FIELDS INSTITUTE MONOGRAPHS

THE FIELDS INSTITUTE FOR RESEARCH IN MATHEMATICAL SCIENCES

Large Deviations

Frank den Hollander

American Mathematical Society
Providence, Rhode Island

Selected Titles in This Series

14 **Frank den Hollander,** Large deviations, 2000

13 **B. V. Rajarama Bhat, George A. Elliott, and Peter A. Fillmore, Editors,** Lectures in operator theory, 2000

12 **Salma Kuhlmann,** Ordered exponential fields, 2000

11 **Tibor Krisztin, Hans-Otto Walther, and Jianhong Wu,** Shape, smoothness and invariant stratification of an attracting set for delayed monotone positive feedback, 1999

10 **Jiří Patera, Editor,** Quasicrystals and discrete geometry, 1998

9 **Paul Selick,** Introduction to homotopy theory, 1997

8 **Terry A. Loring,** Lifting solutions to perturbing problems in C^*-algebras, 1997

7 **S. O. Kochman,** Bordism, stable homotopy and Adams spectral sequences, 1996

6 **Kenneth R. Davidson,** C*-Algebras by example, 1996

5 **A. Weiss,** Multiplicative Galois module structure, 1996

4 **Gérard Besson, Joachim Lohkamp, Pierre Pansu, and Peter Petersen Miroslav Lovric, Maung Min-Oo, and McKenzie Y.-K. Wang, Editors,** Riemannian geometry, 1996

3 **Albrecht Böttcher, Aad Dijksma and Heinz Langer, Michael A. Dritschel and James Rovnyak, and M. A. Kaashoek Peter Lancaster, Editor,** Lectures on operator theory and its applications, 1996

2 **Victor P. Snaith,** Galois module structure, 1994

1 **Stephen Wiggins,** Global dynamics, phase space transport, orbits homoclinic to resonances, and applications, 1993

FIELDS INSTITUTE MONOGRAPHS

THE FIELDS INSTITUTE FOR RESEARCH IN MATHEMATICAL SCIENCES

Large Deviations

Frank den Hollander

American Mathematical Society
Providence, Rhode Island

The Fields Institute for Research in Mathematical Sciences

The Fields Institute is named in honour of the Canadian mathematician John Charles Fields (1863–1932). Fields was a visionary who received many honours for his scientific work, including election to the Royal Society of Canada in 1909 and to the Royal Society of London in 1913. Among other accomplishments in the service of the international mathematics community, Fields was responsible for establishing the world's most prestigious prize for mathematics research—the Fields Medal.

The Fields Institute for Research in Mathematical Sciences is supported by grants from the Ontario Ministry of Education and Training and the Natural Sciences and Engineering Research Council of Canada. The Institute is sponsored by McMaster University, the University of Toronto, the University of Waterloo, and York University and has affiliated universities in Ontario and across Canada.

2000 *Mathematics Subject Classification.* Primary 60-01, 60F10, 60K35; Secondary 82B31, 82B44.

ABSTRACT. This book is an introduction to large deviations and consists of two parts: Part A theory, Part B applications.

Part A describes the basic large deviation theorems for i.i.d. sequences, Markov sequences, and sequences with moderate dependence. It also gives an outline of definitions and theorems in a more abstract context, exposing the unified scheme that gives large deviation theory its overall structure.

Part B describes a selection of applications, most of which are recent and circle around statistical physics and random media.

The book contains some 60 exercises with solutions.

Library of Congress Cataloging-in-Publication Data

Hollander, F. den (Frank)
Large deviations / Frank den Hollander.
p. cm. — (Fields Institute monographs, ISSN 1069-5273 ; 14)
Includes bibliographical references and index.
ISBN 0-8218-1989-5 (alk. paper)
1. Large deviations. I. Title. II. Series.
QA273.67.H65 2000
519.5′34–dc21 99-058913

The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
This publication was prepared by The Fields Institute.
Visit the AMS home page at URL: http://www.ams.org/

10 9 8 7 6 5 4 3 2 1 05 04 03 02 01 00

Contents

PREFACE

Large deviation theory is a part of probability theory that deals with the description of events where a sum of random variables deviates from its mean by more than a "normal" amount, i.e., beyond what is described by the central limit theorem. A precise calculation of the probabilities of such events turns out to be crucial for the study of integrals of exponential functionals of sums of random variables, which come up in a variety of different contexts. Large deviation theory finds application in probability theory, statistics, operations research, ergodic theory, information theory, statistical physics, financial mathematics, and the list goes on.

These lecture notes are an *introduction* to large deviations. Part A (Chapters I–V) describes *theory*, Part B (Chapters VI–X) describes *applications*. A glance at the table of contents shows what topics are covered. I have put much effort into conveying the main ideas without putting too much emphasis on technicalities. Most of the theory is driven by a few "key principles" and once these are understood the rest of the journey is safe sailing, give or take a storm or two. This is not to say that it is easy to grasp the full theoretical panorama. But the reader's patience will be rewarded when the ship enters the harbor of the applications.

Chapters I and II contain the basic large deviation theorems for i.i.d. random variables. Here the goal is to make the reader acquainted with the type of statements that are typical for the theory and to obtain results via explicit calculation of the rate function. Chapter III presents general definitions and theorems in a more abstract context. Here the goal is to expose the unified scheme that gives large deviation theory its overall structure and that can be made to work in many concrete cases. Chapter IV looks at large deviation theorems for Markov chains and explains how these can be obtained from the i.i.d. case via a change-of-measure argument. Chapter V considers random sequences with moderate dependence and shows how many of the results in Chapters I, II and IV can be put under a single heading, which in some sense closes the circle.

Chapters VI–X describe a selection of applications: statistical hypothesis testing; random walk in random environment; heat conduction with random sources and sinks; polymer chains; interacting diffusions. Here the theory comes to life and the reader gets to see the full impact of the results derived earlier. Except for the application in statistical hypothesis testing, which is put in mainly for didactical reasons, all applications are recent. Naturally their choice reflects my personal taste and involvement, since they circle around statistical physics and random media. But I think they offer a good sample of what large deviation theory is able to do in various different contexts. Each application is self-contained and tells a small story.

Many questions that come up during the exposition are posed as exercises to the reader. The solutions to these exercises are given in the Appendix. At the end I have included a list of frequently used words and symbols, with the number of the section where they appear first. This will help the reader to connect the different chapters.

