(x,\nu) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$. Let $(\bar{X}_1^N, \ldots, \bar{X}_N^N)$ be the unique pathwise solution (which is guaranteed due to Conditions 5.1.1 and 5.1.2) on $(\tilde{\Omega}, \tilde{\mathcal{F}}, P^N)$ of the system

$$\begin{aligned}
\bar{X}_j^N(t) &= x_j^N + \int_0^t b\left(\bar{X}_j^N(s), \bar{V}^N(s)\right)\, ds + \varepsilon_N \int_0^t \sigma\left(\bar{V}^N(s)\right)\, dW_j(s) + \int_0^t \sigma\left(\bar{V}^N(s)\right) u_j(s)\, ds, \\
\bar{V}^N(t) &= \frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}_j^N(t)}, \qquad 0 \le t \le T, \qquad 1 \le j \le N.
\end{aligned}$$

Also let $\bar{\mu}^N = \frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}_j^N}$. Now define the sequence $\{Q^N\}$ of $\mathcal{P}(\mathcal{Z})$-valued random variables as

$$Q^N \doteq \frac{1}{N}\sum_{j=1}^N \delta_{\left(\bar{X}_j^N, \rho_j\right)}, \qquad N \in \mathbb{N}.$$

Letting E^N denote expectation on $(\tilde{\Omega}, \tilde{\mathcal{F}}, P^N)$, we note that for a measurable $f : \mathcal{R}_1 \to \mathbb{R}_+$,

$$\int_{\mathcal{R}_1} f(r)\,(\Theta_0)_{(2)}(dr) < \infty \qquad \text{implies} \qquad E^N\left[\frac{1}{N}\sum_{j=1}^N f(\rho_j)\right] \to \int_{\mathcal{R}_1} f(r)\,(\Theta_0)_{(2)}(dr). \tag{5.2.12}$$

Indeed, if $g(x) = \int_{\mathcal{R}_1} f(r)\,\Lambda_0(x,dr)$ for $x \in \mathbb{R}^d$, then

$$E^N\left[\frac{1}{N}\sum_{j=1}^N f(\rho_j)\right] = \frac{1}{N}\sum_{j=1}^N \int_{\mathcal{R}_1} f(r)\,\Lambda_0\left(x_j^N, dr\right) = \frac{1}{N}\sum_{j=1}^N g\left(x_j^N\right),$$

and

$$\begin{aligned}
\int_{\mathbb{R}^d} g(x)\,\mu_0(dx) &= \int_{\mathbb{R}^d}\int_{\mathcal{R}_1} f(r)\,\Lambda_0(x,dr)\,\mu_0(dx) \\
&= \int_{\mathbb{R}^d \times \mathcal{R}_1} f(r)\,\Theta_0 \circ \theta^{-1}(dx,dr) = \int_{\mathcal{R}_1} f(r)\,(\Theta_0)_{(2)}(dr) < \infty.
\end{aligned}$$

Thus, from Condition 5.1.3(ii),

$$\lim_{N\to\infty} E^N\left[\frac{1}{N}\sum_{j=1}^N f(\rho_j)\right] = \lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^N g\left(x_j^N\right) = \int_{\mathbb{R}^d} g(x)\,\mu_0(dx) = \int_{\mathcal{R}_1} f(r)\,(\Theta_0)_{(2)}(dr),$$

which proves (5.2.12). Now, we have

$$\limsup_{N\to\infty} E^N\left[\frac{1}{N}\sum_{j=1}^N\int_0^T |u_j(t)|^2\, dt\right] \le \limsup_{N\to\infty} E^N\left[\frac{1}{N}\sum_{j=1}^N\int_{[0,T]\times\mathbb{R}^m} |y|^2\,\rho_j(dt,dy)\right] \\ = E_{\Theta_0}\left[\int_{[0,T]\times\mathbb{R}^m}|y|^2\,\rho(dt,dy)\right] < \infty, \tag{5.2.13}$$

where the convergence on the second line follows from (5.2.12) on observing that, since $\Theta_0 \in \mathcal{P}_2(\mathcal{Z})$,

$$f(r) = \int_{[0,T]\times\mathbb{R}^m} |y|^2\, r(dt,dy), \qquad r\in\mathcal{R}_1,$$

satisfies

$$\int_{\mathcal{R}_1} f(r)\,(\Theta_0)_{(2)}(dr) = E_{\Theta_0}\left[\int_{[0,T]\times\mathbb{R}^m}|y|^2\,\rho(dt,dy)\right] < \infty.$$

Next, for each $\varphi\in\mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$ define

$$\bar{J}^N(\varphi) \doteq \frac{1}{N}\sum_{j=1}^N\int_0^T \varphi\left(t,\bar{X}_j^N(t)\right)\circ d\bar{X}_j^N(t), \qquad N\in\mathbb{N}.$$

From Lemma 5.2.1, the collection $\{\varphi\mapsto\bar{J}^N(\varphi)\}$ has a pathwise realization $\bar{\mathcal{J}}^N$ in $\mathbf{H}^{-\mathbf{s}}$. Using Lemma 5.2.3 and the moment bound in (5.2.13), we now see that $\{(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N), N\in\mathbb{N}\}$ is tight in $\mathcal{P}_1(\mathcal{X})\times\mathcal{P}(\mathcal{Z})\times\mathbf{H}^{-\mathbf{s}}$. Suppose, without loss of generality, that $(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N)\Rightarrow(\bar{\mu}, Q, \bar{\mathcal{J}})$ in $\mathcal{P}_1(\mathcal{X})\times\mathcal{P}(\mathcal{Z})\times\mathbf{H}^{-\mathbf{s}}$. By Lemma 5.2.3 again, $Q\in\mathcal{P}^*(\bar{\mathcal{J}})$ and $Q_{(1)}=\bar{\mu}$ a.s. Since $Q^N\circ\theta^{-1}=\Lambda^N$, (5.2.11) implies that $Q\circ\theta^{-1}=\Theta_0\circ\theta^{-1}$ a.s., and hence by the weak uniqueness established in Lemma 5.2.4, $Q=\Theta_0$ a.s. Furthermore, from the definition of $\mathcal{P}^*(\bar{\mathcal{J}})$,

$$\langle\bar{\mathcal{J}},\varphi\rangle = G_\varphi(Q) = G_\varphi(\Theta_0) = \langle\mathcal{J}_0,\varphi\rangle$$

for every φ, a.s., and hence $\bar{\mathcal{J}}=\mathcal{J}_0$ a.s. by separability of $\mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$ and its denseness in $\mathbf{H}^{-\mathbf{s}}$.

It follows that $(Q^N,\bar{\mathcal{J}}^N)\Rightarrow(\Theta_0,\mathcal{J}_0)$. Finally,

$$\limsup_{N\to\infty} E^N\left[\frac{1}{2N}\sum_{j=1}^N\int_0^T|u_j(t)|^2\,dt + F\left(\bar{\mu}^N,\bar{\mathcal{J}}^N\right)\right] \\ = \limsup_{N\to\infty} E^N\left[\frac{1}{2N}\sum_{j=1}^N\int_0^T|u_j(t)|^2\,dt + F\left(Q^N_{(1)},\bar{\mathcal{J}}^N\right)\right] \\ \le E_{\Theta_0}\left[\frac{1}{2}\int_{[0,T]\times\mathbb{R}^m}|y|^2\,\rho(dt,dy)\right] + F\left((\Theta_0)_{(1)},\mathcal{J}_0\right),$$

where the last inequality is from (5.2.13) and since F is a bounded continuous function. This shows (5.2.10) and completes the proof of the lower bound in (5.2.2), and part (ii) of Theorem 5.1.3 follows. $\square$

5.2.5 Rate Function Property

In this section we show that the function $I : \mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}} \to [0, \infty]$ defined in (5.1.11) has compact sublevel sets for every $\mathbf{s} \in \mathcal{O}_d$. Fix $\mathbf{s}$, and for each $l < \infty$ consider the level set $\Gamma_l \doteq \{(\mu, \mathcal{J}) \in \mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}} : I(\mu, \mathcal{J}) \leq l\}$. The proof of the following lemma is given in Section 5.3.2.

Lemma 5.2.5. *Suppose Conditions 5.1.1 and 5.1.2 hold. Let* $\mathbf{s} \in \mathcal{O}_d$ *and let* $\{(\mu_k, \Theta_k, \mathcal{J}_k), k \in \mathbb{N}\}$ *be a sequence in* $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z}) \times \mathbf{H}^{-\mathbf{s}}$ *such that for each* k*,* $\Theta_k \in \mathcal{P}^*(\mathcal{J}_k)$*,* $(\Theta_k)_{(1)} = \mu_k$*, and*

$$\sup_{k \geq 1} E_{\Theta_k} \left[\frac{1}{2} \int_{[0,T] \times \mathbb{R}^m} |y|^2 \, \rho(dt, dy) \right] < \infty. \tag{5.2.14}$$

Then the sequence $\{(\mu_k, \Theta_k, \mathcal{J}_k), k \in \mathbb{N}\}$ *is relatively compact in* $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z}) \times \mathbf{H}^{-\mathbf{s}}$.

Now we prove the compactness of Γ_l. Let $\{(\mu_k, \mathcal{J}_k), k \in \mathbb{N}\}$ be a sequence in Γ_l. From the definition of I, for each $k \in \mathbb{N}$ there is a $\Theta_k \in \mathcal{P}^*(\mathcal{J}_k)$ with $(\Theta_k)_{(1)} = \mu_k$ such that

$$E_{\Theta_k} \left[\frac{1}{2} \int_{[0,T] \times \mathbb{R}^m} |y|^2 \, \rho(dt, dy) \right] \leq l + \frac{1}{k}. \tag{5.2.15}$$

From Lemma 5.2.5, $\{(\mu_k, \Theta_k, \mathcal{J}_k)\}$ is relatively compact in $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z}) \times \mathbf{H}^{-\mathbf{s}}$. It is easily checked that if $(\mu, \Theta, \mathcal{J})$ is a limit point along some subsequence, then $\Theta_{(1)} = \mu$ and along the same subsequence $G_\varphi(\Theta_k) \to G_\varphi(\Theta)$ and $\langle \mathcal{J}_k, \varphi \rangle \to \langle \mathcal{J}, \varphi \rangle$ for every $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$. This shows that $\Theta \in \mathcal{P}^*(\mathcal{J})$. Sending $k \to \infty$ in (5.2.15) and using lower semicontinuity of the map $r \mapsto \int_{[0,T] \times \mathbb{R}^m} |y|^2 \, r(dt, dy)$ on $\mathcal{R}_1$, we obtain

$$E_\Theta \left[\frac{1}{2} \int_{[0,T] \times \mathbb{R}^m} |y|^2 \, \rho(dt, dy) \right] \leq l,$$

and hence $(\mu, \mathcal{J})$ lies in Γ_l. Compactness of Γ_l follows. $\square$

5.2.6 Law of Large Numbers

Here we prove Theorem 5.1.2. The model (5.1.1) can be viewed as the controlled equation (5.2.3) with the controls taken to be $u_j^N \equiv 0$ for all $1 \leq j \leq N$ and $N \in \mathbb{N}$. From Lemma 5.2.3 it then follows that $(\mu^N, Q^N, \mathcal{J}^N)$ is tight in $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z}) \times \mathbf{H}^{-\mathbf{s}}$. Suppose that along some subsequence $(\mu^N, Q^N, \mathcal{J}^N) \Rightarrow (\mu, Q, \mathcal{J})$. Then, once again from Lemma 5.2.3, $Q_{(1)} = \mu$ and $Q \in \mathcal{P}^*(\mathcal{J})$ a.s. Furthermore, since $u_j^N \equiv 0$ for all $1 \leq j \leq N$ and $N \in \mathbb{N}$ we see that the second coordinate variable on $\mathcal{Z}$ satisfies $Q(\rho = 0) = 1$ a.s., and thus, under Q, the first coordinate variable on $\mathcal{Z}$ satisfies

$$X(t) = X(0) + \int_0^t b(X(s), V(s))\, ds, \qquad V(t) = Q \circ X(t)^{-1}, \qquad V(0) = \mu_0,$$

for all $0 \le t \le T$. Then, from the unique solvability of (5.1.12), it follows that $\mu = \mu^*$ a.s., and hence we have that μ^N converges in probability in $\mathcal{P}_1(\mathcal{X})$ (along the full sequence) to μ^*. Since $V^N(t) = \mu^N \circ \pi_t^{-1}$ and $V^*(t) = \mu^* \circ \pi_t^{-1}$ for each $0 \le t \le T$, we also have that $V^N \to V^*$ in probability in $\mathcal{V}$. Finally, since $Q \in \mathcal{P}^*(\mathcal{J})$ a.s.,

$$G_\varphi(Q) = \langle \mathcal{J}, \varphi \rangle$$

for all $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$, a.s., and note that

$$\begin{aligned} G_\varphi(Q) &= E_Q\left[\int_0^T \varphi\left(t, X(t)\right) \cdot dX(t)\right] \\ &= E_Q\left[\int_0^T \varphi\left(t, X(t)\right) \cdot b(X(t), V^*(t))\, dt\right] \\ &= \int_0^T \langle V^*(t), \varphi(t, \cdot) \cdot b\left(\cdot, V^*(t)\right)\rangle \, dt. \end{aligned}$$

Thus $\langle \mathcal{J}, \varphi \rangle$ is (a.s.) uniquely characterized for all $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$. From the separability of $\mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$ and its denseness in $\mathbf{H}^{-\mathbf{s}}$ we now see that $\mathcal{J}^N$ converges (along the full sequence) in probability, in $\mathbf{H}^{-\mathbf{s}}$, to the nonrandom limit $\mathcal{J}^*$ characterized as

$$\langle \mathcal{J}^*, \varphi \rangle = \int_0^T \langle V^*(t), \varphi(t, \cdot) \cdot b\left(\cdot, V^*(t)\right)\rangle \, dt.$$

The result follows. □

5.2.7 Equivalent Formulation of the Rate Function

In this section we give the proof of Proposition 5.1.5. Let $m = d$ and $\sigma = \mathrm{Id}$. We first argue that $\tilde{I}_0 \le \tilde{I}$. Fix $(V, \mathcal{J}) \in \mathcal{V} \times \mathbf{H}^{-\mathbf{s}}$ such that $\tilde{I}(V, \mathcal{J}) < \infty$. Fix $\delta > 0$ and let $\Theta \in \mathcal{P}^*(\mathcal{J})$ with $\nu_\Theta = V$ be δ-optimal for $\tilde{I}(V, \mathcal{J})$, namely

$$E_\Theta\left[\frac{1}{2}\int_{[0,T]\times\mathbb{R}^d} |y|^2\, \rho(dt, dy)\right] \le \tilde{I}(V, \mathcal{J}) + \delta. \tag{5.2.16}$$

Disintegrate $\rho(dt, dy) = \rho_t(dy)\, dt$ and define

$$v(t) \doteq \int_{\mathbb{R}^d} y\, \rho_t(dy), \qquad 0 \le t \le T. \tag{5.2.17}$$

Also let $\eta_t \doteq \Theta \circ (X(t), v(t))^{-1} \in \mathcal{P}(\mathbb{R}^{2d})$. Then η_t can be disintegrated as $\eta_t(dx, dy) = V(t, dx)\, \hat{\eta}_t(x, dy)$ for some $\hat{\eta}_t : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^d)$. Define the functions u and h on $[0, T] \times \mathbb{R}^d$ by

$$u(t, x) \doteq \int_{\mathbb{R}^d} y\, \hat{\eta}_t(x, dy), \qquad (5.2.18)$$
$$h(t, x) \doteq u(t, x) + b(x, V(t)).$$

It is easily verified that $h \in L^2(\mathbf{V}, [0, T] \times \mathbb{R}^d, \mathbb{R}^d)$. Under Θ, $V(0) = \mu_0$ and

$$X(t) = X(0) + \int_0^t b(X(s), V(s))\, ds + \int_{[0,t]\times\mathbb{R}^d} y\, \rho_s(dy)\, ds, \qquad \text{a.s.}, \qquad (5.2.19)$$

for each t, and so for $\varphi \in \mathcal{C}_c^\infty((0, T) \times \mathbb{R}^d, \mathbb{R})$,

$$\begin{aligned} 0 &= \varphi(T, X(T)) - \varphi(0, X(0)) \\ &= \int_0^T \left(\frac{\partial}{\partial t}\varphi(t, X(t)) + \nabla\varphi(t, X(t)) \cdot b(X(t), V(t)) + \nabla\varphi(t, X(t)) \cdot v(t) \right) dt, \end{aligned}$$

where v is as in (5.2.17). Taking expectations with respect to Θ,

$$\begin{aligned} 0 &= \int_0^T \left\langle V(t), \frac{\partial}{\partial t}\varphi(t, \cdot) + \nabla\varphi(t, \cdot) \cdot b(\cdot, V(t)) \right\rangle dt + \int_0^T \int_{\mathbb{R}^{2d}} \nabla\varphi(t, x) \cdot y\, \eta_t(dx, dy)\, dt \\ &= \int_0^T \left\langle V(t), \frac{\partial}{\partial t}\varphi(t, \cdot) + \nabla\varphi(t, \cdot) \cdot (b(\cdot, V(t)) + u(t, \cdot)) \right\rangle dt \qquad (5.2.20) \\ &= \int_0^T \left\langle V(t), \frac{\partial}{\partial t}\varphi(t, \cdot) + \nabla\varphi(t, \cdot) \cdot h(t, \cdot) \right\rangle dt. \end{aligned}$$

Similarly it is seen that for $\varphi \in \mathcal{C}_c^\infty((0, T) \times \mathbb{R}^d, \mathbb{R}^d)$,

$$\langle \mathcal{J}, \varphi \rangle = \int_0^T \langle V(t), \varphi(t, \cdot) \cdot h(\cdot, t) \rangle \; dt.$$

Since $V = \nu_\Theta$ and $u(t, x) = h(t, x) - b(x, V(t))$, we now see from the above two identities that

$$\begin{aligned} \tilde{I}_0(V, \mathcal{J}) &\le \frac{1}{2} \int_0^T |\langle V(t), u(t, \cdot) \rangle|^2 \; dt = \frac{1}{2} \int_0^T \left| \int_{\mathbb{R}^{2d}} y\, \hat{\eta}_t(x, dy)\, V(t, dx) \right|^2 dt \\ &\le \frac{1}{2} \int_0^T E_\Theta \left[|v(t)|^2 \right] \, dt \le \frac{1}{2} E_\Theta \left[\int_{[0,T]\times\mathbb{R}^d} |y|^2\, \rho(dt, dy) \right] \le \tilde{I}(V, \mathcal{J}) + \delta, \end{aligned}$$

where the last inequality is from (5.2.16). Since $\delta > 0$ is arbitrary, the inequality $\tilde{I}_0(V, \mathcal{J}) \le \tilde{I}(V, \mathcal{J})$ follows.

We now prove the reverse inequality, namely $\tilde{I}(V,\mathcal{J}) \leq \tilde{I}_0(V,\mathcal{J})$. Once more fix $\delta > 0$ and $(V,\mathcal{J}) \in \mathcal{V} \times \mathbf{H}^{-\mathbf{s}}$ such that $\tilde{I}_0(V,\mathcal{J}) < \infty$, and let $(h,\Theta) \in L^2(\mathbf{V},[0,T]\times\mathbb{R}^d) \times (\mathcal{S}(\mathcal{Z}) \cap \mathcal{P}_2(\mathcal{Z}))$ be δ-optimal for $\tilde{I}_0(V,\mathcal{J})$, namely

$$\frac{1}{2}\int_0^T |\langle V(t), h(t,\cdot) - b(\cdot,V(t))\rangle|^2 \, dt \leq \tilde{I}_0(V,\mathcal{J}) + \delta, \tag{5.2.21}$$

$V = \nu_\Theta$, and $(V,\mathcal{J})$ solves (5.1.18). Under Θ, (5.2.19) is satisfied for the coordinate variable X, and so we have, as in (5.2.20), that for all $\varphi \in \mathcal{C}_c^\infty((0,T)\times\mathbb{R}^d,\mathbb{R})$,

$$0 = \int_0^T \left\langle V(t), \frac{\partial}{\partial t}\varphi(t,\cdot) + \nabla\varphi(t,\cdot)\cdot(b(\cdot,V(t)) + u(t,\cdot))\right\rangle \, dt, \tag{5.2.22}$$

for the random variable u on $(\mathcal{Z},\mathcal{B}(\mathcal{Z}))$ defined as in (5.2.18). Similarly, for all $\varphi \in \mathcal{C}_c^\infty((0,T)\times\mathbb{R}^d,\mathbb{R}^d)$,

$$\langle \mathcal{J}, \varphi\rangle = \int_0^T \langle V(t), \varphi(t,\cdot)\cdot(b(\cdot,V(t)) + u(t,\cdot))\rangle \, dt. \tag{5.2.23}$$

However, since $(V,\mathcal{J})$ solves (5.1.18) with the δ-optimal h chosen as above we must also have

$$\langle \mathcal{J}, \varphi\rangle = \int_0^T \langle V(t), \varphi(t,\cdot)\cdot h(\cdot,t)\rangle \, dt.$$

This says that for $\mathbf{V}$-a.e. $(t,x) \in [0,T]\times\mathbb{R}^d$,

$$h(t,x) = b(x,V(t)) + u(t,x). \tag{5.2.24}$$

Now define an $\mathcal{R}_1$-valued random variable $\tilde{\rho}$ on $(\mathcal{Z},\mathcal{B}(\mathcal{Z}))$ as

$$\tilde{\rho}(dt,dy) = \delta_{v(t)}(dy)\, dt,$$

where v is defined in terms of the coordinate variable ρ as in (5.2.17). Defining $\tilde{\Theta} \in \mathcal{P}(\mathcal{Z})$ as $\tilde{\Theta} \doteq \Theta \circ (X,\tilde{\rho})^{-1}$, we have that $\nu_{\tilde{\Theta}} = \nu_\Theta = V$, and it can be seen from (5.2.19) that $\tilde{\Theta} \in \mathcal{S}(\mathcal{Z})$. Also, observing that for any $\varphi \in \mathcal{C}_c^\infty((0,T)\times\mathbb{R}^d,\mathbb{R}^d)$,

$$\begin{aligned}\langle V(t), \varphi(t,\cdot)\cdot u(t,\cdot)\rangle &= \left\langle V(t), \varphi(t,\cdot)\cdot\int_{\mathbb{R}^d} y\,\hat{\eta}_t(\cdot,dy)\right\rangle = \int_{\mathbb{R}^{2d}} \varphi(t,x)\cdot y\,\eta_t(dx,dy)\\ &= E_\Theta\left[\varphi(t,X(t))\cdot v(t)\right] = E_{\tilde{\Theta}}\left[\varphi(t,X(t))\cdot\int_{\mathbb{R}^d} y\,\rho_t(dy)\right],\end{aligned}$$

we see from (5.2.23) and (5.1.10) that $\langle \mathcal{J}, \varphi \rangle = G_\varphi(\tilde{\Theta})$ for every $\varphi \in \mathcal{C}_c^\infty((0,T) \times \mathbb{R}^d, \mathbb{R}^d)$. Thus, $\tilde{\Theta} \in \mathcal{P}^*(\mathcal{J})$. Finally,

$$\begin{aligned}\tilde{I}(V, \mathcal{J}) &\le E_{\tilde{\Theta}}\left[\frac{1}{2}\int_{[0,T]\times\mathbb{R}^d} |y|^2\, \rho(dt, dy)\right] = E_\Theta\left[\frac{1}{2}\int_0^T |v(t)|^2\, dt\right] = \frac{1}{2}\int_0^T |\langle V(t), u(t,\cdot)\rangle|^2\ dt \\ &= \frac{1}{2}\int_0^T |\langle V(t), h(t,x) - b(x, V(t))\rangle|^2\ dt \le \tilde{I}_0(V, \mathcal{J}) + \delta,\end{aligned}$$

where we used (5.2.24) and (5.2.21). Since $\delta > 0$ is arbitrary, the inequality $\tilde{I}(V, \mathcal{J}) \le \tilde{I}_0(V, \mathcal{J})$ follows and completes the proof of the lemma. □

5.3 Proofs of Key Lemmas

In this section we provide proofs of the results used in showing the Laplace upper and lower bounds. First we establish two estimates that will be used in subsequent sections.