Large deviation theory is a mixture of probability theory, convex analysis, variational calculus and set topology. As such it is mathematically both challenging and captivating. Even so, it is not obvious how to do justice to a vast area like large deviation theory in one hundred pages or so, especially when the goal is to cover both theory and applications. To focus ideas, I have restricted most of the exposition to random sequences, i.e., discrete-time random processes. This is a severe restriction indeed, but it makes the presentation much more user-friendly. The reader can expand his or her skills by turning to the monographs that are listed as references. Here a wealth of refinements and embellishments can be found, as well as beautiful and deep large deviation results for Brownian motion, random dynamical systems, Gibbs measures, interacting particle systems, Brownian motion among random obstacles, etc. These monographs also contain an extensive historical overview of the area.

The material presented here was taught as a graduate course at the Fields Institute for Research in Mathematical Sciences in Toronto, Canada, in the Fall of 1998, as part of the 1998-99 program on "Probability and its Applications". I am grateful to the staff of the Fields Institute for the hospitality I enjoyed as a visitor. I am grateful to the following colleagues for comments during the course: Rami Atar, Siva Athreya, Marek Biskup, Jürgen Gärtner, Takashi Hara, Remco van der Hofstad, Min Kang, Neal Madras, Anders Martin-Löf, George O'Brien, David Rolls, Tom Salisbury, Gordon Slade, Dean Slonowsky, Jan Swart and Stas Volkov.

I am indebted to Siva Athreya, Nina Gantert, Jürgen Gärtner, Andreas Greven, Remco van der Hofstad, Wolfgang König and George O'Brien for reading a draft of the manuscript and for suggesting improvements.

Special thanks go to Marek Biskup, who helped me to prepare the manuscript, both in terms of content and exposition, and who worked hard on the layout and on the figures. Marek was a constant companion in the enterprise of bringing the job to a good end.

Frank den Hollander
Nijmegen, October 1999

Part A

THEORY

CHAPTER I

LARGE DEVIATIONS FOR I.I.D. SEQUENCES: PART 1

Chapters I and II are devoted to large deviation theorems for i.i.d. real-valued random variables. The reader will get acquainted with the basic results via explicit computation.

I.1 Introduction

We begin our journey on familiar territory. Let $X_1, X_2, \ldots$ be i.i.d. random variables on a probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$, where $\mathcal{B}(\mathbb{R})$ is the Borel sigma-field on $\mathbb{R}$. Write $\mathbb{E}$ to denote expectation under $\mathbb{P}$, let

$$\begin{aligned} \mathbb{E}\, X_1 &= \mu \in \mathbb{R}, \\ \operatorname{Var} X_1 &= \sigma^2 \in (0, \infty), \end{aligned}$$

and let $S_n = X_1 + \cdots + X_n$ $(n \in \mathbb{N})$ be the partial sums. In standard textbooks on probability theory two fundamental theorems dealing with such sequences can be found:

Strong Law of Large Numbers (SLLN)

$$\frac{1}{n} S_n \underset{n\to\infty}{\longrightarrow} \mu \qquad \mathbb{P}\text{–a.s.}$$

Central Limit Theorem (CLT)

$$\frac{1}{\sigma\sqrt{n}}(S_n - \mu n) \underset{n\to\infty}{\longrightarrow} Z \qquad \text{in law w.r.t. } \mathbb{P},$$

where Z is a standard normal random variable.

While the SLLN asserts that the empirical average $\frac{1}{n}S_n$ converges to μ as $n \to \infty$, the CLT quantifies the probability that S_n differs from μn by an amount of order $\sqrt{n}$. Deviations of this size are called "normal".

In these lecture notes we deal with events where S_n differs from μn by an amount of order n, so well beyond what is described by the CLT. Deviations of this size are called "large". An example is the event

$$\{S_n \geq (\mu + a)n\}, \qquad a > 0,$$

whose probability tends to zero as $n \to \infty$. It is our task to quantify the *rate* at which this occurs. We will see that, under a certain condition on the tail of the distribution of X_1, the decay is exponential in n:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(S_n \geq (\mu + a)\, n\Big) = -I(a) < 0, \qquad a > 0.$$

This is because, typically, a positive fraction of the components $X_1, \ldots, X_n$ in the sum S_n must deviate from μ to create the large deviation an away from μn. It will turn out that knowledge of the rate function $a \mapsto I(a)$ is crucial for a correct evaluation of integrals of exponential functionals of S_n as $n \to \infty$.

As we go along, we also consider other functionals of $X_1, \dots, X_n$, such as the empirical measure $L_n = \frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n})$, with δ_x the point-measure at $x \in \mathbb{R}$. It turns out that L_n has a large deviation behavior similar to that of $\frac{1}{n}S_n$ with its own rate function. We will go even further and look at large deviations for the empirical measure of words of length two and larger, thus obtaining results at ever higher levels of detail. Our goal will be to describe a theory for handling such large deviation questions in general.

Before we start we make an observation. Given two sequences of positive numbers (α_n) and (β_n), we write

$$\alpha_n \simeq \beta_n \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \frac{1}{n}(\log \alpha_n - \log \beta_n) = 0, \tag{I.1}$$

i.e., the symbol $\simeq$ means that the two sequences are logarithmically equivalent. The following elementary fact plays an important role throughout the exposition:

$$\alpha_n + \beta_n \simeq \alpha_n \vee \beta_n. \tag{I.2}$$

The reader should think of (I.2) as a "largest-exponent-wins" principle. Obviously, (I.2) can be iterated to apply to finitely many sequences.

I.2 An example: coin tossing

We begin with an example that serves as a warm-up.

THEOREM I.3 *Let (X_i) be i.i.d. random variables with $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$. Let $S_n = \sum_{i=1}^n X_i$. Then, for all $a > \frac{1}{2}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -I(a),$$

where

$$I(z) = \begin{cases} \log 2 + z\log z + (1-z)\log(1-z) & \text{if } z \in [0,1], \\ \infty & \text{otherwise.} \end{cases}$$