Lemma 5.3.1. *Suppose Conditions 5.1.1 and 5.1.2 are satisfied. Let* $u^N = (u_1^N, \ldots, u_N^N) \in \mathcal{A}_N$ *and let* $\bar{X}^N$ *be as defined in* (5.2.3). *Then, for each* $N \in \mathbb{N}$,

$$\frac{1}{N}\sum_{j=1}^N E\left[\left\|\bar{X}_j^N\right\|_\infty^2\right] \le c\left(1 + \frac{1}{N}\sum_{j=1}^N \left|x_j^N\right|^2 + E\left[\frac{1}{N}\sum_{j=1}^N\int_0^T \left|u_j^N(t)\right|^2\ dt\right]\right), \tag{5.3.1}$$

and for any $\varepsilon > 0$ *and any* $\{\mathcal{F}(t)\}$*-stopping time* τ *taking values in* $[0, T-\varepsilon]$,

$$\frac{1}{N}\sum_{j=1}^N E\left[\left|\bar{X}_j^N(\tau+\varepsilon) - \bar{X}_j^N(\tau)\right|^2\right] \le c\varepsilon\left(1 + \frac{1}{N}\sum_{j=1}^N \left|x_j^N\right|^2 + E\left[\frac{1}{N}\sum_{j=1}^N\int_0^T \left|u_j^N(t)\right|^2\ dt\right]\right).$$

where $c < \infty$ *does not depend on* N, u^N, *or* ε.

Proof. Condition 5.1.1 (see (5.1.5)) implies

$$\left|b\left(\bar{X}_j^N(t), \bar{V}^N(t)\right)\right|^2 \le 3L^2\left(1 + \left|\bar{X}_j^N(t)\right|^2 + \frac{1}{N}\sum_{j=1}^N \left|\bar{X}_j^N(t)\right|^2\right),$$

and so from (5.2.3) and since $|\sigma| \le L$ and $\varepsilon_N \le 1$, we have

$$\begin{aligned}\left|\bar{X}_j^N(t)\right|^2 \le 4\left|x_j^N\right|^2 &+ 4\left|\int_0^t b\left(\bar{X}_j^N(s), \bar{V}^N(s)\right)\ ds\right|^2 + 4\left|\varepsilon_N\int_0^t \sigma\left(\bar{X}_j^N(s), \bar{V}^N(s)\right)\ dW_j(s)\right|^2 \\ &+ 4\left|\int_0^t \sigma\left(\bar{X}_j^N(s), \bar{V}^N(s)\right) u_j^N(s)\, ds\right|^2\end{aligned}$$

$$\le 4\left|x_j^N\right|^2 + 12L^2T\left(1+\int_0^t \sup_{0\le r\le s}\left|\bar{X}_j^N(r)\right|^2 ds + \frac{1}{N}\sum_{j=1}^N \int_0^t \sup_{0\le r\le s}\left|\bar{X}_j^N(r)\right|^2 ds\right)$$
$$+ 4\sup_{0\le r\le t}\left|\int_0^r \sigma\left(\bar{X}_j^N(s), \bar{V}^N(s)\right) dW_j(s)\right|^2 + 4L^2T\int_0^T \left|u_j^N(s)\right|^2 ds.$$

Hence by The Burkholder-Davis-Gundy inequality, and using boundedness of σ once more,

$$\frac{1}{N}\sum_{j=1}^N E\left[\sup_{0\le s\le t}\left|\bar{X}_j^N(s)\right|^2\right] \le \frac{4}{N}\sum_{j=1}^N \left|x_j^N\right|^2 + 24L^2T\left(1+\int_0^t \frac{1}{N}\sum_{j=1}^N E\left[\sup_{0\le r\le s}\left|\bar{X}_j^N(r)\right|^2\right] ds\right)$$
$$+ 16L^2T + 4L^2TE\left[\frac{1}{N}\sum_{j=1}^N \int_0^T \left|u_j^N(s)\right|^2 ds\right].$$

The first statement in the lemma then follows by Gronwall's inequality (see [41, Theorem A.5.1]) with $c = 24(L^2T + 1)e^{24L^2T^2}$.

Next, for any $t \in [0, T-\varepsilon]$, the linear growth of b, boundedness of σ, and the Cauchy-Schwarz inequality give

$$\left|\bar{X}_j^N(t+\varepsilon) - \bar{X}_j^N(t)\right|^2 \le 4\left|\int_t^{t+\varepsilon} b\left(\bar{X}_j^N(s), \bar{V}^N(s)\right) ds\right|^2 + 4\left|\varepsilon_N \int_t^{t+\varepsilon} \sigma\left(\bar{X}_j^N(s), \bar{V}^N(s)\right) dW_j(s)\right|^2$$
$$+ 4\left|\int_t^{t+\varepsilon} \sigma\left(\bar{X}_j^N(s), \bar{V}^N(s)\right) u_j^N(s)\, ds\right|^2$$
$$\le 12TL^2\varepsilon\left(1+\sup_{0\le s\le T}\left|\bar{X}_j^N(s)\right|^2 + \frac{1}{N}\sum_{j=1}^N \sup_{0\le s\le T}\left|\bar{X}_j^N(s)\right|^2\right)$$
$$+ 4\left|\int_t^{t+\varepsilon} \sigma\left(\bar{X}_j^N(s), \bar{V}^N(s)\right) dW_j(s)\right|^2 + 4L^2\varepsilon\int_0^T \left|u_j^N(s)\right|^2 ds.$$

Since τ is a bounded stopping time, the optional sampling theorem gives

$$E\left|\int_\tau^{\tau+\varepsilon} \sigma\left(\bar{X}_j^N(s), \bar{V}^N(s)\right) dW_j(s)\right|^2 \le L^2\varepsilon,$$

and so

$$\frac{1}{N}\sum_{j=1}^N E\left[\left|\bar{X}_j^N(\tau+\varepsilon) - \bar{X}_j^N(\tau)\right|^2\right]$$
$$\le 24(T+1)L^2\varepsilon\left(1+E\left[\frac{1}{N}\sum_{j=1}^N \left\|\bar{X}_j^N\right\|_\infty^2\right] + E\left[\frac{1}{N}\sum_{j=1}^N \int_0^T \left|u_j^N(s)\right|^2 ds\right]\right).$$

The second estimate in the lemma now follows (with a possibly larger choice of c). □

5.3.1 Proof of Lemma 5.2.3

The following general lemma will be useful in proving the tightness of $\{\bar{\mathcal{J}}^N\}$. The proof is standard (see e.g. [41, Exercise 3.11.18]) and is therefore omitted.

Lemma 5.3.2. *Let $\{Z_k, k \in \mathbb{N}\}$ be a sequence of random variables taking values in a separable Banach space with norm $\|\cdot\|$. Suppose that for each $\varepsilon > 0$ we can write $Z_k = Z_k^\varepsilon + R_k^\varepsilon$ for each $k \in \mathbb{N}$, where $\{Z_k^\varepsilon, k \in \mathbb{N}\}$ is tight and $\sup_{k\geq 1} E\left[\|R_k^\varepsilon\|\right] \leq \varepsilon$. Then $\{Z_k\}$ is tight.*

To prove tightness for the controlled stochastic currents, we will make use of a collection of test functions $\{g_M, M < \infty\}$ defined as follows.

Definition 5.3.1. *Let $\{g_M, M < \infty\}$ be a collection of functions in $\mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R})$ that satisfy $0 \leq g_M(x) \leq 1$ for all $M < \infty$ and $x \in \mathbb{R}^d$, and have the following properties*

(i) For each M, $g_M(x) = 1$ on $|x| \leq M$,

(ii) For each M, $g_M(x) = 0$ on $|x| \geq M + 1$, and

(iii) For every $k \in \mathbb{N}$, there is a constant $B(k) < \infty$ such that $|D^\alpha g_M(x)| \leq B(k)$ for all $x \in \mathbb{R}^d$, all $M < \infty$, and all $|\alpha| \leq k$.

Note that if $\{g_M, M < \infty\}$ is a collection as in Definition 5.3.1 then for every $k \in \mathbb{N}$, there is a constant $L(k) < \infty$ such that

$$|D^\alpha g_M(x) - D^\alpha g_M(y)| \leq L(k)|x - y| \tag{5.3.2}$$

for all $x, y \in \mathbb{R}^d$, all $M < \infty$, and all $|\alpha| \leq k$. We will need the following property of the collection $\{g_M, M < \infty\}$. Proof of the lemma is given in the Appendix.

Lemma 5.3.3. *For any $s > 0$, there is a constant $K = K(s) < \infty$ such that for any $f \in H^s(\mathbb{R}^d, \mathbb{R}^d)$,*

$$\sup_{M<\infty} \|g_M f\|_s \leq K\|f\|_s.$$

The following is a simple extension of the well-known compact embedding result for Sobolev spaces on $\mathbb{R}^d$ known as Rellich's Theorem (see [45, Theorem 9.22]). Although the proof is standard, we provide details in the Appendix. For $\mathbf{s} \in \mathcal{O}_d$, $F \in \mathbf{H}^{-\mathbf{s}}$, and open $U_0 \subset U$, we say $F = 0$ on U_0 if for all $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$ with support in U_0, $\langle F, \varphi\rangle = 0$. The support of F is the complement of the union of all open sets in U on which $F = 0$.

Lemma 5.3.4. *Let $\mathbf{s} = (s_1, s_2)$ and $\mathbf{s}' = (s_1', s_2')$ in $\mathcal{O}_d$ be such that $s_1' < s_1$ and $s_2' < s_2$. Suppose $A \subset \mathbf{H}^{-\mathbf{s}'}$ is such that for some compact $K \subset U \times \mathbb{R}^d$, every $F \in A$ has support contained in K. Suppose also that $\sup_{F\in A} \|F\|_{-\mathbf{s}'} < \infty$. Then A is relatively compact in $\mathbf{H}^{-\mathbf{s}}$.*

Finally, the lemma below establishes the required tightness for the controlled currents.

Lemma 5.3.5. *Suppose Conditions 5.1.1 and 5.1.2 are satisfied. Let $\{g_M, M < \infty\}$ be the collection of functions in $\mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R})$ as in Definition 5.3.1. For each $N \in \mathbb{N}$, $M < \infty$, and $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$, define*

$$\bar{J}^{N,M}(\varphi) \doteq \bar{J}^N(g_M \varphi), \qquad \bar{J}_c^{N,M}(\varphi) \doteq \bar{J}^N(\varphi) - \bar{J}^{N,M}(\varphi).$$

Then, the collections $\{\varphi \mapsto \bar{J}^{N,M}(\varphi)\}$ and $\{\varphi \mapsto \bar{J}_c^{N,M}(\varphi)\}$ have pathwise realizations $\bar{\mathcal{J}}^{N,M}, \bar{\mathcal{J}}_c^{N,M}$ in $\mathbf{H}^{-\mathbf{s}}$ for all $\mathbf{s} \in \mathcal{O}_d$. Furthermore, if

$$\sup_{N \geq 1} E\left[\frac{1}{N}\sum_{j=1}^N \int_0^T \left|u_j^N(t)\right|^2 \, dt\right] < \infty,$$

then for all $\mathbf{s} \in \mathcal{O}_d$,

$$\sup_{M<\infty} \sup_{N \geq 1} E\left[\|\bar{\mathcal{J}}^{N,M}\|_{-\mathbf{s}}\right] < \infty$$

and

$$\lim_{M \to \infty} \sup_{N \geq 1} E\left[\|\bar{\mathcal{J}}_c^{N,M}\|_{-\mathbf{s}}\right] = 0.$$

In particular, $\{\bar{\mathcal{J}}^N, N \in \mathbb{N}\}$ is tight in $\mathbf{H}^{-\mathbf{s}}$ for all $\mathbf{s} \in \mathcal{O}_d$.

Proof. Fix $\mathbf{s} = (s_1, s_2) \in \mathcal{O}_d$, and for each N and $1 \leq j \leq N$, let $C_{j,\mathbf{s}}^N$ be the square-integrable random variable from Lemma 5.2.1, so that $|\bar{J}^N(\varphi)| \leq C_\mathbf{s}^N \|\varphi\|_\mathbf{s}$ a.s. for all $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$, where $C_\mathbf{s}^N = \frac{1}{N}\sum_{j=1}^N C_{j,\mathbf{s}}^N$. As a consequence of Lemma 5.3.3, for some constant $K = K(s_2) < \infty$, we have, for all $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$ and $M < \infty$,

$$\|g_M \varphi\|_\mathbf{s}^2 = \int_U \|g_M \varphi(u,\cdot)\|_{s_2}^2 \, du + \int_U \int_U \frac{\|g_M(\varphi(u,\cdot) - \varphi(v,\cdot))\|_{s_2}^2}{|u-v|^{1+2s_1}} \, du \, dv \leq K^2 \|\varphi\|_\mathbf{s}^2. \tag{5.3.3}$$

Hence,

$$\left|\bar{J}^{N,M}(\varphi)\right| \leq C_\mathbf{s}^N \|g_M \varphi\|_\mathbf{s} \leq K C_\mathbf{s}^N \|\varphi\|_\mathbf{s} \qquad \text{a.s.,}$$

and

$$\left|\bar{J}_c^{N,M}(\varphi)\right| = \left|\bar{J}^N((1-g_M)\varphi)\right| \leq C_\mathbf{s}^N \|(1-g_M)\varphi\|_\mathbf{s} \leq (1+K) C_\mathbf{s}^N \|\varphi\|_\mathbf{s} \qquad \text{a.s.}$$

From [42, Lemma 5] it then follows that, for every $M < \infty$, there are $\mathbf{H}^{-\mathbf{s}}$-valued random variables $\bar{\mathcal{J}}^{N,M}$ and $\bar{\mathcal{J}}_c^{N,M}$ such that, for every $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$ and $M < \infty$,

$$\left\langle \bar{\mathcal{J}}^{N,M}(\omega), \varphi \right\rangle = \left[J^{N,M}(\varphi)\right](\omega) \qquad \text{and} \qquad \left\langle \bar{\mathcal{J}}_c^{N,M}(\omega), \varphi \right\rangle = \left[J_c^{N,M}(\varphi)\right](\omega), \qquad \text{a.e. } \omega \in \Omega.$$

Then, from Lemma 5.2.1,

$$\sup_{M<\infty} \sup_{N\geq 1} E\left[\left\|\bar{\mathcal{J}}^{N,M}\right\|_{-\mathbf{s}}^2\right] \leq K^2 \sup_{N\geq 1} E\left[\left(C_{\mathbf{s}}^N\right)^2\right] < \infty. \tag{5.3.4}$$

Let $\bar{J}_j^N$ be as in (5.2.4) and define the stopping times $\tau_j^{N,M} = \inf\{t > 0 : |\bar{X}_j^N(t)| \geq M\}$. Then,

$$\bar{J}_c^{N,M}(\varphi) = \frac{1}{N}\sum_{j=1}^N \bar{J}_j^N((1-g_M)\varphi) = \frac{1}{N}\sum_{j=1}^N 1_{\left\{\tau_j^{N,M}<T\right\}} \bar{J}_j^N((1-g_M)\varphi),$$

and by Lemma 5.2.1,

$$\left|\bar{J}_j^N((1-g_M)\varphi)\right| \leq C_{j,\mathbf{s}}^N \|(1-g_M)\varphi\|_{\mathbf{s}} \leq (1+K) C_{j,\mathbf{s}}^N \|\varphi\|_{\mathbf{s}}.$$

Thus,

$$\left|\bar{J}_c^{N,M}(\varphi)\right| \leq \left(\frac{1+K}{N}\sum_{j=1}^N 1_{\left\{\tau_j^{N,M}<T\right\}} C_{j,\mathbf{s}}^N\right) \|\varphi\|_{\mathbf{s}} \doteq \tilde{C}_{\mathbf{s}}^N \|\varphi\|_{\mathbf{s}}. \tag{5.3.5}$$

Also, by the Cauchy-Schwarz inequality,

$$E\left[\left(\tilde{C}_{\mathbf{s}}^N\right)^2\right] \leq \frac{(1+K)^2}{N}\left(\sum_{j=1}^N P\left(\tau_j^{N,M} < T\right)\right)\left(\frac{1}{N}\sum_{j=1}^N E\left[\left(C_{j,\mathbf{s}}^N\right)^2\right]\right).$$

By Lemma 5.3.1, Condition 5.1.2, and the assumption that $\sup_{N\in\mathbb{N}} E\left[\frac{1}{N}\sum_{j=1}^N \int_0^T \left|u_j^N(t)\right|^2 dt\right] < \infty$, there is a constant $\tilde{K} < \infty$ such that

$$\sup_{N\geq 1} \frac{1}{N}\sum_{j=1}^N P\left(\tau_j^{N,M} < T\right) \leq \sup_{N\geq 1} \frac{1}{N}\sum_{j=1}^N P\left(\left\|\bar{X}_j^N\right\|_\infty \geq M\right) \leq \frac{\tilde{K}}{M^2}.$$

Thus,

$$E\left[\left\|\bar{\mathcal{J}}_c^{N,M}\right\|_{-\mathbf{s}}^2\right] \leq E\left[\left(\tilde{C}_{\mathbf{s}}^N\right)^2\right] \leq \frac{\tilde{K}(1+K)^2}{M^2} \sup_{N\geq 1} \frac{1}{N}\sum_{j=1}^N E\left[\left(C_{j,\mathbf{s}}^N\right)^2\right],$$

and therefore, from Lemma 5.2.1,

$$\lim_{M\to\infty} \sup_{N\geq 1} E\left[\left\|\bar{\mathcal{J}}_c^{N,M}\right\|_{-\mathbf{s}}^2\right] = 0. \tag{5.3.6}$$

Note that (5.3.4) and (5.3.6) are satisfied for every $\mathbf{s} \in \mathcal{O}_d$. Now for an arbitrary $\mathbf{s} \in \mathcal{O}_d$, choose $\mathbf{s}' = (s_1', s_2') \in \mathcal{O}_d$ such that $s_1' < s_1$ and $s_2' < s_2$. Then applying (5.3.4) for $\mathbf{s}'$ and observing that $\{\bar{\mathcal{J}}^{N,M}, N \in \mathbb{N}\}$ are compactly supported on $[0,T]\times\{|x| \leq M+1\} \subset U\times\mathbb{R}^d$, we see from Lemma 5.3.4 and Markov's inequality that for each fixed M,

$\{\bar{\mathcal{J}}^{N,M}, N \in \mathbb{N}\}$ is a tight collection of $\mathbf{H}^{-\mathbf{s}}$-valued random variables. Finally, observing that $\bar{\mathcal{J}}^N = \bar{\mathcal{J}}^{N,M} + \bar{\mathcal{J}}_c^{N,M}$ for each M and applying (5.3.6) and Lemma 5.3.2, we obtain that $\{\bar{\mathcal{J}}^N, N \in \mathbb{N}\}$ is tight in $\mathbf{H}^{-\mathbf{s}}$. $\square$

The following general lemma will be useful in proving tightness of $\{\bar{\mu}^N\}$.

Lemma 5.3.6. *Let (S, d_S) be a Polish space. If $\{\gamma_k, k \in \mathbb{N}\}$ is a tight sequence of $\mathcal{P}(S)$-valued random variables and for some $x_0 \in S$*

$$\sup_{k\in\mathbb{N}} E\left[\int_S d_S(x,x_0)^2\, \gamma_k(dx)\right] < \infty, \tag{5.3.7}$$

then $\{\gamma_k\}$ is tight as a sequence of $\mathcal{P}_1(S)$-valued random variables.

Proof. Suppose that γ_k converges in distribution, along a subsequence, in $\mathcal{P}(S)$ to some γ, and denote the convergent subsequence once more as $\{\gamma_k\}$. From (5.3.7) it follows that each γ_k is in $\mathcal{P}_1(S)$ a.s. Furthermore, by lower semicontinuity of the map $\mu \mapsto \int_S d_S(x,x_0)^2\, \mu(dx)$ on $\mathcal{P}(S)$ and Fatou's lemma, we see that

$$E\left[\int_S d_S(x,x_0)^2\, \gamma(dx)\right] \le E\left[\liminf_{k\to\infty} \int_S d_S(x,x_0)^2\, \gamma_k(dx)\right] \le \sup_{k\ge 1} E\left[\int_S d_S(x,x_0)^2\, \gamma_k(dx)\right] < \infty,$$

and so in particular $\gamma \in \mathcal{P}_1(S)$ a.s. By appealing to Skorohod's representation theorem we can assume that $\gamma_k \to \gamma$ a.s. in $\mathcal{P}(S)$. Recalling from Section **??** the metric d_{bl} on the space $\mathcal{P}(S)$, we have that $d_{\text{bl}}(\gamma_k, \gamma) \to 0$ a.s.