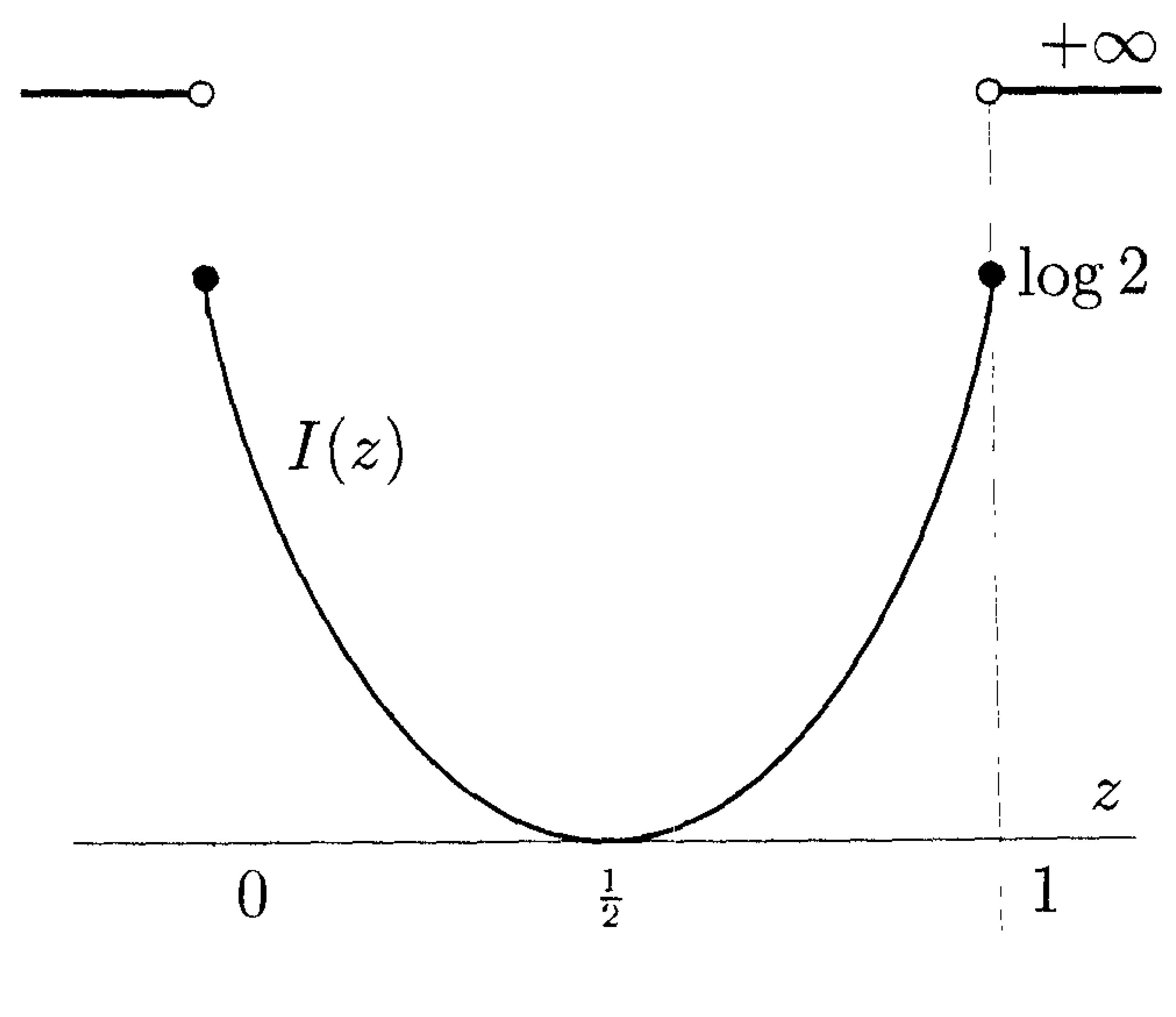

FIG. 1. The rate function for coin tossing

PROOF. The claim is trivial for $a > 1$. For $a \in (\frac{1}{2}, 1]$ we observe that $\mathbb{P}(S_n \geq an) = 2^{-n} \sum_{k \geq an} \binom{n}{k}$, which yields the estimate

$$2^{-n} Q_n(a) \leq \mathbb{P}(S_n \geq an) \leq (n+1) 2^{-n} Q_n(a),$$

where

$$Q_n(a) = \max_{k \geq an} \binom{n}{k}.$$

The maximum is attained at $k = \lceil an \rceil$, the smallest integer $\geq an$. Stirling's formula $n! = n^n e^{-n} \sqrt{2\pi n}\,(1 + O(\frac{1}{n}))$ therefore allows us to infer that

$$\lim_{n \to \infty} \frac{1}{n} \log Q_n(a) = -a \log a - (1-a) \log(1-a).$$

Noting that the upper and the lower bound merge on an exponential scale as $n \to \infty$, we arrive at the desired statement. □

Since $\mathbb{E} X_1 = \frac{1}{2}$ and $a > \frac{1}{2}$, Theorem I.3 deals with large deviations in the upward direction. It is clear from symmetry that the same holds for $\mathbb{P}(S_n \leq an)$ with $a < \frac{1}{2}$. This is manifested by the symmetry relation $I(1-z) = I(z)$.

The function $z \mapsto I(z)$ is called the *rate function*. Note that it is infinite outside $[0, 1]$, finite and strictly convex inside $[0, 1]$, and has a unique zero at $z = \frac{1}{2}$ (see Fig. 1). This zero corresponds to the SLLN. Indeed, Theorem I.3 implies that

$$\sum_{n \in \mathbb{N}} \mathbb{P}\big(|\tfrac{1}{n} S_n - \tfrac{1}{2}| > \delta\big) < \infty \qquad \forall \delta > 0,$$

and so the SLLN follows via the Borel-Cantelli lemma. The curvature of the rate function at $z = \frac{1}{2}$ corresponds to the CLT, as will become clear later.

I.3 Cramér's Theorem for the empirical average

We now formulate the first basic result of large deviation theory, which goes back to Cramér [D12]. This result, which generalizes Theorem I.3, identifies the large deviation behavior of the *empirical average* $\frac{1}{n} S_n$ under a certain condition on the tail of the distribution of X_1.

THEOREM I.4 *Let (X_i) be i.i.d. $\mathbb{R}$-valued random variables satisfying*

$$\varphi(t) = \mathbb{E}\, e^{tX_1} < \infty \qquad \forall t \in \mathbb{R}. \tag{I.5}$$

Let $S_n = \sum_{i=1}^n X_i$. Then, for all $a > \mathbb{E} X_1$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq an) = -I(a),$$

where

$$I(z) = \sup_{t \in \mathbb{R}} [zt - \log \varphi(t)]. \tag{I.6}$$

PROOF. We may suppose without loss of generality that $a = 0$ and $\mathbb{E} X_1 < 0$. Namely, the substitution $X_1 \to X_1 + a$ gives $\varphi(t) \to e^{at} \varphi(t)$ and, consequently, $I(a) \to I(0)$. We may also suppose that X_1 is non-degenerate, because the claim is trivial otherwise.

EXERCISE I.7 *Assume $\mathbb{P}(X_1 = a) = 1$. Check that $I(a) = 0$ and $I(z) = \infty$ for $z \neq a$.*

Henceforth we abbreviate

$$\rho = \inf_{t\in\mathbb{R}} \varphi(t).$$

Note that $I(0) = -\log\rho$ (with $I(0) = \infty$ if $\rho = 0$), so we must prove that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) = \log\rho.$$

Let $F(x) = \mathbb{P}(X_1 \leq x)$ be the distribution function of X_1. It follows from (I.5) that $\varphi \in C^\infty(\mathbb{R})$, the space of smooth functions, with

$$\begin{aligned} \varphi'(t) &= \int_{\mathbb{R}} x e^{tx} dF(x), \\ \varphi''(t) &= \int_{\mathbb{R}} x^2 e^{tx} dF(x). \end{aligned}$$

Hence, φ is strictly convex and $\varphi'(0) = \mathbb{E}\, X_1 < 0$. According to where the mass of $\mathbb{P}$ is situated, we may distinguish three subcases:

(i) $\mathbb{P}(X_1 < 0) = 1$.
Then φ is strictly decreasing and $\lim_{t\to\infty}\varphi(t) = \rho = 0$. Since $\mathbb{P}(S_n \geq 0) = 0$, the claim follows.

(ii) $\mathbb{P}(X_1 \leq 0) = 1$ and $\mathbb{P}(X_1 = 0) > 0$.
Then φ is strictly decreasing and $\lim_{t\to\infty}\varphi(t) = \rho = \mathbb{P}(X_1 = 0) > 0$. Since $\mathbb{P}(S_n \geq 0) = \mathbb{P}(X_1 = \cdots = X_n = 0) = \rho^n$, the claim follows.

(iii) $\mathbb{P}(X_1 < 0) > 0$ and $\mathbb{P}(X_1 > 0) > 0$.
Then $\lim_{t\to\pm\infty}\varphi(t) = \infty$, and since φ is strictly convex there exists a unique $\tau \in \mathbb{R}$, satisfying $\tau > 0$, such that (see Fig. 2)

$$\varphi(\tau) = \rho, \quad \varphi'(\tau) = 0.$$

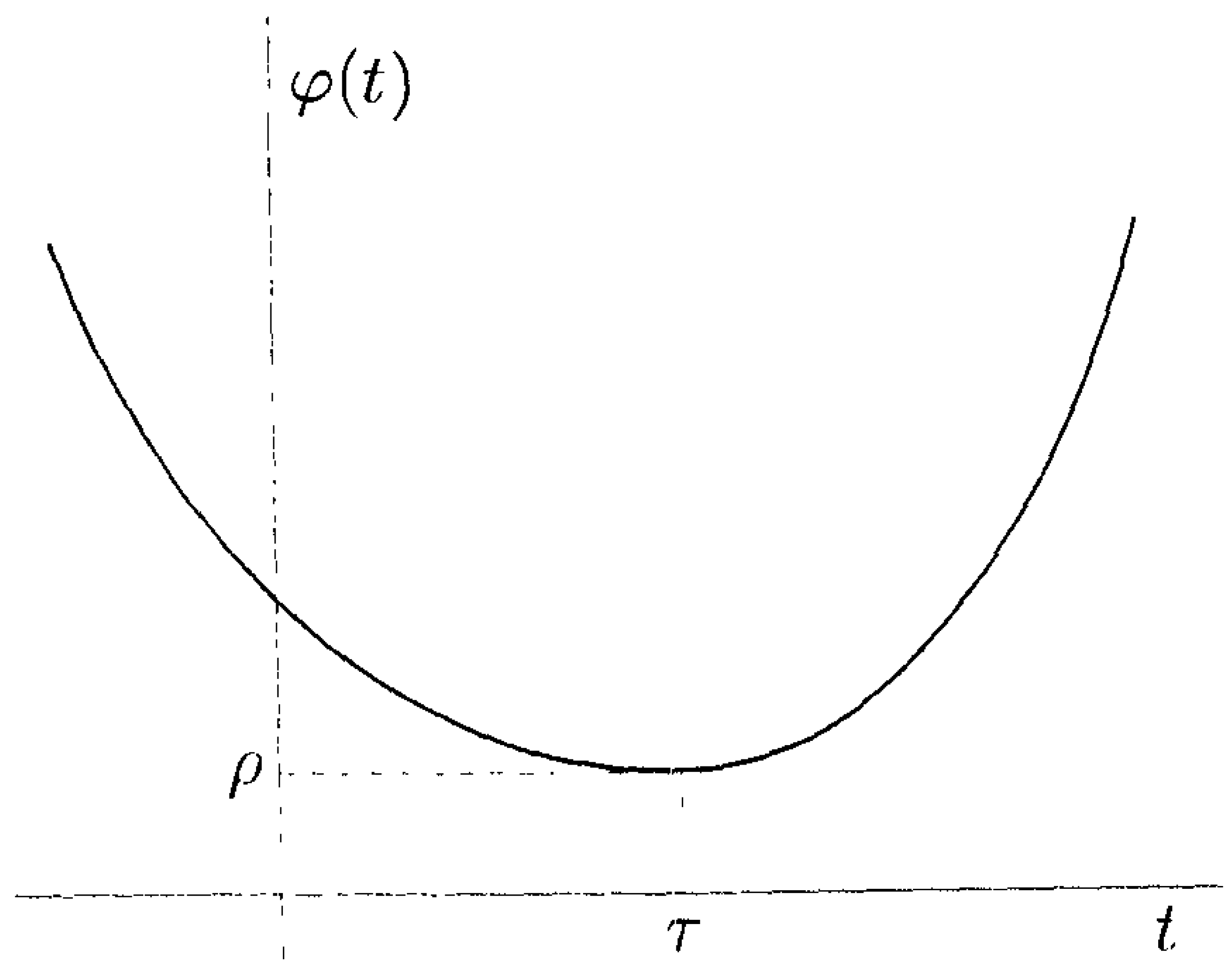

FIG. 2. Relation between φ, τ and ρ

We proceed with the analysis of (iii). By the exponential Chebyshev inequality, we have the following upper bound:

$$\mathbb{P}(S_n \geq 0) = \mathbb{P}(e^{\tau S_n} \geq 1) \leq \mathbb{E}\, e^{\tau S_n} = [\varphi(\tau)]^n = \rho^n.$$

Hence

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \leq \log\rho.$$

To get the lower bound we resort to a technique that will find an interesting generalization later on. Let $(\widehat{X}_i)$ be an i.i.d. sequence of random variables with distribution function

$$\widehat{F}(x) = \frac{1}{\rho}\int_{(-\infty,x]} e^{\tau y} dF(y).$$

The function $\widehat{F}$ is called the *Cramér-transform* of F. Note that $\rho = \varphi(\tau) = \int_{\mathbb{R}} e^{\tau y} dF(y)$. The proof proceeds via three lemmas:

LEMMA I.8 $\mathbb{E}\,\widehat{X}_1 = \widehat{\mu} = 0$ *and* $\operatorname{Var}\widehat{X}_1 = \widehat{\sigma}^2 \in (0,\infty)$.