It suffices now to show that γ_k converges in probability in $\mathcal{P}_1(S)$ to γ. Take $f \in \mathcal{L}(S)$ such that $f(x_0) = 0$. Fix $1 < M < \infty$ and define

$$f_M(x) \doteq \left(\frac{f(x)}{M} \vee (-1)\right) \wedge 1,$$

which is a function bounded by 1 in absolute value whose Lipschitz constant is also bounded by 1. Then,

$$\begin{aligned}
&\left|\int_S f(x)\, \gamma_k(dx) - \int_S f(x)\, \gamma(dx)\right| \\
&\le M \left|\int_S f_M(x)\, \gamma_k(dx) - \int_S f_M(x)\, \gamma(dx)\right| + \int_S |M f_M(x) - f(x)|\, \gamma_k(dx) + \int_S |M f_M(x) - f(x)|\, \gamma(dx) \\
&\le M d_{\text{bl}}(\gamma_k, \gamma) + \int_S 2|f(x)| 1_{\{|f(x)|>M\}}\, \gamma_k(dx) + \int_S 2|f(x)| 1_{\{|f(x)|>M\}}\, \gamma(dx).
\end{aligned}$$

Since the Lipschitz constant of f is bounded by 1 and $f(x_0) = 0$, we have that $|f(x)| \le d_S(x, x_0)$, and so

$$\int_S |f(x)| 1_{\{|f(x)|>M\}}\, \gamma_k(dx) \le \frac{1}{M} \int_S d_S(x,x_0)^2\, \gamma_k(dx),$$

and the equivalent inequality holds for γ. Now, since $\langle \mu, f\rangle - \langle \nu, f\rangle = \langle \mu, f - f(x_0)\rangle - \langle \nu, f - f(x_0)\rangle$ for any $\mu, \nu \in \mathcal{P}_1(S)$ and $f \in \mathcal{L}(S)$, the supremum in the definition of d_1 can be restricted to f such that $f(x_0) = 0$. Thus,

$$\begin{aligned} E\left[d_1(\gamma_k, \gamma)\right] &= E\left[\sup_{f\in\mathcal{L}(S), f(x_0)=0}\left|\int_S f(x)\,\gamma_k(dx) - \int_S f(x)\,\gamma(dx)\right|\right] \\ &\leq M E\left[d_{\text{bl}}(\gamma_k, \gamma)\right] + \frac{2}{M}\sup_{l\in\mathbb{N}} E\left[\int_S d_S(x, x_0)^2\,\gamma_l(dx)\right] + \frac{2}{M} E\left[\int_S d_S(x, x_0)^2\,\gamma(dx)\right]. \end{aligned}$$

Sending first $k \to \infty$ and then $M \to \infty$, we have that $\lim_{k\to\infty} E\left[d_1(\gamma_k, \gamma)\right] = 0$ which completes the proof. ☐

We can now complete the proof of Lemma 5.2.3.

5.3.1.1 Proof of Lemma 5.2.3(i) We begin by arguing that $\{\bar{\mu}^N\}$ is a tight sequence of $\mathcal{P}(\mathcal{X})$-valued random variables. For this it suffices to show (see [21, Theorem 2.11]) that $\{\gamma^N, N \in \mathbb{N}\}$ is a relatively compact set in $\mathcal{P}(\mathcal{X})$, where

$$\gamma^N \doteq E\left[\bar{\mu}^N\right] = \frac{1}{N}\sum_{j=1}^N P\left(\bar{X}_j^N \in \cdot\right).$$

Note that

$$\int_{\mathcal{X}} \|\psi\|_\infty^2\,\gamma^N(d\psi) = \frac{1}{N}\sum_{j=1}^N E\left[\left\|\bar{X}_j^N\right\|_\infty^2\right],$$

and so by Lemma 5.3.1 and the assumption on the controls in Lemma 5.2.3, we see that

$$\sup_{N\geq 1}\int_{\mathcal{X}} \|\psi\|_\infty^2\,\gamma^N(d\psi) = \sup_{N\geq 1}\frac{1}{N}\sum_{j=1}^N E\left[\left\|\bar{X}_j^N\right\|_\infty^2\right] < \infty. \tag{5.3.8}$$

Next, for $\varepsilon > 0$ let $\mathcal{T}_\varepsilon$ denote the collection of all $\{\sigma(X(s) : s \leq t)\}$-stopping times on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ taking values in $[0, T - \varepsilon]$ where $\{X(t)\}$ is the canonical coordinate process on $\mathcal{X}$. Then for each $N \in \mathbb{N}$, there are $\{\sigma(\bar{X}_j^N(s) : s \leq t)\}$-stopping times $\{\tau_j^N, 1 \leq j \leq N\}$ on $(\Omega, \mathcal{F})$ with values in $[0, T - \varepsilon]$, such that

$$\int_{\mathcal{X}} |\psi(\tau + \varepsilon) - \psi(\tau)|^2\,\gamma^N(d\psi) = \frac{1}{N}\sum_{j=1}^N E\left[\left|\bar{X}_j^N\left(\tau_j^N + \varepsilon\right) - \bar{X}_j^N\left(\tau_j^N\right)\right|^2\right].$$

Applying Lemma 5.3.1, we then have

$$\int_{\mathcal{X}} |\psi(\tau + \varepsilon) - \psi(\tau)|^2\,\gamma^N(d\psi) \leq c\varepsilon\left(1 + \sup_{N\geq 1}\frac{1}{N}\sum_{j=1}^N \left|x_j^N\right|^2 + \sup_{N\geq 1} E\left[\frac{1}{N}\sum_{J=1}^N \int_0^T \left|u_j^N(t)\right|^2\,dt\right]\right),$$

and hence

$$\lim_{\varepsilon\to 0}\sup_{N\geq 1}\sup_{\tau\in\mathcal{T}_\varepsilon}\int_{\mathcal{X}} |\psi(\tau + \varepsilon) - \psi(\tau)|^2\,\gamma^N(d\psi) = 0. \tag{5.3.9}$$

The relative compactness of $\{\gamma^N, N \in \mathbb{N}\}$ in $\mathcal{P}(\mathcal{X})$ is immediate from (5.3.8) and (5.3.9) (see [21, Theorem D.4]), which as noted previously shows $\{\bar{\mu}^N\}$ is a tight sequence of $\mathcal{P}(\mathcal{X})$-valued random variables. The tightness of $\{\bar{\mu}^N\}$ as a sequence of $\mathcal{P}_1(\mathcal{X})$-valued random variables now follows from Lemma 5.3.6 and the uniform moment estimate in (5.3.8). Note also that since $\bar{\mu}^N = Q^N_{(1)}$, we have the tightness of the first marginals of $\{Q^N\}$ (as a sequence of $\mathcal{P}(\mathcal{X})$-valued random variables).

That the second marginals $\{Q^N_{(2)}\}$ is a tight sequence of $\mathcal{P}(\mathcal{R}_1)$-valued random variables follows by an argument similar to [22, Lemma 5.1] however we provide the details. Note that the function

$$h(r) = \int_{[0,T]\times\mathbb{R}^m} |y|^2 \, r(dt, dy)$$

has compact level sets on $\mathcal{R}_1$ (recall that $\mathcal{R}_1$ is equipped with the Wasserstein-1 metric). It then follows that

$$H(\theta) = \int_{\mathcal{R}_1} h(r) \, \theta(dr)$$

has relatively compact level sets on $\mathcal{P}(\mathcal{R}_1)$ (see [21, Lemma 2.10]). It now suffices to show $\sup_{N\geq 1} E[H(Q^N_{(2)})] < \infty$ (see [21, Lemmas 2.9]). However this is immediate as

$$\sup_{N\geq 1} E\left[H\left(Q^N_{(2)}\right)\right] = \sup_{N\geq 1} E\left[\frac{1}{N}\sum_{j=1}^N \int_{[0,T]\times\mathbb{R}^m} |y|^2 \rho^N_j(dt, dy)\right] = \sup_{N\geq 1} E\left[\frac{1}{N}\sum_{j=1}^N \int_0^T \left|u^N_j(t)\right|^2 \, dt\right] < \infty. \tag{5.3.10}$$

Thus we have shown that the second marginals of $\{Q^N\}$ are also tight, which in turn shows that $\{\bar{\mu}^N, Q^N\}$ is a tight sequence of $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z})$-valued random variables. Together with Lemma 5.3.5, this finishes the proof of Lemma 5.2.3(i). □

5.3.1.2 Proof of Lemma 5.2.3(ii) Suppose now that $(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N) \Rightarrow (\bar{\mu}, Q, \bar{\mathcal{J}})$ in $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z}) \times \mathbf{H}^{-\mathbf{s}}$, where $(\bar{\mu}, Q, \bar{\mathcal{J}})$ is defined on some probability space. By appealing to Skorokhod's representation theorem, we can assume that $\{(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N)\}$ and $(\bar{\mu}, Q, \bar{\mathcal{J}})$ are defined on a common probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ and that $(\bar{\mu}^N, Q^N, \bar{\mathcal{J}}^N) \to (\bar{\mu}, Q, \bar{\mathcal{J}})$ a.s. Let $\tilde{E}$ denote expectation on this space. The property $Q_{(1)} = \bar{\mu}$ is immediate from the identity $Q^N_{(1)} = \bar{\mu}^N$ for every $N \in \mathbb{N}$. We will complete the remainder of the proof in three steps: step 1 will establish that $Q \in \mathcal{P}_2(\mathcal{Z})$, step 2 that $Q \in \mathcal{S}(\mathcal{Z})$, and step 3 that $Q \in \mathcal{P}^*(\bar{\mathcal{J}})$, from which the result will follow.

Step 1. By Fatou's lemma,

$$\tilde{E}\left[E_Q\left[\int_{\mathbb{R}^m\times[0,T]}|y|^2\rho(dy\,dt)\right]\right] \le \liminf_{N\to\infty}\tilde{E}\left[E_{Q^N}\left[\int_{\mathbb{R}^m\times[0,T]}|y|^2\rho(dy\,dt)\right]\right]$$
$$= \liminf_{N\to\infty} E\left[\frac{1}{N}\sum_{j=1}^N\int_0^T\left|u_j^N(t)\right|^2\,dt\right] < \infty, \tag{5.3.11}$$

and hence $Q \in \mathcal{P}_2(\mathcal{Z})$ a.s.

Step 2. We now show that a.s. $Q \in \mathcal{S}(\mathcal{Z})$, namely it is a weak solution to (5.1.9). Define the generator $\mathcal{A}$ as follows. For each $f \in \mathcal{C}_c^2(\mathbb{R}^d, \mathbb{R})$, let

$$\mathcal{A}f(\nu, x, y) = (b(x,\nu) + \sigma(x,\nu)y)\cdot\nabla f(x), \qquad (\nu, x, y) \in \mathcal{P}_1\left(\mathbb{R}^d\right)\times\mathbb{R}^d\times\mathbb{R}^m.$$

Now fix an $f \in \mathcal{C}_c^2(\mathbb{R}^d, \mathbb{R})$ and and define, for each $V \in \mathcal{V}$, the $\mathbb{R}$-valued process $\{M^V(t), 0 \le t \le T\}$ on the measurable space $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$ by

$$M^V(t,(\xi,r)) = f(\xi(t)) - f(\xi(0)) - \int_{[0,t]\times\mathbb{R}^m}\mathcal{A}f\left(V(s),\xi(s),y\right)\,r(ds,dy), \qquad (\xi,r)\in\mathcal{Z}. \tag{5.3.12}$$

Let $\bar{V} \doteq \nu_Q$. Since f is arbitrary, to establish that $Q \in \mathcal{S}(\mathcal{Z})$ a.s., it suffices to show that for each fixed $0 \le t \le T$ and a.e. $\omega \in \tilde{\Omega}$,

$$M^{\bar{V}(\omega)}(t,(\xi,r)) = 0, \qquad Q(\omega)\text{-a.e. } (\xi,r)\in\mathcal{Z}. \tag{5.3.13}$$

We will supress ω from the notation for the remainder of the proof.

For each $1 \le B < \infty$, let $\psi_B \in \mathcal{C}_c(\mathbb{R}^m, \mathbb{R}^m)$ be such that $\psi_B(y) = y$ on $\{|y| \le B\}$ and $|\psi_B(y)| \le |y| + 1$ everywhere. Note that since $B \ge 1$, this definition implies that

$$|y - \psi_B(y)| \le \frac{|y|(2|y|+1)}{B}1_{\{|y|>B\}} \le \frac{3|y|^2}{B}. \tag{5.3.14}$$

Also let $\eta_B \in \mathcal{C}_c(\mathbb{R}^d, \mathbb{R}^d)$ be such that $\eta_B(x) = x$ on $\{|x| \le B\}$ and $|\eta_B(x)| \le |x| + 1$ everywhere. As with ψ_B, we have that

$$|x - \eta_B(x)| \le \frac{3|x|^2}{B}. \tag{5.3.15}$$

Now define the 'truncated generator' $\mathcal{A}_B$

$$\mathcal{A}_Bf(\nu,x,y) = (\eta_B(b(x,\nu)) + \sigma(x,\nu)\psi_B(y))\cdot\nabla f(x), \qquad (\nu,x,y)\in\mathcal{P}_1\left(\mathbb{R}^d\right)\times\mathbb{R}^d\times\mathbb{R}^m,$$

and for each $V \in \mathcal{V}$, let $\{M_B^V(t)\}$ be the corresponding process defined as in (5.3.12) with $\mathcal{A}_B$ in place of $\mathcal{A}$. Let

$$K \doteq \sup_{x\in\mathbb{R}^d} \left(|f(x)| + |\nabla f(x)| + |D^2 f(x)|\right) < \infty,$$

and note that for all $V \in \mathcal{V}$, $0 \le s \le t$, and $(x, y) \in \mathbb{R}^d \times \mathbb{R}^m$,

$$\begin{aligned} |\mathcal{A}f(V(s), x, y) - \mathcal{A}_B f(V(s), x, y)| &\le K \left(\frac{3\,|b(x, V(s))|^2}{B} + \frac{3L|y|^2}{B} \right) \\ &\le \frac{12K(L+1)^2}{B} \left(1 + |x|^2 + \int_{\mathbb{R}^d} |x'|^2 \; V(s, dx') + |y|^2\right). \end{aligned} \tag{5.3.16}$$

Now fix t, and define the maps Φ and Φ_B on $\mathcal{P}(\mathcal{Z}) \times \mathcal{V}$ by

$$\Phi(\Theta, V) = E_\Theta\left[\left|M^V(t)\right|\right], \qquad \Phi_B(\Theta, V) = E_\Theta\left[\left|M_B^V(t)\right|\right].$$

Note that $\bar{V}^N = \nu_{Q^N}$, were $\bar{V}^N$ is as in Section 5.2.1. We proceed by showing that

(a) Φ_B is bounded and continuous on $\mathcal{P}(\mathcal{Z}) \times \mathcal{V}$,

(b) $\sup_{N\ge1} \tilde{E}\left[\left|\Phi(Q^N, \bar{V}^N) - \Phi_B(Q^N, \bar{V}^N)\right|\right] \to 0$ and $\left|\Phi(Q, \bar{V}) - \Phi_B(Q, \bar{V})\right| \xrightarrow{\tilde{P}} 0$ as $B \to \infty$, and

(c) $\Phi(Q^N, \bar{V}^N) \xrightarrow{\tilde{P}} 0$ as $N \to \infty$.

The convergence $(Q^N, \bar{V}^N) \to (Q, \bar{V})$ then yields that $\Phi(Q, \bar{V}) = 0$ a.s., from which the statement in (5.3.13) is immediate.

We first show (a). Boundedness of Φ_B follows from the boundedness of η_B, ψ_B, σ, f, and ∇f. The continuity of Φ_B follows from the continuity of the map $(V, z) \mapsto M_B^V(t, z)$ on $\mathcal{V} \times \mathcal{Z}$.

For (b), note from (5.3.16) that

$$\begin{aligned} &\tilde{E}\left[\left|\Phi\left(Q^N, \bar{V}^N\right) - \Phi_B\left(Q^N, \bar{V}^N\right)\right|\right] \le \tilde{E}\left[E_{Q^N}\left[\left|M^{\bar{V}^N}(t) - M_B^{\bar{V}^N}(t)\right|\right]\right] \\ &\le \frac{12K(L+1)^2}{B}\tilde{E}\left[E_{Q^N}\left[\int_0^T \left(1 + |X(s)|^2 + \int_{\mathbb{R}^d} |x|^2\, \bar{V}^N(s, dx) + \int_{\mathbb{R}^m} |y|^2\, \rho_s(dy)\right) ds\right]\right] \\ &\le \frac{12K(L+1)^2}{B} \sup_{N\ge1} E\left[T + \frac{2T}{N}\sum_{j=1}^N \left\|\bar{X}_j^N\right\|_\infty^2 + \frac{1}{N}\sum_{j=1}^N \int_0^T \left|u_j^N(s)\right|^2 ds\right]. \end{aligned} \tag{5.3.17}$$

From Lemma 5.3.1 and the assumption on the controls in Lemma 5.2.3, we see that the last term in the above display converges to 0 as $B \to \infty$. Similarly, since $Q \in \mathcal{P}_2(\mathcal{Z})$ a.s., the estimate

$$\left|\Phi(Q, \bar{V}) - \Phi_B(Q, \bar{V})\right| \le E_Q\left[\left|M^{\bar{V}}(t) - M_B^{\bar{V}}(t)\right|\right]$$

$$\leq \frac{12K(L+1)^2}{B}\left(\int_0^T\left(1+2\int_{\mathbb{R}^d}|x|^2\,\bar{V}(s,dx)\right)ds+E_Q\left[\int_{[0,T]\times\mathbb{R}^m}|y|^2\,\rho(ds,dy)\right]\right)$$

implies that

$$\left|\Phi\left(Q,\bar{V}\right)-\Phi_B\left(Q,\bar{V}\right)\right|\to 0 \qquad \text{a.s.,} \qquad \text{as } B\to\infty. \tag{5.3.18}$$

This completes the proof of (b).

We now turn to (c). Note that

$$\begin{aligned}\Phi\left(Q^N,\bar{V}^N\right)&=E_{Q^N}\left[\left|M^{\bar{V}^N}(t)\right|\right]=\frac{1}{N}\sum_{j=1}^N\left|M^{\bar{V}^N}\left(t,(\bar{X}_j^N,\rho_j^N)\right)\right|\\&=\frac{1}{N}\sum_{j=1}^N\left|f\left(\bar{X}_j^N(t)\right)-f\left(x_j^N\right)-\int_0^t\mathcal{A}f\left(\bar{V}^N(s),\bar{X}_j^N(s),u_j^N(s)\right)\,ds\right|.\end{aligned}$$

By Itô's lemma, for each $1\leq j\leq N$,

$$\begin{aligned}f\left(\bar{X}_j^N(t)\right)-f\left(x_j^N\right)=&\int_0^t\mathcal{A}f\left(\bar{V}^N,\bar{X}_j^N(s),u_j^N(s)\right)\,ds\\&+\varepsilon_N\int_0^t\nabla f\left(\bar{X}_j^N(s)\right)\cdot\sigma\left(\bar{X}_j^N(s),\bar{V}^N(s)\right)\,dW_j(s)\\&+\frac{\varepsilon_N^2}{2}\int_0^t\operatorname{Tr}\left[D^2f\left(\bar{X}_j^N(s)\right)\left(\sigma\sigma^{\mathsf{T}}\right)\left(\bar{X}_j^N(s),\bar{V}^N(s)\right)\right]\,ds.\end{aligned}$$

Hence,

$$\begin{aligned}\Phi\left(Q^N,\bar{V}^N\right)=\frac{1}{N}\sum_{j=1}^N\Bigg|&\varepsilon_N\int_0^t\nabla f\left(\bar{X}_j^N(s)\right)\cdot\sigma\left(\bar{X}_j^N(s),\bar{V}^N(s)\right)\,dW_j(s)\\&+\frac{\varepsilon_N^2}{2}\int_0^t\operatorname{Tr}\left[D^2f\left(\bar{X}_j^N(s)\right)\left(\sigma\sigma^{\mathsf{T}}\right)\left(\bar{X}_j^N(s),\bar{V}^N(s)\right)\right]\,ds\Bigg|.\end{aligned}$$

From the boundedness of ∇f, $D^2 f$, and σ, it follows that

$$\tilde{E}\left[\Phi\left(Q^N,\bar{V}^N\right)\right]\leq\left(KLT^{1/2}\right)\varepsilon_N+\frac{KL^2T\varepsilon_N^2}{2}\to 0 \qquad \text{as } N\to\infty.$$

This completes (c), which as noted previously proves the statement in (5.3.13) and which in turn shows that Q is a.s. a weak solution to (5.1.9).

Step 3. To complete the proof of Lemma 5.2.3, it only remains to establish that

$$G_\varphi(Q)=\langle\bar{\mathcal{J}},\varphi\rangle \qquad \text{for all} \qquad \varphi\in\mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d), \qquad \tilde{P}\text{-a.s.} \tag{5.3.19}$$

By considering a countable, dense subset of $\mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$, it suffices to show that for each fixed $\varphi\in\mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$, we have $G_\varphi(Q)=\bar{\mathcal{J}}(\varphi)$ a.s.