PROOF. Let $\widehat{\varphi}(t) = \mathbb{E}\,e^{t\widehat{X}_1}$. Then

$$\widehat{\varphi}(t) = \int_{\mathbb{R}} e^{tx} d\widehat{F}(x) = \frac{1}{\rho}\int_{\mathbb{R}} e^{tx}e^{\tau x} dF(x) = \frac{1}{\rho}\varphi(t+\tau) < \infty \qquad \forall t \in \mathbb{R}.$$

This implies that $\widehat{\varphi} \in C^\infty(\mathbb{R})$ and

$$\mathbb{E}\,\widehat{X}_1 = \widehat{\varphi}'(0) = \tfrac{1}{\rho}\varphi'(\tau) = 0,$$
$$\operatorname{Var}\widehat{X}_1 = \widehat{\varphi}''(0) = \tfrac{1}{\rho}\varphi''(\tau) \in (0,\infty).$$

□

LEMMA I.9 *Let* $\widehat{S}_n = \sum_{i=1}^n \widehat{X}_i$. *Then* $\mathbb{P}(S_n \geq 0) = \rho^n \mathbb{E}\big(e^{-\tau\widehat{S}_n} 1_{\{\widehat{S}_n \geq 0\}}\big)$.

PROOF. Write

$$\begin{aligned}\mathbb{P}(S_n \geq 0) &= \int_{\{x_1+\cdots+x_n \geq 0\}} dF(x_1)\ldots dF(x_n)\\ &= \int_{\{x_1+\cdots+x_n \geq 0\}} [\rho e^{-\tau x_1} d\widehat{F}(x_1)]\ldots[\rho e^{-\tau x_n} d\widehat{F}(x_n)],\end{aligned}$$

which gives the expression. □

LEMMA I.10 $\liminf_{n\to\infty} \frac{1}{n}\log\mathbb{E}\left(e^{-\tau\widehat{S}_n} 1_{\{\widehat{S}_n \geq 0\}}\right) \geq 0$.

PROOF. First we observe that, by Lemma I.8, the CLT can be applied to $\widehat{S}_n$. Pick any number $C > 0$ such that $\frac{1}{\sqrt{2\pi}}\int_0^C e^{-x^2/2}dx > \frac{1}{4}$ and estimate

$$\mathbb{E}\left(e^{-\tau\widehat{S}_n} 1_{\{\widehat{S}_n \geq 0\}}\right) \geq e^{-\tau C\widehat{\sigma}\sqrt{n}}\,\mathbb{P}\Big(\frac{\widehat{S}_n}{\widehat{\sigma}\sqrt{n}} \in [0,C)\Big).$$

The probability in the RHS is at least $\frac{1}{4}$ for n large enough. □

The proof of Theorem I.4 is now concluded by observing that Lemmas I.9 and I.10 provide the desired lower bound for case (iii):

$$\liminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}(S_n \geq 0) \geq \log\rho.$$

□

The following exercises identify the rate function I in (I.6) for a number of choices for the distribution of X_1.

EXERCISE I.11 *Compute I for the following distributions: Poisson, Exponential and Normal.*

EXERCISE I.12 *Deduce Theorem I.3 from Theorem I.4.*

EXERCISE I.13 *Compute I when $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 2) = \mathbb{P}(X_1 = 3) = \frac{1}{3}$. Note that the resulting formula is not as nice as for coin tossing.*

The rate function I in Theorem I.4 has the following properties:

LEMMA I.14 *Assume* (I.5). *Then:*
(i) *I is lower semi-continuous and convex on $\mathbb{R}$.*
(ii) *I has compact level sets.*
(iii) *I is continuous and strictly convex on $\mathrm{int}(\mathcal{D}_I)$, where $\mathcal{D}_I = \{z \in \mathbb{R}\colon I(z) < \infty\}$ and $\mathrm{int}(\mathcal{D}_I)$ is the interior of $\mathcal{D}_I$.*
(iv) *I is smooth on $\mathrm{int}(\mathcal{D}_I)$.*
(v) *$I(z) \geq 0$ with equality if and only if $z = \mu$.*
(vi) *$I''(\mu) = \frac{1}{\sigma^2}$.*

REMARKS I.15
1. The level sets of I are the sets $I^{-1}([0,c]) = \{z \in \mathbb{R}\colon I(z) \leq c\}$ with $c \in [0,\infty)$.
2. Lower semi-continuity is equivalent to the level sets being closed.
3. The convexity of I implies that $\mathcal{D}_I$ is an interval (possibly infinite).

EXERCISE I.16 *Prove Lemma I.14.*

Thus, the typical picture of I is qualitatively similar as in Fig. 1, except that there are various possibilities for $\mathcal{D}_I$ and for the behavior of I near $\partial\mathcal{D}_I$, the boundary of $\mathcal{D}_I$ (see Exercise I.11).

I.4 Comments

Let us have a closer look at what has happened in Section I.3.

(1) The same statement as in Theorem I.4 holds for $\mathbb{P}(S_n \leq an)$ and $a < \mathbb{E}X_1$, with the same formula for I. This is easily checked via the mirror reflection $X_1 \to -X_1$.

(2) Note that the key idea in the proof of Theorem I.4 is the "exponential tilting" of the probability measure (through the Cramér transform) in combination with the CLT. Under the tilted probability measure the large deviation event $\{S_n \geq 0\}$ becomes "typical", as is seen from Lemmas I.8–I.10. This type of argument is very powerful and will recur in later chapters in various guises. In fact, it will play the role of a guiding principle, which will become clear as we go along.