Fix φ, and let

$$K_\varphi \doteq \sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left(|\varphi(t,x)|+\sum_{k,l=1}^d\left|\frac{\partial\varphi_k}{\partial x_l}(t,x)\right|\right)<\infty.$$

Then, a.s.,

$$\begin{aligned}
\left\langle\bar{\mathcal{J}}^N,\varphi\right\rangle &= \frac{1}{N}\sum_{j=1}^N\int_0^T\varphi\left(t,\bar{X}_j^N(t)\right)\circ d\bar{X}_j^N(t)\\
&= \frac{1}{N}\sum_{j=1}^N\int_0^T\varphi\left(t,\bar{X}_j^N(t)\right)\cdot d\bar{X}_j^N(t)+\frac{1}{2N}\sum_{j=1}^N\left\langle\varphi\left(\cdot,\bar{X}_j^N(\cdot)\right),\bar{X}_j^N(\cdot)\right\rangle_T\\
&= \frac{1}{N}\sum_{j=1}^N\int_0^T\varphi\left(t,\bar{X}_j^N(t)\right)\cdot d\bar{X}_j^N(t)\\
&\quad+\frac{\varepsilon_N^2}{2N}\sum_{j=1}^N\int_0^T\sum_{k,l=1}^d\frac{\partial\varphi_k}{\partial x_l}\left(t,\bar{X}_j^N(t)\right)(\sigma\sigma^\mathsf{T})_{lk}\left(\bar{X}_j^N(t),\bar{V}^N(t)\right)\,dt.
\end{aligned}$$

Define

$$G_\varphi^*\left(Q^N\right)\doteq\frac{1}{N}\sum_{j=1}^N\int_0^T\varphi\left(t,\bar{X}_j^N(t)\right)\cdot d\bar{X}_j^N(t).$$

Since $|\sigma|\le L$,

$$\left|\left\langle\bar{\mathcal{J}}^N,\varphi\right\rangle-G_\varphi^*\left(Q^N\right)\right|=\left|\frac{\varepsilon_N^2}{2N}\sum_{j=1}^N\int_0^T\sum_{k,l=1}^d\frac{\partial\varphi_k}{\partial x_l}\left(t,\bar{X}_j^N(t)\right)(\sigma\sigma^\mathsf{T})_{lk}\left(\bar{X}_j^N(t),\bar{V}^N(t)\right)\,dt\right|\le\frac{K_\varphi L^2T\varepsilon_N^2}{2},$$

and hence $|\langle\bar{\mathcal{J}}^N,\varphi\rangle-G_\varphi^*(Q^N)|\to0$ in L^1 as $N\to\infty$. Also, by the dominated convergence theorem,

$$\lim_{N\to\infty}E\left[\left|\left\langle\bar{\mathcal{J}},\varphi\right\rangle-\left\langle\bar{\mathcal{J}}^N,\varphi\right\rangle\right|\wedge1\right]=0.$$

Next, writing

$$\left|\left\langle\bar{\mathcal{J}},\varphi\right\rangle-G_\varphi(Q)\right|\wedge1\le\left|\left\langle\bar{\mathcal{J}},\varphi\right\rangle-\left\langle\bar{\mathcal{J}}^N,\varphi\right\rangle\right|\wedge1+\left|\left\langle\bar{\mathcal{J}}^N,\varphi\right\rangle-G_\varphi^*\left(Q^N\right)\right|+\left|G_\varphi^*\left(Q^N\right)-G_\varphi(Q)\right|,$$

we see that to prove (5.3.19) and thus to complete the proof it suffices to argue that the third term on the right side of the above display converges to 0 in probability.

To this end, define the maps $\tilde{G}_\varphi$ and $\tilde{G}^B_\varphi$ on $\{\Theta \in \mathcal{P}_2(\mathcal{Z}) : \nu_\Theta \in \mathcal{V}\} \times \mathcal{V}$ by

$$\tilde{G}_\varphi(\Theta, V) \doteq E_\Theta \left[\int_0^T \varphi(t, X(t)) \cdot b(X(t), V(t))\, dt + \int_{[0,T]\times\mathbb{R}^m} \varphi(t, X(t)) \cdot \sigma(X(t), V(t)) y\, \rho(dt, dy) \right],$$

$$\begin{aligned}\tilde{G}^B_\varphi(\Theta, V) \doteq E_\Theta &\left[\int_0^T \varphi(t, X(t)) \cdot \eta_B \left(b(X(t), V(t)) \right) dt \right] \\ &+ E_\Theta \left[\int_{[0,T]\times\mathbb{R}^m} \varphi(t, X(t)) \cdot \sigma(X(t), V(t)) \psi_B(y)\, \rho(dt, dy) \right],\end{aligned}$$

for each $1 \le B < \infty$. Note by (5.1.10) that $\tilde{G}_\varphi(\Theta, \nu_\Theta) = G_\varphi(\Theta)$ whenever $\Theta \in \mathcal{S}(\mathcal{Z})$, and hence since $\bar{V} = \nu_Q$ and $Q \in \mathcal{S}(\mathcal{Z})$ a.s., we have that $\tilde{G}_\varphi(Q, \bar{V}) = G_\varphi(Q)$ a.s. Also, since

$$\begin{aligned}\tilde{G}_\varphi\left(Q^N, \bar{V}^N\right) = \frac{1}{N} \sum_{j=1}^N \int_0^T &\varphi\left(t, \bar{X}^N_j(t)\right) \cdot b\left(\bar{X}^N_j(t), \bar{V}^N(t)\right) dt \\ &+ \frac{1}{N} \sum_{j=1}^N \int_0^T \varphi\left(t, \bar{X}^N_j(t)\right) \cdot \sigma\left(\bar{X}^N_j(t), \bar{V}^N(t)\right) u^N_j(t)\, dt\end{aligned}$$

and $\varepsilon_N \to 0$, we see that $|\tilde{G}_\varphi(Q^N, \bar{V}^N) - G^*_\varphi(Q^N)| \overset{\bar{P}}{\to} 0$ as $N \to \infty$. Thus it remains to argue that

$$\left| \tilde{G}_\varphi\left(Q^N, \bar{V}^N\right) - \tilde{G}_\varphi\left(Q, \bar{V}\right) \right| \overset{\bar{P}}{\to} 0 \qquad \text{as } N \to \infty. \tag{5.3.20}$$

Now, since

$$\begin{aligned}\tilde{G}^B_\varphi\left(Q^N, \bar{V}^N\right) = \frac{1}{N} \sum_{j=1}^N \int_0^T &\varphi\left(t, \bar{X}^N_j(t)\right) \cdot \eta_B\left(b\left(\bar{X}^N_j(t), \bar{V}^N(t)\right)\right) dt \\ &+ \frac{1}{N} \sum_{j=1}^N \int_0^T \varphi\left(t, \bar{X}^N_j(t)\right) \cdot \sigma\left(\bar{X}^N_j(t), \bar{V}^N(t)\right) \psi_B\left(u^N_j(t)\right) dt,\end{aligned}$$

and the map

$$(\xi, r, V) \mapsto \int_0^T \varphi\left(t, \xi(t)\right) \cdot \eta_B(b(\xi(t), V(t)))\, dt + \int_{[0,T]\times\mathbb{R}^m} \varphi\left(t, \xi(t)\right) \cdot \sigma(\xi(t), V(t)) \psi_B(y)\, r(dt, dy)$$

is bounded and continuous on $\mathcal{Z} \times \mathcal{V}$, the a.s. convergence $(Q^N, \bar{V}^N) \to (Q, \bar{V})$ in $\mathcal{P}(\mathcal{Z}) \times \mathcal{V}$ implies that

$$\tilde{G}^B_\varphi\left(Q^N, \bar{V}^N\right) \to \tilde{G}^B_\varphi\left(Q, \bar{V}\right) \qquad \text{a.s.,} \qquad \text{as } N \to \infty, \tag{5.3.21}$$

for each B. Also, using (5.3.14) and (5.3.15), as in the proof of (5.3.17), we see

$$\begin{aligned}&\left|\tilde{G}_{\varphi}^{B}\left(Q^{N},\bar{V}^{N}\right)-\tilde{G}_{\varphi}\left(Q^{N},\bar{V}^{N}\right)\right|\\&\leq\frac{18K_{\varphi}L^{2}T}{B}\left(1+\frac{1}{N}\sum_{j=1}^{N}\left\|\bar{X}_{j}^{N}\right\|_{\infty}^{2}\right)+\frac{3K_{\varphi}L}{BN}\sum_{j=1}^{N}\int_{0}^{T}\left|u_{j}^{N}(t)\right|^{2}dt,\end{aligned}$$

which in view of Lemma 5.3.1 and the assumption on the controls in Lemma 5.2.3 shows that

$$\sup_{N\geq1}\tilde{E}\left[\left|\tilde{G}_{\varphi}^{B}\left(Q^{N},\bar{V}^{N}\right)-\tilde{G}_{\varphi}\left(Q^{N},\bar{V}^{N}\right)\right|\right]\to0\qquad\text{as } B\to\infty. \tag{5.3.22}$$

Finally, along the same lines as in the proof of (5.3.18),

$$\left|\tilde{G}_{\varphi}^{B}\left(Q,\bar{V}\right)-\tilde{G}_{\varphi}\left(Q,\bar{V}\right)\right|\to0\qquad\text{a.s.,}\qquad\text{as } B\to\infty.$$

Combining the above convergence with (5.3.21) and (5.3.22) shows (5.3.20), which as noted previously establishes that $Q\in\mathcal{P}^{*}(\bar{\mathcal{J}})$ a.s. and thus completes the proof of the lemma. ☐

5.3.2 Proof of Lemma 5.2.5

We first prove an estimate similar to that in Lemma 5.3.1 for the coordinate process $X(t)$ on the space $(\mathcal{Z},\mathcal{B}(\mathcal{Z}),\Theta)$ for each $\Theta\in\mathcal{P}_2(\mathcal{Z})\cap\mathcal{S}(\mathcal{Z})$. By the definition of $\mathcal{S}(\mathcal{Z})$, the coordinate maps (X,ρ) satisfy

$$dX(t)=b\left(X(t),\nu_{\Theta}(t)\right)\,dt+\int_{\mathbb{R}^{m}}\sigma\left(X(t),\nu_{\Theta}(t)\right)y\,\rho_{t}(dy)\,dt\qquad\Theta\text{-a.s.,} \tag{5.3.23}$$

with $X(0)\sim\mu_0$. By Condition 5.1.1,

$$\begin{aligned}\left|b\left(X(t),\nu_{\Theta}(t)\right)\right|^{2}&\leq3L^{2}\left(1+|X(t)|^{2}+\int_{\mathbb{R}^{d}}|x|^{2}\,\nu_{\Theta}(t,dx)\right)\\&=3L^{2}\left(1+|X(t)|^{2}+E_{\Theta}\left[|X(t)|^{2}\right]\right)\end{aligned} \tag{5.3.24}$$

Applying the above bound in (5.3.23), taking expectation, using $|\sigma|\leq L$, and applying Gronwall's inequality, we have

$$E_{\Theta}\left[\|X\|_{\infty}^{2}\right]\leq\tilde{c}\left(1+\int_{\mathbb{R}^{d}}|x|^{2}\,\mu_{0}(dx)+E_{\Theta}\left[\int_{[0,T]\times\mathbb{R}^{m}}|y|^{2}\,\rho(dt,dy)\right]\right)<\infty, \tag{5.3.25}$$

for some $\tilde{c}=\tilde{c}(L,T)<\infty$.

Now fix $\mathbf{s} \in \mathcal{O}_d$ and let $\{(\mu_k, \Theta_k, \mathcal{J}_k)\}$ be a sequence in $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z}) \times \mathbf{H}^{-\mathbf{s}}$ that satisfies the hypotheses of the lemma. Note that, by (5.3.25),

$$\begin{aligned}\sup_{k\geq 1}\int_{\mathcal{X}} \|\psi\|_\infty^2\, \mu_k(d\psi) &= \sup_{k\geq 1}\int_{\mathcal{X}} \|\psi\|_\infty^2\, (\Theta_k)_{(1)}(d\psi) = \sup_{k\geq 1} E_{\Theta_k}\left[\|X\|_\infty^2\right] \\ &\leq \tilde{c}\left(1 + \int_{\mathbb{R}^d} |x|^2\, \mu_0(dx) + \sup_{k\geq 1} E_{\Theta_k}\left[\int_{[0,T]\times\mathbb{R}^m} |y|^2\, \rho(dt, dy)\right]\right) < \infty.\end{aligned} \tag{5.3.26}$$

If τ is a $\{\sigma(X(s), s \leq t)\}$-stopping time on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$ taking values in $[0, T-\varepsilon]$, then for any $\varepsilon > 0$,

$$\begin{aligned}|X(\tau+\varepsilon) - X(\tau)|^2 &\leq 2\left|\int_\tau^{\tau+\varepsilon} b(X(t), \nu_{\Theta_k}(t))\, dt\right|^2 + 2\left|\int_\tau^{\tau+\varepsilon}\int_{\mathbb{R}^m} \sigma(X(t), \nu_{\Theta_k}(t)) y\, \rho_t(dy)\, dt\right|^2 \\ &\leq 6L^2\varepsilon \int_0^T \left(1 + |X(t)|^2 + E_{\Theta_k}\left[|X(t)|^2\right]\right) dt + 2L^2\varepsilon \int_0^T \int_{\mathbb{R}^m} |y|^2\, \rho_t(dy)\, dt,\end{aligned}$$

Θ_k-a.s. for each k. Hence, using the bound in (5.3.25),

$$E_{\Theta_k}\left[|X(\tau+\varepsilon) - X(\tau)|^2\right] \leq 12L^2(1+\tilde{c})\varepsilon\left(1 + \int_{\mathbb{R}^d} |x|^2\, \mu_0(dx) + \sup_{k\geq 1} E_{\Theta_k}\left[\int_{[0,T]\times\mathbb{R}^m} |y|^2\, \rho(dt, dy)\right]\right).$$

If $\mathcal{T}_\varepsilon$ denotes the collection of all such stopping times τ, it follows that

$$\begin{aligned}\sup_{k\geq 1}\sup_{\tau\in\mathcal{T}_\varepsilon}\int_{\mathcal{X}} |\psi(\tau+\varepsilon) - \psi(\tau)|^2\, \mu_k(d\psi) &= \sup_{k\geq 1}\sup_{\tau\in\mathcal{T}_\varepsilon}\int_{\mathcal{X}} |\psi(\tau+\varepsilon) - \psi(\tau)|^2\, (\Theta_k)_{(1)}(d\psi) \\ &= \sup_{k\geq 1}\sup_{\tau\in\mathcal{T}_\varepsilon} E_{\Theta_k}\left[|X(\tau+\varepsilon) - X(\tau)|^2\right] \\ &\to 0\end{aligned}$$

as $\varepsilon \to 0$. This and (5.3.26) prove relative compactness of $\{\mu_k\}$ (and hence of $\{(\Theta_k)_{(1)}\}$) in $\mathcal{P}(\mathcal{X})$. By Lemma 5.3.6 and (5.3.26), we in fact get relative compactness of $\{\mu_k\}$ in $\mathcal{P}_1(\mathcal{X})$.

For the second marginals $\{(\Theta_k)_{(2)}\}$, we recall from the proof of Lemma 5.2.3 that

$$H(\theta) = \int_{\mathcal{R}_1}\int_{[0,T]\times\mathbb{R}^m} |y|^2\, r(dt, dy)\, \theta(dr)$$

has relatively compact level sets on $\mathcal{P}(\mathcal{R}_1)$. Hence, we have relative compactness of $\{(\Theta_k)_{(2)}\}$ in $\mathcal{P}(\mathcal{R}_1)$ on observing that

$$\sup_{k\geq 1} H\left((\Theta_k)_{(2)}\right) = \sup_{k\geq 1} E_{\Theta_k}\left[\int_{[0,T]\times\mathbb{R}^m} |y|^2 \rho(dt, dy)\right] < \infty.$$

This establishes that $\{\Theta_k\}$ is relatively compact in $\mathcal{P}(\mathcal{Z})$.

For $\{\mathcal{J}_k\}$, we employ the following lemma, the proof of which is saved for the Appendix.

Lemma 5.3.7. *Suppose Conditions 5.1.1 and 5.1.2 are satisfied. Also suppose that for some* $\mathbf{s} \in \mathcal{O}_d$ *and* $(\mu, \mathcal{J}) \in \mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}}$, $I(\mu, \mathcal{J}) < \infty$*. Then, for each* $\mathbf{s}' \in \mathcal{O}_d$*, there is a constant* $C_{\mathbf{s}'} < \infty$ *such that for any* $\Theta \in \mathcal{P}^*(\mathcal{J})$ *with* $\Theta_{(1)} = \mu$*, and for all* $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$*,*

$$|\langle \mathcal{J}, \varphi \rangle|^2 \leq E_\Theta \left[\left| \int_0^T \varphi(t, X(t)) \cdot dX(t) \right|^2 \right] \leq C_{\mathbf{s}'} \left(1 + E_\Theta \left[\int_{[0,T]\times\mathbb{R}^m} |y|^2 \, \rho(dt, dy) \right] \right) \|\varphi\|_{\mathbf{s}'}^2,$$

where $C_{\mathbf{s}'}$ *does not depend on* $\mathcal{J}$*,* φ*, or* Θ*. In particular,* $\mathcal{J} \in \mathbf{H}^{-\mathbf{s}'}$ *for all* $\mathbf{s}' \in \mathcal{O}_d$*.*

Recall the collection of test functions $\{g_M, M < \infty\}$ from Definition 5.3.1, which by Lemma 5.3.3 (see (5.3.3)) satisfy

$$\|g_M \varphi\|_{\mathbf{s}} \leq K \|\varphi\|_{\mathbf{s}}, \tag{5.3.27}$$

for all $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$ and $\mathbf{s} \in \mathcal{O}_d$, with $K < \infty$ depending only on $\mathbf{s}$. For each $k \geq 1$ and $M < \infty$, define $\mathcal{J}_k^M, \mathcal{J}_k^{M,c} \in \mathbf{H}^{-\mathbf{s}}$ by

$$\left\langle \mathcal{J}_k^M, \varphi \right\rangle \doteq \langle \mathcal{J}_k, g_M \varphi \rangle, \quad \left\langle \mathcal{J}_k^{M,c}, \varphi \right\rangle \doteq \langle \mathcal{J}_k, \varphi \rangle - \left\langle \mathcal{J}_k^M, \varphi \right\rangle, \quad \varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d).$$

Fix some $\mathbf{s}' \in \mathcal{O}_d$ such that $s_1' < s_1$ and $s_2' < s_2$. Since $\Theta_k \in \mathcal{P}^*(\mathcal{J}_k)$ for each k and (5.2.14) holds, $I(\mu_k, \mathcal{J}_k) < \infty$ for each k, so by Lemma 5.3.7, $\mathcal{J}_k \in \mathbf{H}^{-\mathbf{s}'}$ for each k. Then for each k and M, in view of (5.3.27), $\mathcal{J}_k^M$ and $\mathcal{J}_k^{M,c}$ are in $\mathbf{H}^{-\mathbf{s}'}$ as well, and furthermore,

$$\begin{aligned} \left| \left\langle \mathcal{J}_k^M, \varphi \right\rangle \right|^2 &\leq C_{\mathbf{s}'} \left(1 + E_\Theta \left[\int_{[0,T]\times\mathbb{R}^m} |y|^2 \, \rho(dt, dy) \right] \right) \|g_M \varphi\|_{\mathbf{s}'}^2 \\ &\leq C_{\mathbf{s}'} K^2 \left(1 + E_{\Theta_k} \left[\int_{[0,T]\times\mathbb{R}^m} |y|^2 \, \rho(dt, dy) \right] \right) \|\varphi\|_{\mathbf{s}'}^2, \end{aligned}$$

and hence

$$\sup_{M<\infty, k\geq 1} \left\| \mathcal{J}_k^M \right\|_{-\mathbf{s}'}^2 \leq C_{\mathbf{s}'} K^2 \left(1 + \sup_{k\geq 1} E_{\Theta_k} \left[\int_{[0,T]\times\mathbb{R}^m} |y|^2 \, \rho(dt, dy) \right] \right) < \infty. \tag{5.3.28}$$

Noting that for each M, $\{\mathcal{J}_k^M\}$ are all supported on $[0,T] \times \{|x| \leq M+1\} \subset U \times \mathbb{R}^d$, by Lemma 5.3.4, $\{\mathcal{J}_k^M, k \geq 1\}$ is relatively compact in $\mathbf{H}^{-\mathbf{s}}$. Now define the collection of stopping times $\{\tau^M, M < \infty\}$ on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$ by $\tau^M \doteq \inf\{t > 0 : |X(t)| \geq M\}$. Note that

$$\begin{aligned} \left\langle \mathcal{J}_k^{M,c}, \varphi \right\rangle &= E_{\Theta_k} \left[\int_0^T (1 - g_M(X(t))) \varphi(t, X(t)) \cdot dX(t) \right] \\ &= E_{\Theta_k} \left[1_{\{\tau^M < T\}} \int_0^T (1 - g_M(X(t))) \varphi(t, X(t)) \cdot dX(t) \right], \end{aligned}$$

and so by Lemma 5.3.7 and (5.3.27),

$$\begin{aligned}\left|\left\langle \mathcal{J}_k^{M,c}, \varphi \right\rangle\right|^2 &\leq \Theta_k\left(\tau^M < T\right) E_{\Theta_k}\left[\left|\int_0^T (1 - g_M(X(t)))\varphi(t, X(t)) \cdot dX(t)\right|^2\right] \\ &\leq \Theta_k\left(\tau^M < T\right) C_{\mathbf{s}}\left(1 + E_{\Theta_k}\left[\int_{[0,T]\times\mathbb{R}^m} |y|^2\, \rho(dt, dy)\right]\right) \|(1 - g_M)\varphi\|_{\mathbf{s}}^2 \\ &\leq 2\Theta_k\left(\tau^M < T\right) C_{\mathbf{s}}\left(1 + K^2\right)\left(1 + E_{\Theta_k}\left[\int_{[0,T]\times\mathbb{R}^m} |y|^2\, \rho(dt, dy)\right]\right) \|\varphi\|_{\mathbf{s}}^2,\end{aligned}$$

and hence

$$\begin{aligned}\sup_{k\geq 1}\left\|\mathcal{J}_k^{M,c}\right\|_{-\mathbf{s}}^2 &\leq 2\sup_{k\geq 1}\Theta_k\left(\tau^M < T\right) C_{\mathbf{s}}\left(1 + K^2\right)\left(1 + \sup_{k\geq 1} E_{\Theta_k}\left[\int_{[0,T]\times\mathbb{R}^m} |y|^2\, \rho(dt, dy)\right]\right) \\ &\leq \frac{2}{M^2}\sup_{k\geq 1} E_{\Theta_k}\left[\|X\|_\infty^2\right] C_{\mathbf{s}}\left(1 + K^2\right)\left(1 + \sup_{k\geq 1} E_{\Theta_k}\left[\int_{[0,T]\times\mathbb{R}^m} |y|^2\, \rho(dt, dy)\right]\right) \\ &\to 0\end{aligned} \tag{5.3.29}$$

as $M \to \infty$, by (5.3.25). Then by Lemma 5.3.2 (applied to the constant random variables $\mathcal{J}_k = \mathcal{J}_k^M + \mathcal{J}_k^{M,c}$ on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$), we obtain from (5.3.28) and (5.3.29) that $\{\mathcal{J}_k\}$ is relatively compact in $\mathbf{H}^{-\mathbf{s}}$. Lemma 5.2.5 now follows on combining the above with the relative compactness of $\{(\mu_k, \Theta_k)\}$ in $\mathcal{P}_1(\mathcal{X}) \times \mathcal{P}(\mathcal{Z})$ shown previously. □

5.3.3 Proof of Lemma 5.2.4

Recall that we assume that Conditions 5.1.1, 5.1.2 and 5.1.3 hold. In particular, $\sigma(x, \mu) = \sigma(\mu)$. Let $\Theta_1, \Theta_2 \in \mathcal{S}(\mathcal{Z}) \cap \mathcal{P}_2(\mathcal{Z})$ be such that $\Theta_1 \circ \theta^{-1} = \Theta_2 \circ \theta^{-1}$, and let $\Lambda = \Theta_1 \circ \theta^{-1}$. Then for $j = 1, 2$, we can disintegrate Θ_j as

$$\Theta_j(dx, dr) = \widetilde{\Theta}_j(x_0, r, dx)\, \Lambda(dx_0, dr)$$

for some measurable map $\widetilde{\Theta}_j : \mathbb{R}^d \times \mathcal{R}_1 \to \mathcal{P}(\mathcal{X})$. Define the probability measure Ξ on the space $\mathbb{R}^d \times \mathcal{R}_1 \times \mathcal{X} \times \mathcal{X}$ as

$$\Xi(dx_0, dr, dx_1, dx_2) = \widetilde{\Theta}_1(x_0, r, dx_1)\, \widetilde{\Theta}_2(x_0, r, dx_2)\, \Lambda(dx_0, dr),$$

and let (ξ_0, ρ, X_1, X_2) denote the coordinate maps on this space. Then, $X_1(0) = X_2(0) = \xi_0$, and to prove the lemma it suffices to show that $X_1 = X_2$ Ξ-a.s.

Letting $u(t) = \int_{\mathbb{R}^m} y\, \rho_t(dy)$ and $V_j(t) = \Xi \circ (X_j(t))^{-1}$, we have that $E_\Xi\left[\int_0^T |u(t)|^2\, dt\right] < \infty$ and

$$X_j(t) = \xi_0 + \int_0^t b\left(X_j(s), V_j(s)\right)\, ds + \int_0^t \sigma\left(V_j(s)\right) u(s)\, ds, \qquad j = 1, 2.$$

By the Lipschitz property of the coefficients and the fact that

$$d_1\left(V_1(s), V_2(s)\right)^2 \leq \left(E_\Xi\left[|X_1(s) - X_2(s)|\right]\right)^2 \leq E_\Xi\left[\sup_{0\leq r\leq s}|X_1(r) - X_2(r)|^2\right],$$

it follows from Condition 5.1.1 that for every $0 \leq t \leq T$,

$$\begin{aligned}
|X_1(t) - X_2(t)|^2 &\leq 2T\int_0^t \left|b\left(X_1(s), V_1(s)\right) - b\left(X_2(s), V_2(s)\right)\right|^2 ds \\
&\quad + 2\left(\int_0^T |u(s)|^2\, ds\right)\int_0^T \left|\sigma\left(V_1(s)\right) - \sigma\left(V_2(s)\right)\right|^2 ds \\
&\leq 2L^2T\int_0^t \left(|X_1(s) - X_2(s)| + d_1\left(V_1(s), V_2(s)\right)\right)^2 ds \\
&\quad + 2L^2\left(\int_0^t |u(s)|^2\, ds\right)\int_0^T d_1\left(V_1(s), V_2(s)\right)^2 ds \\
&\leq 4L^2T\int_0^t \sup_{0\leq r\leq s}|X_1(r) - X_2(r)|^2\, ds \\
&\quad + 2L^2\left(2T + \int_0^T |u(t)|^2\, dt\right)\int_0^t E_\Xi\left[\sup_{0\leq r\leq s}|X_1(r) - X_2(r)|^2\right] ds.
\end{aligned}$$

Then taking expectation with respect to Ξ, for all $0 \leq t \leq T$,

$$\begin{aligned}
&E_\Xi\left[\sup_{0\leq s\leq t}|X_1(s) - X_2(s)|^2\right] \\
&\leq 2L^2\left(4T + E_\Xi\left[\int_0^T |u(s)|^2\, ds\right]\right)\int_0^t E_\Xi\left[\sup_{0\leq r\leq s}|X_1(r) - X_2(r)|^2\right] ds.
\end{aligned}$$

Gronwall's inequality now shows that $E_\Xi\left[\|X_1 - X_2\|_\infty^2\right] = 0$, which completes the proof. □

5.4 Proofs of Sobolev Space Results

In this section we provide proofs of some Sobolev space results that are used in our work. It will be convenient to introduce an alternate norm on $\mathbf{H}^\mathbf{s}$ equivalent to (5.1.8), and which is similar to norms used in [73] and [88]. Let $\{e_1, \ldots, e_d\}$ denote the canonical basis in $\mathbb{R}^d$, recall that $U = (a, b) \supset [0, T]$, let $\mathcal{I} \doteq \mathbb{Z} \times \mathbb{R}^d \times \{1, \ldots, d\}$, and define the functions $e_{n,\xi}^k : U \times \mathbb{R}^d \to \mathbb{R}^d$ for $(n, \xi, k) \in \mathcal{I}$ by

$$e_{n,\xi}^k(t, x) = \frac{1}{b-a}e^{2\pi int/(b-a)}e^{2\pi i\xi\cdot x}e_k.$$

Consider the Fourier coefficients of $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$ given by

$$\hat{\varphi}(n,\xi) = \left(\hat{\varphi}_1(n,\xi), \ldots, \hat{\varphi}_d(n,\xi)\right), \qquad \hat{\varphi}_k(n,\xi) = \int_U \int_{\mathbb{R}^d} e^k_{-n,-\xi}(t,x) \cdot \varphi(t,x)\, dx\, dt. \tag{5.4.1}$$

Then an equivalent norm on $\mathbf{H}^{\mathbf{s}}$, $\mathbf{s} = (s_1, s_2) \in \mathbb{R}^2_+$, is given by

$$\|\varphi\|^2_{*,\mathbf{s}} = \sum_{n\in\mathbb{Z}} \int_{\mathbb{R}^d} |\hat{\varphi}(n,\xi)|^2 \left(1+n^2\right)^{s_1} \left(1+|\xi|^2\right)^{s_2}\, d\xi. \tag{5.4.2}$$

5.4.1 Proof of Lemma 5.2.1

From the equivalence of the norms, it suffices to prove the statement in the lemma with $\|\cdot\|_{\mathbf{s}}$ replaced with $\|\cdot\|_{*,\mathbf{s}}$. In what follows, we will abuse notation and denote $\|\cdot\|_{*,\mathbf{s}}$ once more as $\|\cdot\|_{\mathbf{s}}$. Recall that for $N \in \mathbb{N}$, $1 \le j \le N$, and $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$,

$$\bar{J}_j^N(\varphi) = \int_0^T \varphi\left(t, \bar{X}_j^N(t)\right) \circ d\bar{X}_j^N(t).$$

Any such φ can be written in terms of its Fourier coefficients as

$$\varphi(t,x) = \sum_{k=1}^d \sum_{n\in\mathbb{Z}} \int_{\mathbb{R}^d} \hat{\varphi}_k(n,\xi) e^k_{n,\xi}(t,x)\, d\xi.$$

As in [42, Lemma 8] it follows that

$$\bar{J}_j^N(\varphi) = \sum_{k=1}^d \sum_{n\in\mathbb{Z}} \int_{\mathbb{R}^d} \hat{\varphi}_k(n,\xi) Z^N_{j,k}(n,\xi)\, d\xi,$$

where

$$Z^N_{j,k}(n,\xi) \doteq \int_0^T e^k_{n,\xi}\left(t, \bar{X}_j^N(t)\right) \circ d\bar{X}_j^N(t).$$

Note that

$$\begin{aligned}
Z^N_{j,k}(n,\xi) &= \int_0^T e^k_{n,\xi}\left(t, \bar{X}_j^N(t)\right) \cdot d\bar{X}_j^N(t) + \frac{1}{2}\left\langle e^k_{n,\xi}\left(\cdot, \bar{X}_j^N(\cdot)\right), \bar{X}_j^N(\cdot)\right\rangle_T \\
&= \int_0^T e^k_{n,\xi}\left(t, \bar{X}_j^N(t)\right) \cdot b\left(\bar{X}_j^N(t), \bar{V}^N(t)\right)\, dt + \int_0^T e^k_{n,\xi}\left(t, \bar{X}_j^N(t)\right) \cdot \sigma\left(\bar{X}_j^N(t), \bar{V}^N(t)\right) u_j^N(t)\, dt \\
&\quad + \varepsilon_N \int_0^T e^k_{n,\xi}\left(t, \bar{X}_j^N(t)\right) \cdot \sigma\left(\bar{X}_j^N(t), \bar{V}^N(t)\right)\, dW_j(t) \\
&\quad + \pi i \varepsilon_N^2 \xi_k \int_0^T \left(e^k_{n,\xi}\right)_k \left(t, \bar{X}_j^N(t)\right) \left(\sigma\sigma^\mathsf{T}\right)_{kk} \left(\bar{X}_j^N(t), \bar{V}^N(t)\right)\, dt,
\end{aligned}$$

since the kth component $(e^k_{n,\xi})_k$ is the only nonzero component of $e^k_{n,\xi}$. By the Cauchy-Schwarz inequality, for all $\varphi \in \mathcal{C}^\infty_c(U \times \mathbb{R}^d, \mathbb{R}^d)$,

$$\left|\bar{J}^N_j(\varphi)\right|^2 \le \|\varphi\|^2_{\mathbf{s}} \sum_{k=1}^d \sum_{n\in\mathbb{Z}} \int_{\mathbb{R}^d} \frac{\left|Z^N_{j,k}(n,\xi)\right|^2}{(1+n^2)^{s_1}\,(1+|\xi|^2)^{s_2}}\, d\xi = \|\varphi\|^2_{\mathbf{s}} \left(C^N_{j,\mathbf{s}}\right)^2, \tag{5.4.3}$$

where

$$C^N_{j,\mathbf{s}} \doteq \left(\sum_{k=1}^d \sum_{n\in\mathbb{Z}} \int_{\mathbb{R}^d} \frac{\left|Z^N_{j,k}(n,\xi)\right|^2}{(1+n^2)^{s_1}\,(1+|\xi|^2)^{s_2}}\, d\xi\right)^{1/2}.$$

Since $|e^k_{n,\xi}| \le T^{-1}$ and $|\sigma| \le L$, the Burkholder-Davis-Gundy inequality gives

$$\begin{aligned} E\left[\left|Z^N_{j,k}(n,\xi)\right|^2\right] \le 4E\left[\int_0^T \left|b\left(\bar{X}^N_j(t), \bar{V}^N(t)\right)\right|^2 dt\right] + \frac{4L^2}{T} E\left[\int_0^T \left|u^N_j(t)\right|^2 dt\right] + \frac{4\varepsilon^2_N L^2}{T} \\ + \frac{4\pi^2\varepsilon^4_N L^4 \xi^2_k}{T}. \end{aligned} \tag{5.4.4}$$

By the linear growth property of b from Condition 5.1.1,

$$\left|b\left(\bar{X}^N_j(t), \bar{V}^N(t)\right)\right|^2 \le 3L^2 \left(1 + \left|\bar{X}^N_j(t)\right|^2 + \frac{1}{N}\sum_{l=1}^N \left|\bar{X}^N_l(t)\right|^2\right),$$

and from Lemma 5.3.1, $E\left[\|\bar{X}^N_j\|^2_\infty\right] < \infty$ for each $N \in \mathbb{N}$ and $1 \le j \le N$. Using the last two estimates and (5.4.4), we see that

$$\sup_{(n,\xi,k)\in\mathcal{I}} E\left[\left|Z^N_{j,k}(n,\xi)\right|^2\right] < \infty.$$

Thus, for each $N \in \mathbb{N}$ and $1 \le j \le N$, $E[|C^N_{j,\mathbf{s}}|^2] < \infty$ for any $\mathbf{s} \in \mathcal{O}_d$. Following [42], we now have from (5.4.3) the existence of a pathwise realization $\bar{\mathcal{J}}^N$ of $\{\varphi \mapsto \bar{J}^N(\varphi)\}$ in $\mathbf{H}^{-\mathbf{s}}$ for every $N \in \mathbb{N}$ and any $\mathbf{s} \in \mathcal{O}_d$. This proves the first part of the lemma.

For the second part, note that by Lemma 5.3.1,

$$\begin{aligned} E\left[\left|b\left(\bar{X}^N_j(t), \bar{V}^N(t)\right)\right|^2\right] \le 4L^2(c+1)\Bigg(1 + \left|x^N_j\right|^2 + E\left[\int_0^T \left|u^N_j(t)\right|^2 dt\right] + \frac{1}{N}\sum_{l=1}^N \left|x^N_l\right|^2 \\ + E\left[\frac{1}{N}\sum_{l=1}^N \int_0^T \left|u^N_l(t)\right|^2 dt\right]\Bigg). \end{aligned}$$

Thus for some constant $K < \infty$ depending only on d, T, and L,

$$\frac{1}{N}\sum_{j=1}^N \sum_{k=1}^d E\left[\left|Z^N_{j,k}(n,\xi)\right|^2\right] \le K\left(1 + |\xi|^2 + \frac{1}{N}\sum_{j=1}^N \left|x^N_j\right|^2 + E\left[\frac{1}{N}\sum_{j=1}^N \int_0^T \left|u^N_j(t)\right|^2 dt\right]\right).$$

Letting $C_{\mathbf{s}}^N = \frac{1}{N}\sum_{j=1}^N C_{j,\mathbf{s}}^N$, we have from (5.4.3) that, for all $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$,

$$\left|\bar{J}^N(\varphi)\right| \leq \frac{1}{N}\sum_{j=1}^N \left|\bar{J}_j^N(\varphi)\right| \leq C_{\mathbf{s}}^N \|\varphi\|_{\mathbf{s}}.$$

Finally,

$$\begin{aligned} E\left[\left(C_{\mathbf{s}}^N\right)^2\right] &\leq \frac{1}{N}\sum_{j=1}^N E\left[\left(C_{j,\mathbf{s}}^N\right)^2\right] \\ &\leq \sum_{n\in\mathbb{Z}}\int_{\mathbb{R}^d} \frac{K}{(1+n^2)^{s_1}(1+|\xi|^2)^{s_2}}\left(1+|\xi|^2+\sup_{N\geq 1}\frac{1}{N}\sum_{j=1}^N\left|x_j^N\right|^2+\sup_{N\geq 1}E\left[\frac{1}{N}\sum_{j=1}^N\int_0^T\left|u_j^N(t)\right|^2\,dt\right]\right)d\xi, \end{aligned}$$

which is finite by Condition 5.1.2 and (5.2.5) since $\mathbf{s} \in \mathcal{O}_d$. □

5.4.2 Proof of Lemma 5.3.7

As in the proof of Lemma 5.2.1, it suffices to prove the statement in the lemma with $\|\cdot\|_{\mathbf{s}}$ replaced with $\|\cdot\|_{*,\mathbf{s}}$, and once again, abusing notation, we will denote $\|\cdot\|_{*,\mathbf{s}}$ as $\|\cdot\|_{\mathbf{s}}$. Suppose that $\mathbf{s} \in \mathcal{O}_d$ and $(\mu, \mathcal{J}) \in \mathcal{P}_1(\mathcal{X}) \times \mathbf{H}^{-\mathbf{s}}$ are such that $I(\mu, \mathcal{J}) < \infty$. Then there is some $\Theta \in \mathcal{P}^*(\mathcal{J})$ such that $\Theta_{(1)} = \mu$ and

$$\langle \mathcal{J}, \varphi\rangle = G_\varphi(\Theta) = E_\Theta\left[\int_0^T \varphi(t, X(t))\cdot dX(t)\right],$$

for all $\varphi \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$. Furthermore, the estimate (5.3.25) holds for this Θ. By an argument as in the proof of Lemma 5.2.1,

$$\int_0^T \varphi(t, X(t))\cdot dX(t) = \sum_{k=1}^d\sum_{n\in\mathbb{Z}}\int_{\mathbb{R}^d}\hat{\varphi}_k(n,\xi)Z_k(n,\xi)\,d\xi \qquad \Theta\text{-a.s.},$$

where $\hat{\varphi}_k$ is defined in (5.4.1) and

$$\begin{aligned} Z_k(n,\xi) &\doteq \int_0^T e_{n,\xi}^k\left(t, X(t)\right)\cdot dX(t) \\ &= \int_0^T e_{n,\xi}^k\left(t, X(t)\right)\cdot b\left(X(t), \nu_\Theta(t)\right)\,dt + \int_{[0,T]\times\mathbb{R}^m} e_{n,\xi}^k\left(t, X(t)\right)\cdot\sigma\left(X(t),\nu_\Theta(t)\right)y\,\rho(dt,dy) \end{aligned}$$

Θ-a.s. Since $|e_{n,\xi}^k| \leq T^{-1}$, using (5.3.24) we have

$$|Z_k(n,\xi)|^2 \leq \frac{6L^2}{T}\int_0^T\left(1+|X(t)|^2+E_\Theta\left[|X(t)|^2\right]\right)\,dt + \frac{2L^2}{T}\int_{[0,T]\times\mathbb{R}^m}|y|^2\,\rho(dt,dy),$$

and then the bound in (5.3.25) gives

$$\sup_{(n,\xi,k)\in\mathcal{I}} E_\Theta\left[|Z_k(n,\xi)|^2\right] \le c'\left(1+\int_{\mathbb{R}^d}|x|^2\,\mu_0(dx)+E_\Theta\left[\int_{[0,T]\times\mathbb{R}^m}|y|^2\,\rho(dt,dy)\right]\right),$$

for some $c'<\infty$. Thus by the Cauchy-Schwarz inequality, for any $\mathbf{s}'=(s_1',s_2')\in\mathcal{O}_d$ and $\varphi\in\mathcal{C}_c^\infty(U\times\mathbb{R}^d,\mathbb{R}^d)$,

$$\begin{aligned}
|\langle\mathcal{J},\varphi\rangle|^2 &\le E_\Theta\left[\left|\int_0^T\varphi(t,X(t))\cdot dX(t)\right|^2\right]\\
&\le E_\Theta\left[\sum_{k=1}^d\sum_{n\in\mathbb{Z}}\int_{\mathbb{R}^d}\frac{|Z_k(n,\xi)|^2}{(1+n^2)^{s_1'}(1+|\xi|^2)^{s_2'}}\,d\xi\right]\|\varphi\|_{\mathbf{s}'}^2\\
&\le c'\sum_{n\in\mathbb{Z}}\int_{\mathbb{R}^d}\frac{d\xi}{(1+n^2)^{s_1'}(1+|\xi|^2)^{s_2'}}\left(1+\int_{\mathbb{R}^d}|x|^2\,\mu_0(dx)+E_\Theta\left[\int_{[0,T]\times\mathbb{R}^m}|y|^2\,\rho(dt,dy)\right]\right)\|\varphi\|_{\mathbf{s}'}^2\\
&\le C_{\mathbf{s}'}^2\left(1+E_\Theta\left[\int_{[0,T]\times\mathbb{R}^m}|y|^2\,\rho(dt,dy)\right]\right)\|\varphi\|_{\mathbf{s}'}^2
\end{aligned}$$

where

$$C_{\mathbf{s}'}^2\doteq c'\left(1+\int_{\mathbb{R}^d}|x|^2\,\mu_0(dx)\right)\sum_{n\in\mathbb{Z}}\int_{\mathbb{R}^d}\frac{d\xi}{(1+n^2)^{s_1'}(1+|\xi|^2)^{s_2'}}<\infty,$$

since $\mathbf{s}'=(s_1',s_2')\in\mathcal{O}_d$. The result follows. □

5.4.3 Proof of Lemma 5.3.3

We will only consider the case where s is not an integer, the proof for the case when s is an integer is a simpler version of the proof given below. An equivalent norm to $\|\cdot\|_s$ in (5.1.7) can be given as follows (see [39, page 527]): write $s=k+r$ where $k\in\mathbb{N}$ and $r\in(0,1)$. Then, for $h\in H^s(\mathbb{R}^d,\mathbb{R}^d)$, define

$$\|h\|_{*,s}^2\doteq\|h\|_k^2+\sum_{|\alpha|=k}\|D^\alpha h\|_r^2,$$

where $\|\cdot\|_k$ is the usual integer Sobolev norm

$$\|h\|_k^2=\sum_{0\le|\alpha|\le k}\|D^\alpha h\|_{L^2}^2,$$

and $\|\cdot\|_r$ is the fractional Gagliardo-type Sobolev norm

$$\|h\|_r^2=\|h\|_{L^2}^2+[h]_r^2=\int_{\mathbb{R}^d}|h(x)|^2\,dx+\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{|h(x)-h(y)|^2}{|x-y|^{d+2r}}\,dx\,dy. \tag{5.4.5}$$