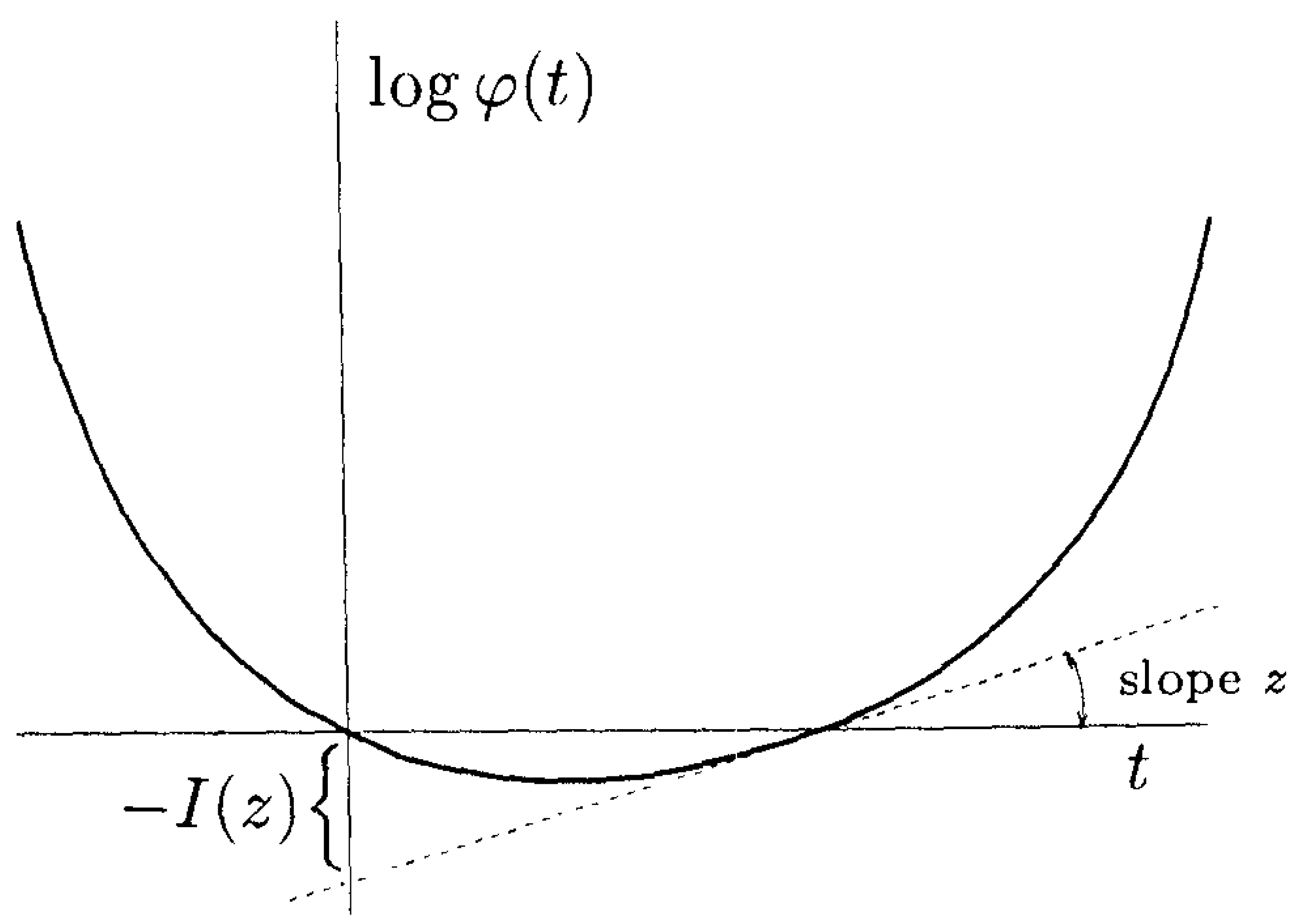

FIG. 3. Geometrical interpretation of the Legendre transform

(3) Equation (I.6) says that the rate function I is the *Legendre transform* of the cumulant generating function $\log\varphi$ (see Fig. 3). This important relationship will turn up later in various different guises. See, in particular, Chapter V.

(4) The requirement (I.5) in Theorem I.4 can be weakened to the condition that

$$0 \in \text{int}(\mathcal{D}_\varphi) \quad \text{with} \quad \mathcal{D}_\varphi = \{t \in \mathbb{R}\colon \varphi(t) < \infty\}. \tag{I.17}$$

We will not prove this generalization (see e.g. Dembo and Zeitouni [A2] Section 2.2.1). However, the reader can easily check that the proof in Section I.3 (and its mirror version for the large deviations in the downward direction) carries over when, in addition to (I.17), we have

$$\lim_{t\to\partial\mathcal{D}_\varphi\colon t\in\mathcal{D}_\varphi} \big|[\log\varphi]'(t)\big| = \infty, \tag{I.18}$$

a condition referred to as "$\log\varphi$ is steep at $\partial\mathcal{D}_\varphi$". Indeed, this property implies that $t \mapsto e^{at}\varphi(t)$ achieves a minimum in $\text{int}(\mathcal{D}_\varphi)$ for every $a \in \mathbb{R}$, which is needed to handle case (iii) after the substitution $X_1 \to X_1 + a$. See also Section V.2.

EXERCISE I.19 *Give examples for which* $\log\varphi$ *is: (i) steep; (ii) not steep.*

EXERCISE I.20 *Show that steepness of* $\log\varphi$ *implies* $\mathcal{D}_I = \mathbb{R}$.

(5) Theorem I.4 is in fact true even without (I.17), as can be shown with the help of a truncation argument (Bahadur [B1]). However, without (I.17) it looses much of its power. For instance, if $\mathcal{D}_\varphi = \{0\}$, then $I \equiv 0$. But in that case all Theorem I.4 tells us is that the large deviation probabilities decay slower than exponential. It gives us no information on how slow the decay actually is. Similarly, if $\mathcal{D}_\varphi$ contains a left-neighborhood of 0 but not a right-neighborhood, then $I(x) = 0$ for all $x \geq 0$, in which case Theorem I.4 gives interesting information only for large deviations that go downwards. Vinogradov [A9] gives a detailed analysis of all the subexponential cases. The decay in these cases depends heavily on the law of X_1 and in general is not captured by as elegant a formula as the Legendre transform in Theorem I.4.

(6) The following needs no condition at all.

EXERCISE I.21 *Use Hölder's inequality and Fatou's lemma to show that* $\log\varphi$ *is convex and lower-semicontinuous on* $\mathbb{R}$.

Because of these two properties we have the inverse formula (see Rockafellar [C7] Theorem 12.2)

$$\log\varphi(t) = \sup_{z\in\mathbb{R}}[tz - I(z)], \tag{I.22}$$

which is the dual of the formula in (I.6). This fact shows that φ and I are in a one-to-one relation. As soon as (I.17) holds, the law of X_1 is uniquely determined by φ. Thus, each law has its own unique rate function.