The norm $\|\cdot\|_{*,s}$ is equivalent to the norm $\|\cdot\|_s$ in (5.1.7) and thus it suffices to prove Lemma 5.3.3 with $\|\cdot\|_s$ replaced with $\|\cdot\|_{*,s}$. Henceforth, abusing notation, we will denote this new norm once more as $\|\cdot\|_s$. Now let f and g_M be as in the statement of the lemma. With $B(k)$ as in Definition 5.3.1(iii), the Leibniz product formula gives, for a multi-index α with $|\alpha| \leq k$,

$$|D^\alpha g_M(x) f(x)| = \left| \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} D^{\alpha-\beta} g_M(x) D^\beta f(x) \right| \leq B(k) \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \left| D^\beta f(x) \right|,$$

and hence for all $M < \infty$

$$\|g_M f\|_k^2 = \sum_{0 \leq |\alpha| \leq k} \int_{\mathbb{R}^d} |D^\alpha g_M(x) f(x)|^2 \, dx \leq c_1 \sum_{0 \leq |\alpha| \leq k} \int_{\mathbb{R}^d} \left| D^\beta f(x) \right|^2 \, dx = c_1 \|f\|_k^2, \tag{5.4.6}$$

for some $c_1 = c_1(k) < \infty$. For the r term we follow the proof of [39, Lemma 5.3]. If $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R})$ is such that $0 \leq \psi \leq B_\psi < \infty$ and $h \in H^r(\mathbb{R}^d, \mathbb{R}^d)$ for some $0 < r < 1$, then $\|\psi h\|_{L^2}^2 \leq B_\psi^2 \|h\|_{L^2}^2$. If L_ψ denotes the Lipschitz constant of ψ, then

$$\begin{aligned}
[\psi h]_r^2 &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|\psi(x)h(x) - \psi(y)h(y)|^2}{|x-y|^{d+2r}} \, dx \, dy \\
&\leq 2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|\psi(x)h(x) - \psi(x)h(y)|^2}{|x-y|^{d+2r}} \, dx \, dy + 2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|\psi(x)h(y) - \psi(y)h(y)|^2}{|x-y|^{d+2r}} \, dx \, dy \\
&\leq 2B_\psi^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|h(x) - h(y)|^2}{|x-y|^{d+2r}} \, dx \, dy + 2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|\psi(x) - \psi(y)|^2 |h(y)|^2}{|x-y|^{d+2r}} \, dx \, dy \\
&\leq 2B_\psi^2 [h]_r^2 + 2L_\psi^2 \int_{\mathbb{R}^d} \int_{\{|x-y| \leq 1\}} \frac{|h(y)|^2}{|x-y|^{d+2(r-1)}} \, dx \, dy + 8B_\psi^2 \int_{\mathbb{R}^d} \int_{\{|x-y| > 1\}} \frac{|h(y)|^2}{|x-y|^{d+2r}} \, dx \, dy \\
&\leq 2B_\psi^2 [h]_r^2 + 2 \left(L_\psi^2 + 4B_\psi^2 \right) c_2 \|h\|_{L^2}^2,
\end{aligned}$$

for $c_2 = c_2(r) < \infty$. In the last line, we used the fact that for some $c_3, c_4 < \infty$ depending on r,

$$\int_{\mathbb{R}^d} \int_{\{|x-y| \leq 1\}} \frac{|h(y)|^2}{|x-y|^{d+2(r-1)}} \, dx \, dy \leq \int_{\mathbb{R}^d} \left(\int_{\{|z| \leq 1\}} \frac{1}{|z|^{d+2(r-1)}} \, dz \right) |h(y)|^2 \, dy \leq c_3 \|h\|_{L^2}^2,$$

since $d + 2(r-1) < d$, and

$$\int_{\mathbb{R}^d} \int_{\{|x-y| > 1\}} \frac{|h(y)|^2}{|x-y|^{d+2r}} \, dx \, dy \leq \int_{\mathbb{R}^d} \left(\int_{\{|z| > 1\}} \frac{1}{|z|^{d+2r}} \, dz \right) |h(y)|^2 \, dy \leq c_4 \|h\|_{L^2}^2,$$

since $d + 2r > d$. Thus we have that

$$\|\psi h\|_r^2 \leq 8 \left(B_\psi^2 + L_\psi^2 \right) (c_2 + 1) \|h\|_r^2.$$

Then, with $B(k)$ as in Definition 5.3.1 and $L(k)$ as in (5.3.2), we obtain that for $|\alpha| = k$,

$$\begin{aligned}
\|D^\alpha g_M f\|_r^2 &= \left\| \sum_{\beta\le\alpha} \binom{\alpha}{\beta} D^{\alpha-\beta} g_M D^\beta f \right\|_r^2 \\
&\le 2^{\alpha!} \sum_{\beta\le\alpha} \binom{\alpha}{\beta}^2 \left\|D^{\alpha-\beta} g_M D^\beta f\right\|_r^2 \\
&\le 2^{\alpha!} 8 \left(B(k)^2 + L(k)^2\right)(c_2+1) \sum_{\beta\le\alpha} \binom{\alpha}{\beta}^2 \left\|D^\beta f\right\|_r^2 .
\end{aligned}$$

Next, for $|\beta| < k$ and some constant $c_5 = c_5(r) < \infty$, we have that

$$\left\|D^\beta f\right\|_r^2 \le c_5 \left\|D^\beta f\right\|_1^2 = c_5 \left\|D^\beta f\right\|_{L^2}^2 + c_5 \sum_{|\alpha|=1} \left\|D^\alpha D^\beta f\right\|_{L^2}^2 \le c_5 \left\|D^\beta f\right\|_{L^2}^2 + c_5 \sum_{|\alpha|=|\beta|+1} \|D^\alpha f\|_{L^2}^2 ,$$

and hence for some $c_6 = c_6(k, r) < \infty$ and all $M < \infty$,

$$\sum_{|\alpha|=k} \|D^\alpha g_M f\|_r^2 \le c_6 \sum_{|\alpha|=k} \|D^\alpha f\|_r^2 + c_6\|f\|_k^2. \tag{5.4.7}$$

Finally, from (5.4.6) and (5.4.7), for all $M < \infty$,

$$\|g_M f\|_s^2 = \|g_M f\|_k^2 + \sum_{|\alpha|=k} \|D^\alpha g_M f\|_r^2 \le (c_1+c_6)\|f\|_k^2 + c_6 \sum_{|\alpha|=k} \|D^\alpha f\|_r^2 \le K\|f\|_s^2,$$

where $K = c_1 + c_6$. □

5.4.4 Proof of Lemma 5.3.4

Let $\mathbf{s}, \mathbf{s}'$, A and K be as in the statement of the lemma. In particular $A \subset \mathbf{H}^{-\mathbf{s}'}$ is such that

$$B \doteq \sup_{F\in A} \|F\|_{-\mathbf{s}'} < \infty, \tag{5.4.8}$$

and every $F \in A$ has support contained in K. Recall the functions $e^k_{n,\xi}$ for $(n, \xi, k) \in \mathcal{I}$ introduced above (5.4.1). Let $\{F^l\}_{l\in\mathbb{N}}$ be a sequence in A, and for $l \in \mathbb{N}$ and $(n, \xi) \in \mathbb{Z} \times \mathbb{R}^d$, let

$$\hat{F}^l(n,\xi) \doteq \left(\hat{F}^l_1(n,\xi), \ldots, \hat{F}^l_d(n,\xi)\right), \qquad \hat{F}^l_k(n,\xi) \doteq \left\langle F^l, e^k_{-n,-\xi}\right\rangle, \qquad 1 \le k \le d. \tag{5.4.9}$$

Since F^l has compact support, the evaluation on the right side of the second equality above is indeed meaningful (see e.g. [45, Theorem 9.8]) and for each $l \in \mathbb{N}$ and $n \in \mathbb{Z}$, $\xi \mapsto \hat{F}^l(n, \xi)$ is in $\mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}^d)$. Also, using (5.4.8) and the compact support property, one can verify (see [45, Theorem 9.22]) that for each $n \in \mathbb{Z}$,

$$\sup_{l\geq 1}\sup_{\xi\in\mathbb{R}^d}\left|\hat{F}^l(n,\xi)\right| < \infty \qquad \text{and} \qquad \sup_{l\geq 1}\sup_{\xi\in\mathbb{R}^d}\left|D_\xi\hat{F}^l(n,\xi)\right| < \infty.$$

Thus, for each $n \in \mathbb{Z}$, $\{\hat{F}^l(n,\cdot), l \in \mathbb{N}\}$ is relatively compact in $\mathcal{C}(\mathbb{R}^d,\mathbb{R}^d)$. By a standard diagonalization procedure, we can pick a subsequence $\{l_j\}$ such that $\{\hat{F}^{l_j}(n,\cdot), j \in \mathbb{N}\}$ converges in $\mathcal{C}(\mathbb{R}^d,\mathbb{R}^d)$ for every n to a limit. We will now show that F^{l_j} is Cauchy in $\mathbf{H}^{-\mathbf{s}}$ which will complete the proof.

By an argument similar to [45, Proposition 9.16], there are constants $c_1(\mathbf{t},K), c_2(\mathbf{t},K) < \infty$ for $\mathbf{t} = \mathbf{s},\mathbf{s}'$ such that for any $F \in \mathbf{H}^{-\mathbf{s}'} \subset \mathbf{H}^{-\mathbf{s}}$ supported on the compact set K and both $\mathbf{t} = (t_1,t_2) = \mathbf{s},\mathbf{s}'$,

$$c_1(\mathbf{t},K)\left\|F\right\|_{-\mathbf{t}}^2 \leq \sum_{n\in\mathbb{Z}}\int_{\mathbb{R}^d}\left|\hat{F}(n,\xi)\right|^2\left(1+n^2\right)^{-t_1}\left(1+|\xi|^2\right)^{-t_2}\,d\xi \leq c_2(\mathbf{t},K)\left\|F\right\|_{-\mathbf{t}}^2, \tag{5.4.10}$$

where $\hat{F}(n,\xi)$ is defined as in (5.4.9). In particular, for $j,m \in \mathbb{N}$,

$$c_1(\mathbf{s},K)\left\|F^{l_j}-F^{l_m}\right\|_{-\mathbf{s}}^2 \leq \sum_{n\in\mathbb{Z}}\int_{\mathbb{R}^d}\left|\hat{F}^{l_j}(n,\xi)-\hat{F}^{l_m}(n,\xi)\right|^2\left(1+n^2\right)^{-s_1}\left(1+|\xi|^2\right)^{-s_2}\,d\xi.$$

Fix $M \in \mathbb{N}$. Then, using $(1+|\xi|^2)^{-s_2} \leq (1+|\xi|^2)^{-s_2'}$, we have

$$\begin{aligned}
c_1(\mathbf{s},K)\left\|F^{l_j}-F^{l_m}\right\|_{-\mathbf{s}}^2 \leq & \sum_{-M\leq n\leq M}\int_{\mathbb{R}^d}\left|\hat{F}^{l_j}(n,\xi)-\hat{F}^{l_m}(n,\xi)\right|^2\left(1+n^2\right)^{-s_1}(1+|\xi|^2)^{-s_2}\,d\xi \\
& + \sum_{|n|>M}\int_{\mathbb{R}^d}\left|\hat{F}^{l_j}(n,\xi)-\hat{F}^{l_m}(n,\xi)\right|^2\left(1+n^2\right)^{-s_1}\left(1+|\xi|^2\right)^{-s_2}\,d\xi \\
\leq & \sum_{-M\leq n\leq M}\int_{\mathbb{R}^d}\left|\hat{F}^{l_j}(n,\xi)-\hat{F}^{l_m}(n,\xi)\right|^2\left(1+n^2\right)^{-s_1}\left(1+|\xi|^2\right)^{-s_2}\,d\xi \\
& + c_2(\mathbf{s}',K)\left\|F^{l_j}-F^{l_m}\right\|_{\mathbf{s}'}^2\frac{1}{(1+(M+1)^2)^{s_1-s_1'}} \\
\leq & \sum_{-M\leq n\leq M}\int_{\mathbb{R}^d}\left|\hat{F}^{l_j}(n,\xi)-\hat{F}^{l_m}(n,\xi)\right|^2\left(1+n^2\right)^{-s_1}\left(1+|\xi|^2\right)^{-s_2}\,d\xi \\
& + \frac{4B^2c_2(\mathbf{s}',K)}{(1+(M+1)^2)^{s_1-s_1'}}.
\end{aligned}$$

Next, for each $|n| \leq M$ and $R < \infty$, there is a $C(R) < \infty$ such that

$$\begin{aligned}
&\int_{\mathbb{R}^d}\left|\hat{F}^{l_j}(n,\xi)-\hat{F}^{l_m}(n,\xi)\right|^2\left(1+n^2\right)^{-s_1}\left(1+|\xi|^2\right)^{-s_2}\,d\xi \\
&= \int_{\{|\xi|\leq R\}}\left|\hat{F}^{l_j}(n,\xi)-\hat{F}^{l_m}(n,\xi)\right|^2\left(1+n^2\right)^{-s_1}\left(1+|\xi|^2\right)^{-s_2}\,d\xi
\end{aligned}$$

$$+ \int_{\{|\xi|>R\}} \left|\hat{F}^{l_j}(n,\xi) - \hat{F}^{l_m}(n,\xi)\right|^2 \left(1+n^2\right)^{-s_1} \left(1+|\xi|^2\right)^{-s_2} \, d\xi$$

$$\leq C(R) \sup_{|\xi|\leq R} \left|\hat{F}^{l_j}(n,\xi) - \hat{F}^{l_m}(n,\xi)\right|^2 + \frac{c_2(\mathbf{s}',K)}{(1+R^2)^{s_2-s_2'}} \left\|F^{l_j} - F^{l_m}\right\|_{\mathbf{s}'}^2$$

$$\leq C(R) \sup_{|\xi|\leq R} \left|\hat{F}^{l_j}(n,\xi) - \hat{F}^{l_m}(n,\xi)\right|^2 + \frac{4B^2 c_2(\mathbf{s}',K)}{(1+R^2)^{s_2-s_2'}}.$$

Combining the above estimates and sending $j, m \to \infty$, since $\{\hat{F}^{l_j}(n,\cdot)\}$ converges for every n, we get

$$\limsup_{j,m\to\infty} \left\|F^{l_j} - F^{l_m}\right\|_{\mathbf{s}}^2 \leq \frac{4B^2(2M+1)c_2(\mathbf{s}',K)}{c_1(\mathbf{s},K)\,(1+R^2)^{s_2-s_2'}} + \frac{4B^2 c_2(\mathbf{s}',K)}{c_1(\mathbf{s},K)\,(1+(M+1)^2)^{s_1-s_1'}}.$$

The result now follows on first sending $R \to \infty$ and then $M \to \infty$. $\square$

APPENDIX

Some supplementary results related to this dissertation work are presented below for completeness.

A.1 Proofs of Background Theorems

Here we provide the proofs of Theorems 1.1.4, 1.1.5, and 1.2.3 from the Introduction.

Proof of Theorem 1.1.4. The proof of part (i) relies on an application of Komolgorov's extension theorem (see [69, Theorems 14.32 and 14.35]), which in its simplest form says that if $\{(\Omega_i, \mathcal{G}_i) : i \in \mathbb{N}\}$ is a collection of Borel spaces and $\{P_i : i \in \mathbb{N}\}$ is collection of probability measures that is consistent in the sense that

$$P_i(A) = P_j(A \times \Omega_{i+1} \times \cdots \times \Omega_j)$$

whenever $i \leq j$ and $A \in \bigotimes_{k=0}^{i} \mathcal{G}_k$, the product σ-field of $\mathcal{G}_0, \ldots, \mathcal{G}_i$, then there exists a unique probability measure P_∞ such that

$$P_\infty\left(A \times \bigtimes_{k=i+1}^{\infty} \Omega_k\right) = P_i(A)$$

for $A \in \bigotimes_{k=0}^{i} \mathcal{G}_k$.

For each $n \in \mathbb{N}$, define P_n^* on $(\Omega, \mathcal{F}_n)$ by $P_n^*(A) = E[1(A)L_n]$. The martingale property implies consistency in this case: if $k \leq n$ and $A \in \mathcal{F}_k$, then

$$P_n^*(A) = E[1(A)L_n] = E\left[1(A)E\left[L_n \middle| \mathcal{F}_k\right]\right] = E[1(A)L_k] = P_k^*(A).$$

This completes (i).

For part (ii), fix $n \in \mathbb{N}$ and suppose first that $A \subset \{\tau \leq n\}$. By the martingale property,

$$L_\tau 1(\tau \leq n) = E[L_n | \mathcal{F}_\tau] 1(\tau \leq n),$$

and hence

$$P(A) = E\left[1(A)L_\tau L_\tau^{-1}\right] = E\left[1(A)E[L_n|\mathcal{F}_\tau]L_\tau^{-1}\right] = E\left[1(A)L_n L_\tau^{-1}\right] = E^*\left[1(A)L_\tau^{-1}\right]$$

by the construction of P^*. Now, for arbitrary $\mathcal{F}_\tau$-measurable $A \subset \{\tau < \infty\}$, the monotone convergence theorem gives

$$P(A) = \lim_{n\to\infty} P(A \cap \{\tau \leq n\}) = \lim_{n\to\infty} E^*\left[1(A \cap \{\tau \leq n\})L_\tau^{-1}\right] = E^*\left[1(A)L_\tau^{-1}\right]. \qquad \square$$

Proof of Theorem 1.1.5. The idea is to use the representation (1.1.7) to derive a renewal equation and apply Theorem 1.1.2. Note first that for $t > 0$, $\tau_+ \le \tau(t)$ and $\tau_+ = \tau(t)$ on $\{S_{\tau_+} > t\}$. In particular, since the increments of the random walk are independent, $S_{\tau(t)} - S_{\tau_+}$ is independent of S_{τ_+} and for any $s \le t$, $(S_{\tau(t)} - S_{\tau_+} | S_{\tau_+} = s) \stackrel{\mathcal{D}}{=} S_{\tau(t-s)}$ under $\widetilde{P}$. Letting $K(t) = e^{\alpha t} P(W > t)$, we have

$$\begin{aligned} K(t) &= \widetilde{E}\left[e^{-\alpha(S_{\tau(t)}-t)}\right] \\ &= \widetilde{E}\left[e^{-\alpha(S_{\tau_+}-t)} 1(S_{\tau_+} > t)\right] + \widetilde{E}\left[e^{-\alpha(S_{\tau(t)}-t)} 1(S_{\tau_+} \le t)\right] \\ &= \widetilde{E}\left[e^{-\alpha(S_{\tau_+}-t)} 1(S_{\tau_+} > t)\right] + \int_0^t \widetilde{E}\left[e^{-\alpha(S_{\tau(t)}-t)} \middle| S_{\tau_+} = s\right] \widetilde{P}\left(S_{\tau_+} \in ds\right) \\ &= \widetilde{E}\left[e^{-\alpha(S_{\tau_+}-t)} 1(S_{\tau_+} > t)\right] + \int_0^t \widetilde{E}\left[e^{-\alpha(S_{\tau(t-s)}-(t-s))}\right] \widetilde{P}\left(S_{\tau_+} \in ds\right) \\ &= \widetilde{E}\left[e^{-\alpha(S_{\tau_+}-t)} 1(S_{\tau_+} > t)\right] + \int_0^t K(t-s)\, \widetilde{P}\left(S_{\tau_+} \in ds\right). \end{aligned}$$

Consequently, the key renewal theorem gives

$$\lim_{t\to\infty} e^{\alpha t} P(W > t) = \frac{1}{\widetilde{E}\left[S_{\tau_+}\right]} \int_0^\infty \widetilde{E}\left[e^{-\alpha(S_{\tau_+}-s)} 1(S_{\tau_+} > s)\right] ds = \frac{\widetilde{E}\left[1 - e^{-\alpha S_{\tau_+}}\right]}{\alpha \widetilde{E}\left[S_{\tau_+}\right]}.$$

Finally, by Wald's identity, $\widetilde{E}\left[S_{\tau_+}\right] = \widetilde{E}[\tau_+]\widetilde{E}[X_1] = m'(\alpha)\widetilde{E}[\tau_+]$. □

Proof of Theorem 1.2.3. First assume that $\{X_n\}$ satisfies the LDP. Fix $F \in \mathcal{C}_b(S)$ and let $M = \sup_{x\in S} |F(x)| \in (0, \infty)$. Let $\delta > 0$ and $L > 0$. Because F is continuous and I is lower semi-continuous, for each x such that $I(x) \le L$, we can find $r > 0$ such that

$$\sup_{y\in B(x,r)} (-F(y)) \le -F(x) + \frac{\delta}{2} \tag{A.1.1}$$

and

$$\inf_{y\in B(x,r)} I(y) \ge I(x) - \frac{\delta}{2}, \tag{A.1.2}$$

where $B(x, r) = \{y \in S : d(x, y) < r\}$ is the open ball around x with radius r in S. let $\overline{B}(x, r)$ denote the closure of $B(x, r)$ in S. Then, since $\{X_n\}$ satisfies the large deviation upper bound,

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{a_n} \log E\left[e^{-a_n F(X_n)} 1(X_n \in B(x, r))\right] &\le \limsup_{n\to\infty} \frac{1}{a_n} \log\left(e^{a_n(-F(x)+\delta)} P(X_n \in B(x, r))\right) \\ &\le -F(x) + \frac{\delta}{2} + \limsup_{n\to\infty} \frac{1}{a_n} \log P\left(X_n \in \overline{B}(x, r)\right) \\ &\le -F(x) + \frac{\delta}{2} - \inf_{y\in\overline{B}(x,r)} I(y) \\ &\le -(F(x) + I(x)) + \delta. \end{aligned}$$

Now, since $\{x \in S : I(x) \leq L\}$ is compact, we can let $\{B(x_j, r_j) : 1 \leq j \leq k\}$ be a finite cover such that (A.1.1) and (A.1.2) hold for each (x_j, r_j) pair. Let $A = \bigcap_{j=1}^{k} B(x_j, r_j)^c$ and note that A is a closed set and $I(y) > L$ for every $y \in A$. Then,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{a_n} \log E\left[e^{-a_n F(X_n)}\right] \\
&\leq \limsup_{n\to\infty} \frac{1}{a_n} \log \left(\sum_{j=1}^{k} E\left[e^{-a_n F(X_n)} 1(X_n \in B(x_j, r_j))\right] + E\left[e^{-a_n F(X_n)} 1\left(X_n \in A\right)\right] \right) \\
&\leq \limsup_{n\to\infty} \frac{1}{a_n} \log \left((k+1) \max_{1\leq j\leq k} E\left[e^{-a_n F(X_n)} 1(X_n \in B(x_j, r_j))\right] \vee E\left[e^{-a_n F(X_n)} 1\left(X_n \in A\right)\right] \right) \\
&= \limsup_{n\to\infty} \left(\max_{1\leq j\leq k} \frac{1}{a_n} \log E\left[e^{-a_n F(X_n)} 1(X_n \in B(x_j, r_j))\right] \right) \vee \frac{1}{a_n} \log E\left[e^{-a_n F(X_n)} 1\left(X_n \in A\right)\right] \\
&\leq \max_{1\leq j\leq k} \left(-(F(x_j) + I(x_j)) + \delta\right) \vee \left(M - \inf_{y\in A} I(y) \right) \\
&\leq \left(-\inf_{x\in S} (F(x) + I(x)) + \delta \right) \vee (M - L)\,. \qquad \text{(A.1.3)}
\end{aligned}$$

In the above estimates, we used the fact that for a finite collection $\{f_1, \ldots, f_k\}$ of functions

$$\limsup_{n\to\infty} \max_{1\leq j\leq k} f_j(n) = \max_{1\leq j\leq k} \limsup_{n\to\infty} f_j(n).$$

Sending $\delta \to 0$ and $L \to \infty$ in (A.1.3) gives that

$$\limsup_{n\to\infty} \frac{1}{a_n} \log E\left[e^{-a_n F(X_n)}\right] \leq -\inf_{x\in S} (F(x) + I(x)). \qquad \text{(A.1.4)}$$

For the reverse inequality, let $x \in S$ and $\delta > 0$. Since F is continuous, we can find $r > 0$ such that

$$\inf_{y\in B(x,r)} (-F(y)) \geq -F(x) - \delta.$$

Then, since $\{X_n\}$ satisfies the large deviation lower bound,

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{a_n} \log E\left[e^{-a_n F(X_n)}\right] &\geq \liminf_{n\to\infty} \frac{1}{a_n} \log E\left[e^{-a_n F(X_n)} 1(X_n \in B(x,r))\right] \\
&\geq -F(x) - \delta \liminf_{n\to\infty} \frac{1}{a_n} \log P(X_n \in B(x,r)) \\
&\geq -F(x) - \delta - \inf_{y\in B(x,r)} I(y) \\
&\geq -(F(x) + I(x)) - \delta.
\end{aligned}$$

Taking the supremum over x and letting $\delta \to 0$ we obtain

$$\liminf_{n\to\infty} \frac{1}{a_n} \log E\left[e^{-a_n F(X_n)}\right] \geq -\inf_{x\in S}(F(x) + I(x)),$$

which along with (A.1.4) establishes that $\{X_n\}$ satisfies the Laplace principle.