(7) Lemma I.14(v) is, of course, related to the SLLN (see the end of Section I.2). Lemma I.14(vi), on the other hand, is related to the CLT. Indeed, replacing a by $a_n = \mu + \frac{b\sigma}{\sqrt{n}}$ ($b > 0$) and pretending that the result in Theorem I.4 is still applicable, we find

$$\mathbb{P}\Big(\frac{S_n - \mu n}{\sigma\sqrt{n}} \in db\Big) = e^{-nI(\mu+\frac{b\sigma}{\sqrt{n}}) + \text{c.t.}}db = e^{-b^2/2 + \text{c.t.}}db \qquad n \to \infty,$$

which is the standard normal density (modulo the normalization, which is hidden in the correction term). This argument can be made more precise by appealing to Lemmas I.8–I.10.

(8) Bahadur and Ranga Rao [D2] prove that if (I.17) holds and the law of X_1 is non-lattice (i.e., is not concentrated on a periodic grid), then the following refined estimate holds:

$$\mathbb{P}(S_n \geq an) = \frac{1}{\sqrt{2\pi\widehat{\sigma}(a)^2 n}}\, e^{-nI(a)}\,(1+o(1)), \qquad a \in \operatorname{int}(\mathcal{D}_I),\ a > \mu,$$

with $\widehat{\sigma}(a)^2 = \frac{[I'(a)]^2}{I''(a)}$, and similarly for $a \in \operatorname{int}(\mathcal{D}_I)$, $a < \mu$. Again, this behavior can be traced back to Lemmas I.8–I.10. In these lecture notes we will not address such "higher-order" results and focus on the leading exponential term only.

(9) By Lemma I.14(i,v), if $a > \mu$, then $I(z) \geq I(a)$ for all $z \geq a$. Hence the result of Theorem I.4 can be rewritten as the statement

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n} S_n \in A\Big) = -\inf_{z\in A} I(z) \qquad \text{with } A = [a,\infty). \tag{I.23}$$

This formula should be interpreted as saying that the large deviation $\{\frac{1}{n}S_n \in A\}$ is essentially carried by the event where $\frac{1}{n}S_n$ is close to $\bar{z}$, with $\bar{z}$ the minimizer of $I(z)$ on A (which happens to be $\bar{z} = a$ when $a > \mu$). Namely, the latter event costs $\exp[-nI(\bar{z}) + o(n)]$ and therefore is the cheapest realization of A. This fact illustrates a key principle in large deviation theory:

ANY LARGE DEVIATION IS DONE IN THE LEAST UNLIKELY OF ALL THE UNLIKELY WAYS!

Equation (I.23) of course suggests that the same statement holds when $A = [a,\infty)$ is replaced by a sufficiently nice subset of $\mathbb{R}$. This is indeed the case, but we will postpone such topological considerations to Chapter III.

EXERCISE I.24 *Show that* (I.23) *holds: (i) for all* $A = [a,\infty)$ *with* $a \in \mathbb{R}$*; (ii) for all* $A = (a,\infty)$ *with* $a \in \mathbb{R} \setminus \partial\mathcal{D}_I$*; (iii) for all* $A = (a,b)$ *with* $a,b \in \mathbb{R} \setminus \partial\mathcal{D}_I$.

I.5 A toy application of Cramér

Let us demonstrate how Theorem I.4 can be used to study integrals of exponential functionals of S_n. For that we cook up the following artificial example. A more serious application will be described in Chapter VI.

Let $\mathbb{P}(X_1 = \frac{1}{2}) = \mathbb{P}(X_1 = \frac{3}{2}) = \frac{1}{2}$. Suppose that we are interested in the behavior of $\mathbb{E}\left[(\frac{1}{n}S_n)^n\right]$ for large n. Naively we might think that $\frac{1}{n}\log\mathbb{E}\left[(\frac{1}{n}S_n)^n\right] \to 0$ because $\frac{1}{n}S_n \to 1$ by the SLLN. However, this intuition is wrong. To see why, we compute (recall the notation introduced in (I.1))

$$\begin{aligned}
\mathbb{E}\left[(\tfrac{1}{n}S_n)^n\right] &= \int_0^\infty a^n \, \mathbb{P}(\tfrac{1}{n}S_n \in da) \\
&= \int_0^\infty n a^{n-1} \mathbb{P}(\tfrac{1}{n}S_n \geq a)\, da \\
&\simeq \int_0^\infty \exp\{n[\log a - J(a)]\}\, da \\
&\simeq \exp\left\{n \sup_{a>0}[\log a - J(a)]\right\},
\end{aligned}$$

where J is the rate function given by Cramér's Theorem, i.e., $J(a) = I(a - \frac{1}{2})$ with I as in Theorem I.3. Hence we find

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[(\tfrac{1}{n}S_n)^n\right] = b \quad \text{with} \quad b = \sup_{a>0}[\log a - J(a)].$$

EXERCISE I.25 *Let a^* be the maximizer of the variational expression in the last display. Show that $a^* \neq 1$ and $b > 0$.*

Thus, the expectation $\mathbb{E}\left[(\frac{1}{n}S_n)^n\right]$ is not dominated by the almost sure behavior, but rather by the rare event where $\frac{1}{n}S_n$ is in the vicinity of a^*.

The above calculation, which is nothing other than an example of Laplace's method for exponential integrals, can easily be made rigorous by appealing to (I.2). Indeed, cut $[\frac{1}{2}, \frac{3}{2}]$ into finitely many small intervals and apply (I.2) iteratively, in combination with the fact that I is continuous on $[\frac{1}{2}, \frac{3}{2}]$. We leave the details to the reader. In Section III.3 we will encounter a more powerful way to approach this kind of argument.

CHAPTER II

LARGE DEVIATIONS FOR I.I.D. SEQUENCES: PART 2

Throughout this chapter, we restrict ourselves to the situation where the X_i take values in a *finite* set:

$$\begin{aligned}&X_i \in \Gamma = \{1, \dots, r\} \subset \mathbb{N},\\ &X_1, X_2, \dots \text{ are i.i.d. with marginal law } \rho = (\rho_s)_{s\in\Gamma},\\ &\rho_s > 0 \ \forall s \in \Gamma.\end{aligned} \tag{II.1}$$

In Section II.7 we will worry about how to relax the condition that Γ be finite.