Now suppose that $\{X_n\}$ satisfies the Laplace principle. We want to show that it also satisfies the large deviation upper and lower bounds. We begin with the former. Given a closed set A, define the function

$$\varphi(x) = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{if } x \notin A, \end{cases}$$

and note that φ is nonnegative and lower semi-continuous. For $j \in \mathbb{N}$ let

$$F_j(x) = j(d(x, A) \wedge 1),$$

where $d(x, A) = \inf_{y\in A} d(x, y)$, so that F_j is a bounded continuous function and F_j increases to φ pointwise as $j \to \infty$. It follows by the Laplace upper bound that for each j,

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{a_n} \log P(X_n \in A) &= \limsup_{n\to\infty} \frac{1}{a_n} \log E\left[e^{-a_n \varphi(X_n)}\right] \\ &\leq \limsup_{n\to\infty} \frac{1}{a_n} \log E\left[e^{-a_n F_j(X_n)}\right] \\ &\leq -\inf_{x\in S}(F_j(x) + I(x)). \end{aligned} \tag{A.1.5}$$

Note that for each j,

$$\inf_{x\in S}(F_j(x) + I(x)) \leq \inf_{x\in S}(\varphi(x) + I(x)) = \inf_{x\in S} I(x). \tag{A.1.6}$$

Furthermore, if $\inf_{x\in S} I(x) = 0$, then trivially $\inf_{x\in S}(F_j(x) + I(x)) \geq \inf_{x\in S} I(x)$. If $\inf_{x\in S} I(x) > 0$, then

$$\liminf_{j\to\infty} \inf_{x\in A^c}(F_j(x) + I(x)) \geq \inf_{x\in A} I(x). \tag{A.1.7}$$

To see why, suppose not, that is, for some $L < \inf_{x\in A} I(x)$ we have $\liminf_{j\to\infty} \inf_{x\in A^c}(F_j(x) + I(x)) < L$. In particular, this implies the existence of a subsequence $\{j\} \subset \mathbb{N}$ and $0 < \varepsilon < L$ such that or each j there is $x_j \in A^c$ such that

$$F_j(x_j) + I(x_j) \leq L - \varepsilon. \tag{A.1.8}$$

(A.1.8) implies that $d(x_j, A) \to 0$ as $j \to \infty$ along the subsequence, since otherwise there would exist a subsubsequence on which $F_j(x_j) = j(d(x_j, A) \wedge 1) \to \infty$, a contradiction.

Now, $d(x_j, A) \to 0$ implies the existence of a sequence $\{y_j\} \subset A$ such that $d(x_j, y_j) \to 0$ as $j \to \infty$, and since $\sup_j I(x_j) \leq L - \varepsilon$ from (A.1.8) and the fact that I has compact level sets, there must be a further subsequence $\{x'_j\} \subset \{x_j\}$ and

$$x^* \in \{x \in S : I(x) \leq L - \varepsilon\} \tag{A.1.9}$$

such that $d(x'_j, x^*) \to 0$. This in turn implies that $d(y_j, x^*) \to 0$, and since $\{y_j\} \subset A$ and A is a closed set, $x^* \in A$. Consequently, $I(x^*) \geq \inf_{y \in A} I(y)$, however by (A.1.9) and the fact that $L < \inf_{x \in A} I(x)$, this leads to a contradiction, establishing the inequality (A.1.7).

Then, since $F_j(x) = 0$ when $x \in A$,

$$\inf_{x \in S}(F_j(x) + I(x)) = \inf_{x \in A} I(x) \wedge \inf_{x \in A^c}(F_j(x) + I(x)),$$

and from (A.1.7),

$$\liminf_{j \to \infty} \inf_{x \in S}(F_j(x) + I(x)) \geq \inf_{x \in A} I(x),$$

which establishes the large deviation upper bound when combined with (A.1.5) and (A.1.6).

It remains to show that $\{X_n\}$ satisfies the large deviation lower bound. Let $G \subset S$ be open and suppose that $\inf_{x \in G} I(x) < \infty$, since otherwise the bound is trivial. Let $x \in G$ such that $I(x) < \infty$ and let $M > I(x)$. Choose r small enough so that $B(x, r) \subset G$, which can be done since G is open. Let

$$F(y) = M\left(\frac{d(x, y)}{r} \wedge 1\right),$$

which defines a bounded, continuous function. In particular, $0 \leq F \leq M$ and $F(y) = M$ for $y \notin B(x, r)$, so

$$E\left[e^{-a_n F(X_n)}\right] \leq e^{-a_n M} + P(X_n \in B(x, r)),$$

and hence

$$\log E\left[e^{-a_n F(X_n)}\right] \leq \log\left(e^{-a_n M} + P(X_n \in B(x, r))\right) \leq (-M) \vee \log P(X_n \in B(x, r)).$$

The Laplace principle then implies that

$$(-M) \vee \liminf_{n \to \infty} \frac{1}{a_n} \log P(X_n \in B(x, r)) \geq \liminf_{n \to \infty} \frac{1}{a_n} \log E\left[e^{-a_n F(X_n)}\right]$$

$$
\begin{aligned}
&= -\inf_{y\in S}(F(y)+I(y)) \\
&\geq -(F(x)+I(x)) \\
&= -I(x),
\end{aligned}
$$

since $F(x) = 0$ by construction. Since $M > I(x)$ It follows that

$$
\liminf_{n\to\infty} \frac{1}{a_n} \log P(X_n \in G) \geq \liminf_{n\to\infty} \frac{1}{a_n} \log P(X_n \in B(x,r)) \geq -I(x).
$$

taking the supremum over all x such that $I(x) < \infty$ completes the proof. □

A.2 Useful Lemmas

The following four results are used frequently in Chapters 4 and 5, and we state them here for reference.

Lemma A.2.1 (Grönwall's Inequality)**.** *For $T \leq \infty$, suppose $u : [0,T] \to \mathbb{R}$ is nondecreasing and $v : [0,T] \to \mathbb{R}$ is continuous. If $f : [0,T] \to \mathbb{R}$ is continuous and satisfies*

$$
f(t) \leq u(t) + \int_0^t v(s) f(s)\, ds, \qquad 0 \leq t \leq T,
$$

then

$$
f(t) \leq u(t) e^{\int_0^t v(s)\, ds}, \qquad 0 \leq t \leq T.
$$

For a proof of the following, see [67, Theorem 3.28].

Lemma A.2.2 (Burkholder-Davis-Gundy Inequality)**.** *Let $p \geq 1$. There exist positive constants c_p and C_p such that for any local martingale $\{X_t : t \geq 0\}$ with $X_0 = 0$ a.s. and any stopping time τ,*

$$
c_p E\left[\langle X\rangle_\tau^{p/2}\right] \leq E\left[\sup_{0\leq t\leq\tau} |X_t|^p\right] \leq C_p E\left[\langle X\rangle_\tau^{p/2}\right],
$$

where $\langle X\rangle_t$ denotes the quadratic variation of X at time t.

The next result is due to Aldous [2]. For a proof, see [70, Theorem 2.7].

Lemma A.2.3 (Aldous-Kurtz Criterion)**.** *Let $\mathcal{D}([0,T],S)$ be the space of right continuous functions with finite left limits from $[0,T]$ into a Polish space (S,d). Let $\{X_n : n \in \mathbb{N}\}$ be a sequence of processes with paths in $\mathcal{D}([0,T],S)$, and for each n let $\mathcal{F}_n(t) = \sigma(X_n(s) : 0 \leq s \leq t)$. If*

(a) $\{X_n(t) : n \in \mathbb{N}\}$ is tight in S for each rational $t \in [0,T]$, and

(b) for any sequence $\{\tau_n : n \in \mathbb{N}\}$ such that τ_n is an $\mathcal{F}_n(t)$-stopping time bounded by T and any nonnegative sequence $\{\delta_n : n \in \mathbb{N}\}$ such that $\delta_n \to 0$ as $n \to \infty$,

$$d\left(X_n(\tau_n + \delta_n), X_n(\tau_n)\right) \xrightarrow{P} 0$$

as $n \to \infty$,

then $\{X_n : n \in \mathbb{N}\}$ is tight in $\mathcal{D}([0,T], S)$. The result is also true when $\mathcal{D}([0,T], S)$ is replaced with $\mathcal{C}([0,T], S)$.

The following is a corollary of the above lemma which gives the form of the criterion employed in Chapters 4 and 5.

Corollary A.2.4. *Let $(S, \|\cdot\|_S)$ be a separable Banach space (i.e. a normed Polish space), let $\mathcal{C} = \mathcal{C}([0,T], S)$, and let $\{\mu_n : n \in \mathbb{N}\}$ be a $\mathcal{P}(\mathcal{C})$-valued sequence of random variables. Recall the norm $\|\psi\|_\infty = \sup_{0 \le t \le T} \|\psi(t)\|_S$ on $\mathcal{C}$, and let $\nu_n = E[\mu_n]$ for each n. Let $\{X(t) : 0 \le t \le T\}$ denote the coordinate process on the space $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$, namely*

$$X(t, \psi) = \psi(t), \qquad \psi \in \mathcal{C},$$

let $\mathcal{F}(t) = \sigma(X(s) : 0 \le s \le t)$, and for each $\delta > 0$ let $\mathcal{T}_\delta$ be the set of $\mathcal{F}(t)$-stopping times bounded by $T - \delta$. If

(a) $\sup_{n \ge 1} \int_{\mathcal{C}} \|\psi\|_\infty \, \nu_n(d\psi) < \infty$, and

(b) $\lim_{\delta \to 0} \sup_{n \ge 1} \sup_{\tau \in \mathcal{T}_\delta} \int_{\mathcal{C}} \|\psi(\tau + \delta) - \psi(\tau)\|_S \wedge 1 \, \nu_n(d\psi) = 0$,

then $\{\mu_n\}$ is tight.

Proof. By [21, Theorem 2.11], tightness of $\{\mu_n\}$ as random variables follows from relative compactness of $\{\nu_n\}$ in $\mathcal{P}(\mathcal{C})$, which is equivalent to tightness of the $\mathcal{C}$-valued process $\{X_n\}$, where $X_n \sim \nu_n$ for each n. From (a),

$$\sup_{n \ge 1} \nu_n\left(\|\psi(t)\|_S > M\right) \le \frac{1}{M} \sup_{n \ge 1} \int_{\mathcal{C}} \|\psi\|_\infty \, \nu_n(d\psi) \to 0$$

as $M \to \infty$, and hence $\{X_n(t)\}$ is tight in S for each t, which satisfies part (a) of Lemma A.2.3.

Now note that for any sequence $\{Y_m\}$ of nonnegative random variables, $Y_m \xrightarrow{P} 0$ if and only if $E[Y_m \wedge 1] \to 0$. Then, if $\{\delta_m\}$ is a sequence such that $\delta_m \downarrow 0$ and $\{\tau_m\}$ is a squence of $\mathcal{F}(t)$ stopping times such that $\tau_m \le T - \delta_m$ for each m, then from (b),

$$\begin{aligned} E\left[\|X_m(\tau_m + \delta_m) - X_m(\tau_m)\|_S \wedge 1\right] &\le \sup_{n \ge 1} \sup_{\tau \in \mathcal{T}_{\delta_m}} E\left[\|X_n(\tau + \delta_m) - X_n(\tau)\|_S \wedge 1\right] \\ &= \sup_{n \ge 1} \sup_{\tau \in \mathcal{T}_{\delta_m}} \int_{\mathcal{C}} \|\psi(\tau + \delta) - \psi(\tau)\|_S \wedge 1 \, \nu_n(d\psi) \end{aligned}$$

$$\to 0$$

as $m \to \infty$, which satisfies part (b) of Lemma A.2.3. $\square$

A.3 The Space $\mathbf{H}^{\mathbf{s}}$

Here we give an alternate characterization of the Hilbert Sobolev space $\mathbf{H}^{\mathbf{s}}$ defined in Chapter 5. As the next lemma shows, when $\mathbf{s} \in (0,1) \times \mathbb{R}_+$, we can identify

$$\mathbf{H}^{\mathbf{s}} = H^{s_1}\left(U, H^{s_2}\left(\mathbb{R}^d, \mathbb{R}^d\right)\right) = H^{s_1}\left(U, \mathbb{R}\right) \otimes H^{s_2}\left(\mathbb{R}^d, \mathbb{R}^d\right),$$

where $\otimes$ denotes the Hilbert tensor product. When we identify two Hilbert spaces $H = K$, we mean that there exists an isomorphism $\Phi : H \to K$, i.e. a one-to-one and onto map that preserves inner product, also called a unitary transformation.

This characterization also verifies that the norm $\|\cdot\|_{*,\mathbf{s}}$ defined in (5.4.2) is equivalent to the norm $\|\cdot\|_{\mathbf{s}}$ used in most of Chapter 5 when $\mathbf{s} \in (0,1) \times \mathbb{R}_+$, since $\|\cdot\|_{*,\mathbf{s}}$ is the tensor product norm.

Lemma A.3.1. *Let K be a Hilbert space over $\mathbb{R}$ with inner product $\langle \cdot, \cdot \rangle_K$, and for $s \in (0,1)$ and any domain $U \subset \mathbb{R}$, let $H^s(U,K)$ be the Hilbert space of functions in $L^2(U,K)$ with inner product*

$$\langle f, g \rangle_{H^s(U,K)} = \int_U \langle f(u), g(u) \rangle_K \, du + \int_U \int_U \frac{\langle f(u) - f(v), g(u) - g(v) \rangle_K}{|u-v|^{1+2s}} \, du \, dv.$$

Also let $H^s(U,\mathbb{R})$ be the Hilbert space of functions in $L^2(U,\mathbb{R})$ with the inner product the same as above with $\langle \cdot, \cdot \rangle_K$ replaced with the standard product on $\mathbb{R}$. Then we can identify

$$H^s(U,K) = H^s(U,\mathbb{R}) \otimes K,$$

where $\otimes$ denotes the Hilbert tensor product.

Proof. It suffices to find a unitary transformation $\Psi : H^s(U,\mathbb{R}) \otimes K \to H^s(U,K)$. For a pure tensor $f \otimes x \in H^s(U,\mathbb{R}) \otimes K$, let

$$\Psi(f \otimes x) = f(\cdot)x \in H^s(U,K).$$

The bilinearity of $\otimes$ implies that Ψ is linear on the span of all the pure tensors, i.e. all finite linear combinations of the form $\sum_{j=1}^n f_j \otimes x_j$, and this is a dense subset of $H^s(U,\mathbb{R}) \otimes K$. To see that Ψ is an isometry, note that for any

$f \otimes x, g \otimes y \in H^s(U, \mathbb{R}) \otimes K$,

$$\begin{aligned}
&\langle \Psi(f \otimes x), \Psi(g \otimes y) \rangle_{H^s(U,K)} \\
&= \langle f(\cdot)x, g(\cdot)y \rangle_{H^s(U,K)} \\
&= \int_U \langle f(u)x, g(u)y \rangle_K \, du + \int_U \int_U \frac{\langle f(u)x - f(v)x, g(u)y - g(v)y \rangle_K}{|u-v|^{1+2s}} \, du \, dv \\
&= \left(\int_U f(u)g(u) \, du \right) \langle x, y \rangle_K + \left(\int_U \int_U \frac{(f(u)x - f(v))(g(u) - g(v))}{|u-v|^{1+2s}} \, du \, dv \right) \langle x, y \rangle_K \\
&= \langle f, g \rangle_{H^s(U,\mathbb{R})} \langle x, y \rangle_K \\
&= \langle f \otimes x, g \otimes y \rangle_{H^s(U,\mathbb{R}) \otimes K}.
\end{aligned}$$

By linearity, this extends to all finite linear combinations of pure tensors.

Now, we can approximate any function in $H^s(U, K) \subset L^2(U, K)$ by an elementary function of the form $h = \sum_{j=1}^n 1_{U_j}(\cdot)x_j$ for $U_j \subset U$ measurable and $x_j \in K$. Note that

$$h = \sum_{i=1}^n 1_{U_j}(\cdot)x_j = \sum_{j=1}^n \Psi(1_{U_j} \otimes x_j) = \Psi\left(\sum_{j=1}^n 1_{U_j} \otimes x_j \right),$$

and hence the range of Ψ on a dense subset of $H^s(U, \mathbb{R}) \otimes K$ is dense in $H^s(U, K)$. Thus Ψ can be extended to an isometry on the whole space which is one-to-one and onto. □

At various points in Chapter 5, the fact that $\mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$ is dense in $\mathbf{H}^\mathbf{s}$ when $\mathbf{s} \in \mathbb{R}_+^2$ is used. The characertization of the previous lemma allows this to be verified from the well-known dense inclusions $\mathcal{C}_c^\infty(U, \mathbb{R}) \subset H^{s_1}(U, \mathbb{R})$ and $\mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R}^d) \subset H^{s_2}(\mathbb{R}^d, \mathbb{R}^d)$, as shown in the following lemma.

Lemma A.3.2. *$\mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$ is dense in $\mathbf{H}^\mathbf{s}$ for $s_1, s_2 > 0$.*

Proof. Since $\mathbf{H}^\mathbf{s} = H^{s_1}(U, \mathbb{R}) \otimes H^{s_2}(\mathbb{R}^d, \mathbb{R}^d)$, for any $f \in \mathbf{H}^\mathbf{s}$ and $\varepsilon \in (0, 1)$, there are $n \in \mathbb{N}$, $\{u_j\} \subset H^{s_1}(U, \mathbb{R})$ and $\{v_j\} \subset H^{s_2}(\mathbb{R}^d, \mathbb{R}^d)$ such that

$$\left\| f - \sum_{j=1}^n u_j v_j \right\|_\mathbf{s} < \frac{\varepsilon}{2}.$$

Let

$$M = \left(\max_{1 \le j \le n} \|u_j\|_{s_1} + \max_{1 \le j \le n} \|v_j\|_{s_2} \right) \vee 1 < \infty,$$

where, with some abuse of notation, $\|\cdot\|_{s_1}$ is the norm on $H^{s_1}(U, \mathbb{R})$ and $\|\cdot\|_{s_2}$ is the norm on $H^{s_2}(\mathbb{R}^d, \mathbb{R}^d)$. Since $\mathcal{C}_c^\infty(U, \mathbb{R})$ is dense in $H^{s_1}(U, \mathbb{R})$ and $\mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R}^d)$ is dense in $H^{s_2}(\mathbb{R}^d, \mathbb{R}^d)$, for each j there are $g_j \in \mathcal{C}_c^\infty(U, \mathbb{R})$ and

$h_j \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R}^d)$ such that

$$\|u_j - g_j\|_{s_1} < \frac{\varepsilon}{6nM} \qquad \text{and} \qquad \|v_j - h_j\|_{s_2} < \frac{\varepsilon}{6nM}.$$

Then,

$$\begin{aligned}
\left\| f - \sum_{j=1}^n g_j h_j \right\|_{\mathbf{s}} &\leq \left\| f - \sum_{j=1}^n u_j v_j \right\|_{\mathbf{s}} + \left\| \sum_{j=1}^n u_j v_j - \sum_{j=1}^n g_j h_j \right\|_{\mathbf{s}} \\
&< \frac{\varepsilon}{2} + \sum_{j=1}^n \|u_j v_j - g_j h_j\|_{\mathbf{s}} \\
&\leq \frac{\varepsilon}{2} + \sum_{j=1}^n \|u_j\|_{s_1} \|v_j - h_j\|_{s_2} + \sum_{j=1}^n \|u_j - g_j\|_{s_1} \|h_j\|_{s_2} \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{6} + \frac{\varepsilon}{6nM} \sum_{i=1}^n (\|v_j - h_j\|_{s_2} + \|v_j\|_{s_2}) \\
&< \frac{2\varepsilon}{3} + \frac{\varepsilon}{6M} \left(\frac{\varepsilon}{6nM} + M \right) \\
&\leq \varepsilon.
\end{aligned}$$

Noting that $\sum_{j=1}^n g_j h_j \in \mathcal{C}_c^\infty(U \times \mathbb{R}^d, \mathbb{R}^d)$ completes the proof. □

The tensor product characterization of $\mathbf{H}^{\mathbf{s}}$ when $\mathbf{s} \in \mathbb{R}_+^2$ also allows for an alternate, more general proof of the compact embedding result of Lemma 5.3.4, which uses the explicit form of the norm on $\mathbf{H}^{\mathbf{s}}$. Rellich's theorem [45, Theorem 9.22] says that $H^{s_1}(U, \mathbb{R})$ is compactly embedded in $H^r(U, \mathbb{R})$ when $r < s_1$ and $H^{s_2}(\mathbb{R}^d, \mathbb{R}^d)$ is compactly embedded in $H^r(\mathbb{R}^d, \mathbb{R}^d)$ when $r < s_2$. Hence, the following theorem proves the compact embedding $\mathbf{H}^{\mathbf{s}} \subset \mathbf{H}^{\mathbf{r}}$ whenever $s_1 > r_1 > 0$ and $s_2 > r_2 > 0$.

Theorem A.3.3. *Let $H_1 \subset K_1$ and $H_2 \subset K_2$ be continuous embeddings of Hilbert spaces, and let $H_1 \otimes H_2$ and $K_1 \otimes K_2$ be their tensor product spaces. If every bounded sequence in H_1 has a convergent subsequence in K_1 and every bounded sequence in H_2 has a convergent subsequence in K_2, then every bounded sequence in $H_1 \otimes H_2$ has a convergent subsequence in $K_1 \otimes K_2$.*

This result follows from the following applied to the inclusion operators $\iota_1 : H_1 \to K_1$ and $\iota_2 : H_2 \to K_2$. A compact operator between two Banach spaces is a linear operator such that its image of any bounded set is relatively compact. For two Hilbert space linear, bounded operators $S : H_1 \to K_1$ and $T : H_2 \to K_2$, the tensor product operator $S \otimes T$ is the unique linear, bounded operator from $H_1 \otimes H_2$ into $K_1 \otimes K_2$ such that $S \otimes T(x \otimes y) = S(x) \otimes T(y)$ for all $x \in H_1$ and $y \in H_2$.

Lemma A.3.4. *Let H_1, H_2, K_1, and K_2 be Hilbert spaces, and let $H = H_1 \otimes H_2$ and $K = K_1 \otimes K_2$. If $S : H_1 \to K_1$ and $T : H_2 \to K_2$ are compact linear operators, then $S \otimes T : H \to K$ is compact.*

Proof. We assume that the Hilbert spaces are infinite dimensional, since any finite-rank operator is compact. We use the fact that operator-norm limits of finite-rank operators are compact, and also the fact that $R : A \to B$ for two infinite-dimensional Hilbert spaces A and B is compact if and only if it can be written in the form

$$R(x) = \sum_{i=1}^{\infty} a_i \langle e_i, x \rangle f_i, \qquad x \in A,$$

where $\{e_i\}$ is a complete orthonormal set in A and $\{f_i\}$ is a complete orthonormal set in B. Noting that $R(e_j) = a_j f_j$ for any j, we see that when $i \neq j$, $\langle R(e_i), R(e_j) \rangle = a_i a_j \langle f_i, f_j \rangle = 0$, and hence R preserves the orthogonality of $\{e_i\}$. Furthermore, the partial sums $R_N = \sum_{i=1}^{n} a_i \langle e_i, \cdot \rangle f_i$ converge to R in the operator norm. As usual, the operator norm is given by

$$\|R\|_{op} = \sup_{x \neq 0} \frac{\|R(x)\|_B}{\|x\|_A}.$$

Hence, there exist finite-rank linear operators $\{S_n\}$ and $\{T_n\}$ such that $S_n \to S$ and $T_n \to T$ as $n \to \infty$ in the operator norm, and there exists some $M < \infty$ such that

$$\|S\|_{op} + \|T\|_{op} + \sup_{n \geq 1} \|S_n\|_{op} + \sup_{n \geq 1} \|T_n\|_{op} \leq M.$$

It suffices to show that $S_n \otimes T_n \to S \otimes T$ in the operator norm.

Let $\varepsilon > 0$, and consider a nonzero pure tensor $x \otimes y \in H$. Since $S_n \to S$ and $T_n \to T$, for all n large enough we have

$$\|S_n - S\|_{op} \leq \frac{\varepsilon}{2M} \qquad \text{and} \qquad \|T_n - T\|_{op} \leq \frac{\varepsilon}{2M}.$$

Then,

$$\begin{aligned}
\|S_n \otimes T_n(x \otimes y) - S \otimes T(x \otimes y)\|_K &= \|S_n(x) \otimes T_n(y) - S(x) \otimes T(y)\|_K \\
&\leq \|S_n(x) - S(x)\|_{K_1} \|T_n(y)\|_{K_2} + \|S(x)\|_{K_1} \|T_n(y) - T(y)\|_{K_2} \\
&\leq \|S_n - S\|_{op} \|x\|_{H_1} \cdot \|T_n\|_{op} \|y\|_{H_2} + \|S\|_{op} \|x\|_{H_1} \cdot \|T_n - T\|_{op} \|y\|_{H_2} \\
&\leq \varepsilon \|x\|_{H_1} \|y\|_{H_2} \\
&= \varepsilon \|x \otimes y\|_H.
\end{aligned}$$

Now let $z \in H$ be arbitrary. Then we can write

$$z = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} e_i \otimes f_j,$$

where $\{e_i\}$ is a complete orthonormal system in H_1 and $\{f_j\}$ is a complete orthonormal system in H_2, so that $\{e_i \otimes f_j\}$ is a complete orthonormal system in H, and $\{a_{ij}\} \subset \mathbb{R}$ are unique. Note that Parseval's identity gives $\|z\|_H^2 = \sum_i \sum_j a_{ij}^2 < \infty$. Because of the discussion above, we may assume that $\{S(e_i)\}$ is orthogonal in K_1 and $\{T(f_j)\}$ is orthogonal in K_2, and hence $\{S \otimes T(e_i \otimes f_j)\} = \{S(e_i) \otimes T(f_j)\}$ is orthogonal in K. Similarly, $\{S_n \otimes T_n(e_i \otimes f_j)\}$ is orthogonal in K. By the above calculation for pure tensors, for all n large enough,

$$\|S_n \otimes T_n(e_i \otimes f_j) - S \otimes T(e_i \otimes f_j)\|_K \leq \varepsilon \|e_i \otimes f_j\|_H = \varepsilon \qquad \text{for all } i, j.$$

Then,

$$\begin{aligned} \|S_n \otimes T_n(z) - S \otimes T(z)\|_K^2 &= \left\| \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} \left(S_n \otimes T_n(e_i \otimes f_j) - S \otimes T(e_i \otimes f_j) \right) \right\|_K^2 \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij}^2 \|S_n \otimes T_n(e_i \otimes f_j) - S \otimes T(e_i \otimes e_j)\|_K^2 \\ &\leq \varepsilon^2 \|z\|_K^2. \end{aligned}$$

Since the choice of z was arbitrary, $\|S_n \otimes T_n - S \otimes T\|_{op} \leq \varepsilon$. □

Finally, from the following we get the compact embedding $\mathbf{H^s} \subset \mathbf{H^r}$ whenever $s_1 > r_1$ and $s_2 > r_2$ for any $\mathbf{s} \in \mathbb{R}^2$.

Lemma A.3.5. *Let $E_1 \subset E_2$ be Banach spaces with norms $\|\cdot\|_1$ and $\|\cdot\|_2$, respectively, such that the closed unit ball $\{x : \|x\|_1 \leq 1\}$ is compact in E_2. Let $E_2' \subset E_1'$ be their dual spaces equipped with the operator norms*

$$\|T\|_i' = \sup_{x \in E_i \setminus \{0\}} \frac{|T(x)|}{\|x\|_i} = \sup_{x \in E_i : \|x\|_i = 1} |T(x)|, \quad i = 1, 2.$$

Then, every bounded sequence in E_2' has a subsequence that converges in E_1'.

Proof. Let $\{T_k\} \subset E_2'$ such that $M = \sup_k \|T_k\|_2' < \infty$. For any $x, y \in E_2$, we have that

$$|T_k(x) - T_k(y)| \leq M \|x - y\|_2,$$

and hence $\{T_k\}$ is equicontinuous and pointwise bounded. The Arzela-Ascoli lemma then guarantees that there is a subsequence $\{T_{k_j}\}$ that converges uniformly on compact sets in E_2. Since $\{\|x\|_1 \leq 1\}$ is compact in E_2,

$$\|T_{k_j} - T_{k_m}\|_1' = \sup_{x \in E_1 : \|x\|_1 = 1} |T_{k_j}(x) - T_{k_m}(x)| \leq \sup_{x \in E_2 : \|x\|_1 \leq 1} |T_{k_j}(x) - T_{k_m}(x)| \to 0$$

as $k, m \to \infty$. Hence $\{T_{k_j}\}$ is Cauchy and thus convergent in the complete space E_1'. □

REFERENCES

[1] R. A. Adams and J. F. Fournier. *Sobolev Spaces*. Academic Press, Second edition, 2003.

[2] D. Aldous. Stopping times and tightness. *Ann. Probab.*, 6:335–340, 1978.

[3] D. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equation. *Ann. Appl. Probab.*, 15(2):1047–1110, 2005.

[4] G. Alsmeyer. On the markov renewal theorem. *Stochastic Process. Appl.*, 50:37–56, 1994.

[5] N. Antunes, C. Fricker, P. Robert, and D. Tibi. Stochastic networks with multiple stable points. *Ann. Probab.*, 36:255–278, 2008.

[6] V. Araman and P. Glynn. Tail asymptotics for the maximum of perturbed random walk. *Ann. Appl. Probab.*, 16:1411–1431, 2006.

[7] S. Asmussen. *Applied Probability and Queues*. Springer, New York, 2003.

[8] S. Asmussen and P. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Stochastic Modelling and Applied Probability. Springer New York, 2007.

[9] K. Athreya, D. McDonald, and P. Ney. Limit theorems for semi-Markov processes and renewal theory for Markov chains. *Ann. Probab.*, 6(5):788–797, 1978.

[10] J. Baladron, D. Fasoli, O. Faugeras, and J. Touboul. Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *J. Math. Neurosci.*, 2(1):10, 2012.

[11] J. Biggins. Martingale convergence in the branching random walk. *J. Appl. Prob.*, 14(1):25–37, 1977.

[12] J. Biggins. Lindley-type equations in the branching random walk. *Stochastic Process. Appl.*, 75:105–133, 1998.

[13] J. Blanchet and H. Lam. State-dependent importance sampling for rare event simulation: An overview and recent advances. *Surv. Oper. Res. Manag. Sci.*, 17:38–59, 2012.

[14] J. Blanchet, H. Lam, and B. Zwart. Efficient rare event simulation for perpetuities. *Stochastic Process. Appl.*, 122:3361–3392, 2012.

[15] C. Bordenave, D. Macdonald, and A. Proutiere. Performance of random medium access control, an asymptotic approach. In *Proc. ACM Sigmetrics*, pages 1–12, 2008.

[16] C. Bordenave, D. McDonald, and A. Proutiere. Random multi-access algorithms in networks with partial interaction: A mean field analysis. *Netw. Heterog. Media*, 5(1):31–62, 2010.

[17] M. Boué and P. Dupuis. A variational representation for certain functionals of Brownian motion. *Ann. Probab.*, 26:1641–1659, 1998.

[18] W. Braun and K. Hepp. The Vlasov dynamics and its fluctuations in the 1/N limit of interacting classical particles. *Comm. Math. Phys.*, 56(2):101–113, 1977.

[19] J. Bucklew. *Introduction to Rare-event Simulation*. Springer-Verlag, New York, 2004.

[20] A. Budhiraja and P. Dupuis. A variational representation for positive functionals of infinite dimensional Brownian motion. *Probab. Math. Statist.*, 20:39–61, 2000.

[21] A. Budhiraja and P. Dupuis. *Analysis and Approximation of Rare Events: Representations and Weak Convergence Methods*. Springer, New York, 2019.

[22] A. Budhiraja, P. Dupuis, and M. Fischer. Large deviation properties of weakly interacting processes via weak convergence methods. *Ann. Probab.*, 40:74–102, 2012.

[23] N. Chen, N. Litvak, and M. Olvera-Cravioto. Generalized PageRank on directed configuration networks. *Random Struct. Algor.*, 51(2):237–274, 2017.

[24] N. Chen and M. Olvera-Cravioto. Efficient simulation for branching linear recursions. *Proceedings of the 2015 Winter Simulation Conference*, pages 2716–2717, 2015.

[25] J. Cohen. *The Single Server Queue*. North-Holland Publishing Co., Amsterdam, 1982.

[26] J. Collamore. Importance sampling techniques for the multidimensional ruin problem for general markov additive sequences of random vectors. *Ann. Appl. Probab.*, 12:382–421, 2002.

[27] J. Collamore, G. Diao, and A. Vidyashankar. Rare event simulation for processes generated via stochastic fixed point equations. *Ann. Appl. Probab.*, 24:2143–2175, 2014.

[28] J. Collamore and A. Vidyashankar. Tail estimates for stochastic fixed point equations via nonlinear renewal theory. *Stochastic Process. Appl.*, 123:3378–3429, 2013.

[29] J. Collamore, A. Vidyashankar, and J. Xu. Rare event simulation for stochastic fixed point equations related to the smoothing transformation. *Proc. of the 2013 Winter Simulation Conference*, pages 555–563, 2013.

[30] J. Cvitanić, J. Ma, and J. Zhang. The law of large numbers for self-exciting correlated defaults. *Stochastic Process. Appl.*, 122(8):2781–2810, 2012.

[31] E. D. D. Buraczewski and T. Mikosch. *Stochastic Models with Power-Law Tails: The Equation* $X = AX + B$. Springer, 2016.

[32] D. A. Dawson. Critical dynamics and fluctuations for a mean-field model of cooperative behavior. *J. Stat. Phys.*, 31(1):29–85, 1983.

[33] D. A. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

[34] P. Del Moral. *Feynman-Kac formulae*. Probability and its Applications (New York). Springer-Verlag, New York, 2004. Genealogical and interacting particle systems with applications.

[35] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston, 1993.

[36] G. Dos Reis, W. Salkeld, J. Tugaut, et al. Freidlin–Wentzell LDP in path space for McKean–Vlasov equations and the functional iterated logarithm law. *Ann. Appl. Probab.*, 29(3):1487–1540, 2019.

[37] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, New York, 1997.

[38] P. Dupuis and K. Spiliopoulos. Large deviations for multiscale diffusions via weak convergence methods. *Stochastic Process. Appl.*, 122:1947–1987, 2012.

[39] G. P. E. Di Nezza and E. Valdinoci. Hitchhiker's guide to fractional Sobolev spaces. *Bull. Sci. Math.*, 2011.

[40] N. Enriquez, C. Sabot, and O. Zindy. A probabilistic representation of constants in Kesten's renewal theorem. *Probab. Theory Related Fields*, 144:581–613, 2009.

[41] S. Ethier and T. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.

[42] M. G. F. Flandoli, M. Gubinelli and V. M. Tortorelli. Stochastic currents. *Stocastic Process. Appl.*, 115(9):1583–1601, 2005.

[43] O. Faugeras, J. Touboul, and B. Cessac. A constructive mean-field analysis of multi population neural networks with random synaptic weights and stochastic inputs. *Front. Comput. Neurosc.*, 3:1–28, 2009.

[44] F. Flandoli and C. A. Tudor. Brownian and fractional Brownian stochastic currents via Malliavin calculus. *J. Funct. Anal.*, 258(9):279–306, 2010.

[45] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Wiley, 2 edition, 1999.

[46] J. Garnier, G. Papanicolaou, and T.-W. Yang. Large deviations for a mean field model of systemic risk. *SIAM J. Financ. Math.*, 4(1):151–184, 2013.

[47] K. Giesecke, K. Spiliopoulos, and R. B. Sowers. Default clustering in large portfolios: Typical events. *Ann. Appl. Probab.*, 23(1):348–385, 2013.

[48] P. Glynn and D. Iglehart. Importance sampling for stochastic simulations. *IEEE Trans. Automat. Control*, 35:1367–1392, 1989.

[49] C. Goldie. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.*, 1(1):126–166, 1991.

[50] J. Goméz-Serrano, C. Graham, and J.-Y. L. Boudec. The bounded confidence model of opinion dynamics. *Math. Models Methods Appl. Sci.*, 22(02):1150007, 2012.

[51] C. Graham and S. Méléard. Stochastic particle approximations for generalized Boltzmann models and convergence estimates. *Ann. Probab.*, 25(1):115–132, 1997.

[52] C. Graham and P. Robert. A multi-class mean-field model with graph structure for TCP flows. In *Progress in industrial mathematics at ECMI 2008*, pages 125–131. Springer, 2010.

[53] D. Grey. Regular variation in the tail behaviour of solutions of random difference equations. *Ann. Appl. Probab.*, 4:169–183, 1994.

[54] M. Gubinelli. Controlling rough paths. *J. Funct. Anal.*, 216(1):86–140, 2004.

[55] Y. Guivarc'h and E. L. Page. Spectral gap properties for linear random walks and Pareto's asymptotics for affine stochastic recursions. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52:503–574, 2016.

[56] S. Herrmann, P. Imkeller, and D. Peithmann. Large deviations and a Kramers' type law for self-stabilizing diffusions. *Ann. Appl. Probab.*, 18(4):1379–1423, 2008.

[57] S. Herrmann and J. Tugaut. Mean-filed limit versus small-noise limit for some interacting particle systems. *Commun. Stoch. Anal.*, 10(1):39–55, 2008.

[58] W. Hu, M. Salins, and K. Spiliopoulos. Large deviations and averaging for systems of slow-fast stochastic reaction-diffusion equations. *Stoch. PDE: Anal. Comp.*, 7:808–874.

[59] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second edition, 1989.

[60] A. Iksanov. Elementary fixed points of the BRW smoothing transforms with infinite number of summands. *Stochastic Process. Appl.*, 114:27–50, 2004.

[61] A. Iksanov. *Renewal Theory for Perturbed Random Walks and Similar Processes*. Springer, 2016.

[62] P. Jelenković and M. Olvera-Cravioto. Implicit renewal theory and power tails on trees. *Adv. Appl. Prob.*, 44(2):528–561, 2012.

[63] P. Jelenković and M. Olvera-Cravioto. Implicit renewal theory for trees with general weights. *Stochastic Process. Appl.*, 122(9):3209–3238, 2012.

[64] P. Jelenković and M. Olvera-Cravioto. Maximums on trees. *Stochastic Process. Appl.*, 125:217–232, 2015.

[65] P. R. Jelenković and M. Olvera-Cravioto. Convergence rates in the implicit renewal theorem on trees. *J. Appl. Probab.*, 50(4):1077–1088, 2013.

[66] M. Kac. Foundations of kinetic theory. In *Proceedings of the Berkeley SStatistics and Probability*, volume 3, pages 171–197, 1956.

[67] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 1988.

[68] F. Karpelevich, M. Kelbert, and Y. Suhov. Higher-order Lindley equations. *Stochastic Process. Appl.*, 53:65–96, 1994.

[69] A. Klenke. *Probability Theory: A Comprehensive Course*. Springer, 2 edition, 2013.

[70] T. Kurtz. *Approximation of Population Processes*, volume 36 of *CBMS-NSF Regional Conf. Series in Appl. Math.* SIAM, Philadelphia, 1981.

[71] T. G. Kurtz and J. Xiong. Particle representations for a class of nonlinear SPDEs. *Stochastic Process. Appl.*, 83(1):103–126, 1999.

[72] A. Kyprianou. Martingale convergence and the stopped branching random walk. *Probab. Theory Related Fields*, 116:405–419, 2000.

[73] P. B. L. Bertini and A. Pisante. Stochastic Allen-Cahn approximation of the mean curvature flow: large deviations upper bound. *Arch. Ration. Mec. Anal.*, 224(2):659–707, 2017.

[74] Q. Liu. Fixed points of a generalized smoothing transformation and applications to the branching random walk. *Adv. Appl. Prob.*, 30:85–112, 1998.

[75] Q. Liu. On generalized multiplicative cascades. *Stochastic Process. Appl.*, 86:263–286, 2000.

[76] R. Lyons. A simple path to Biggins' martingale convergence for branching random walk. In K. B. Athreya and P. Jagers, editors, *Classical and Modern Branching Processes*, volume 84 of *IMA Vol. Math. Appl.*, pages 217–222. Springer, New York, 1997.

[77] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of $l \log l$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23(3):1125–1138, 1995.

[78] T. J. Lyons. Differential equations driven by rough signals. *Reev. Mat. Iberoam*, 14(2):215–310, 1998.

[79] C. O. M. Fornasier, S. Lisini and G. Savaré. Mean-filed optimal control as Gamma-limit of finite agent controls. *Preprint arXiv:1803.04689*, 2018.

[80] G. M. M. Giaquinta and J. Souček. *Cartesian Currents in the Calculus of Variations I*. Springer, 1998.

[81] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour XXVII, 1997)*, pages 93–191. Springer-Verlag, Berlin, 1999.

[82] H. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA*, 56(6):1907, 1966.

[83] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 42–95. Springer, Berlin, 1996.

[84] M. Mezard and A. Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.

[85] K. Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.*, 12(2):458–479, 1984.

[86] M. Olvera-Cravioto et al. Convergence of the population dynamics algorithm in the Wasserstein metric. *Electron. J. Probab.*, 24(61):1–27, 2019.

[87] M. Olvera-Cravioto and O. Ruiz-Lacedelli. Stationary waiting time in parallel queues with synchronization. *Math. Oper. Res.*, to appear, 2020.

[88] C. Orrieri. Large deviations for interacting particle systems: joint mean-field and small-noise limit. *Electron. J. Probab.*, 25, 2020.

[89] Y.-X. Ren and R. Song. Spine decomposition for branching Markov processes and its applications. *Preprint arXiv:2007.12495*, 2020.

[90] U. Rösler. The weighted branching process. In *Dynamics of Complex and Irregular Systems (Bielefeld, 1991)*, Bielefeld Encount. Math. Phys. VIII, pages 154–165. World Sci. Publishing, River Edge, NJ, 1993.

[91] T. Shiga and H. Tanaka. Central limit theorem for a system of Markovian particles with mean field interactions. *Probab. Theory Related Fields*, 69(3):439–459, 1985.

[92] D. Siegmund. Importance sampling in the Monte Carlo study of sequential tests. *Ann. Stat.*, 3(1):673–684, 1976.

[93] X. Sun, R. Wang, L. Xu, and X. Yang. Large deviations for two-time-scale stochastic Burgers equation. *Preprint arXiv:1811.00290*, 2018.

[94] A.-S. Sznitman. Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. *J. Funct. Anal.*, 56(3):311–336, 1984.

[95] A.-S. Sznitman. *Topics in propagation of chaos*, volume 1464, pages 167–251. Springer, Berlin, 1991.

[96] H. Tanaka. Limit theorems for certain diffusion processes with interaction. In *Stochastic analysis (Katata/Kyoto, 1982)*, volume 32 of *North-Holland Math. Library*, pages 469–488. North-Holland, Amsterdam, 1984.

[97] H. Tanaka and M. Hitsuda. Central limit theorem for a simple diffusion model of interacting particles. *Hiroshima Mathematical Journal*, 11(2):415–423, 1981.

[98] S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, pages 261–286, 1966.

[99] C. Villani. *Optimal Transport, Old and New*. Springer, 2009.