II.1 Sanov's Theorem for the empirical measure

In this section we generalize the result of Theorem I.4 to a statement about the limiting frequencies at which the random variables $X_1, X_2, \dots$ take their values in Γ. Theorem II.2 below goes back to Sanov [D52].

The values that occur along the sequence $X_1, \dots, X_n$ are recorded by means of the *empirical measure*

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

with δ_x denoting the point-mass at $x \in \mathbb{R}$. Note that L_n is a random probability measure on Γ. We write

$$\mathfrak{M}_1(\Gamma) = \Big\{\nu = (\nu_1, \dots, \nu_r) \in [0,1]^r \colon \sum_{s=1}^{r} \nu_s = 1\Big\}$$

to denote the probability simplex in $\mathbb{R}^r$, which may be identified with the set of probability measures on Γ. On $\mathfrak{M}_1(\Gamma)$ we define the total variation distance

$$d(\mu, \nu) = \frac{1}{2}\sum_{s=1}^{r} |\mu_s - \nu_s|,$$

which turns $\mathfrak{M}_1(\Gamma)$ into a Polish space (i.e., a complete separable metric space).

According to the SLLN,

$$d(L_n, \rho) \underset{n\to\infty}{\longrightarrow} 0 \qquad \mathbb{P}\text{–a.s.},$$

where $\rho = (\rho_1, \dots, \rho_r)$. The following theorem is a statement about the large deviations of L_n away from ρ.

THEOREM II.2 *Let* (X_i) *be i.i.d. random variables satisfying* (II.1). *Let* $L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$. *Then, for all* $a > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(L_n \in B_a^c(\rho)\big) = -\inf_{\nu \in B_a^c(\rho)} I_\rho(\nu),$$

where $B_a(\rho) = \{\nu \in \mathfrak{M}_1(\Gamma)\colon\ d(\nu,\rho) \le a\}$, $B_a^c(\rho) = \mathfrak{M}_1(\Gamma) \setminus B_a(\rho)$, *and*

$$I_\rho(\nu) = \sum_{s=1}^{r} \nu_s \log\left(\frac{\nu_s}{\rho_s}\right) \tag{II.3}$$

(with the convention $\inf_\emptyset I_\rho = \infty$*).*

PROOF. Let

$$K_n = \Big\{k = (k_1, \dots, k_r) \in \mathbb{N}_0^r\colon\ \sum_{s=1}^{r} k_s = n\Big\},$$

and note that $\frac{1}{n}K_n \subset \mathfrak{M}_1(\Gamma)$ for all $n \in \mathbb{N}$. Then L_n has the multinomial distribution

$$\mathbb{P}\Big(L_n(s) = \frac{k_s}{n}\ \forall s\Big) = n! \prod_{s=1}^{r} \frac{\rho_s^{k_s}}{k_s!}, \qquad k \in K_n.$$

For $k \in K_n$, let $\nu_n(k) = \frac{1}{n}k \in \mathfrak{M}_1(\Gamma)$. In analogy with the proof of Theorem I.1, let us put

$$Q_n(a) = \max_{k\in K_n\colon \nu_n(k)\in B_a^c(\rho)} \Big(n! \prod_{s=1}^{r} \frac{\rho_s^{k_s}}{k_s!}\Big).$$

Then, clearly,

$$Q_n(a) \le \mathbb{P}\big(L_n \in B_a^c(\rho)\big) \le |K_n| Q_n(a).$$

Stirling's formula gives

$$\frac{1}{n}\log\Big(n! \prod_{s=1}^{r} \frac{\rho_s^{k_s}}{k_s!}\Big) = \sum_{s=1}^{r} \frac{k_s}{n}\Big(\log\rho_s - \log\frac{k_s}{n}\Big) + O\Big(\frac{\log n}{n}\Big) \quad \text{uniformly on } K_n,$$

where we use that $\sum_{s=1}^{r} k_s = n$. Since the sum in the RHS equals $-I_\rho(\nu_n(k))$ and since $|K_n| = \binom{n+r}{r-1} = O(n^{r-1})$, we find that

$$\begin{aligned} \tfrac{1}{n}\log\mathbb{P}\big(L_n \in B_a^c(\rho)\big) &= O\Big(\tfrac{\log n}{n}\Big) + \tfrac{1}{n}\log Q_n(a) \\ &= O\Big(\tfrac{\log n}{n}\Big) - \min_{k\in K_n\colon \nu_n(k)\in B_a^c(\rho)} I_\rho\big(\nu_n(k)\big). \end{aligned}$$

In order to complete the proof, it now remains to observe that

(i) $\bigcup_{n\in\mathbb{N}} \{\nu_n(k)\colon k \in K_n\}$ is dense in $\mathfrak{M}_1(\Gamma)$,

(ii) $\nu \mapsto I_\rho(\nu)$ is continuous on $\mathfrak{M}_1(\Gamma)$,

which are both readily checked. Indeed, (i) and (ii) guarantee that for every $\nu \in \mathfrak{M}_1(\Gamma)$ there exists a sequence $(k_n)_{n\in\mathbb{N}}$, with $k_n \in K_n$ for all n, such that

$$\lim_{n\to\infty} d\big(\nu_n(k_n), \nu\big) = 0, \qquad \lim_{n\to\infty} I_\rho\big(\nu_n(k_n)\big) = I_\rho(\nu).$$

Since $B_a^c(\rho)$ is an open set, this implies that

$$\limsup_{n\to\infty} \min_{k\in K_n\colon \nu_n(k)\in B_a^c(\rho)} I_\rho\big(\nu_n(k)\big) \le I_\rho(\nu) \qquad \forall \nu \in B_a^c(\rho).$$

Optimizing over ν, we get

$$\limsup_{n\to\infty} \min_{k\in K_n\colon \nu_n(k)\in B_a^c(\rho)} I_\rho\big(\nu_n(k)\big) \le \inf_{\nu\in B_a^c(\rho)} I_\rho(\nu).$$

However, the reverse inequality is trivial, and so the statement follows. □

Equation (II.3) says that $I_\rho(\nu) = H(\nu|\rho)$, the *relative entropy of* ν *with respect to* ρ. It has the following properties:

LEMMA II.4 *Assume* (II.1).
(i) I_ρ *is finite, continuous and strictly convex on* $\mathfrak{M}_1(\Gamma)$.
(ii) $I_\rho(\nu) \geq 0$ *with equality if and only if* $\nu = \rho$.

EXERCISE II.5 *Prove Lemma* II.4.

Relative entropy is a key notion in ergodic theory, information theory and statistical physics (for background, see e.g. Petersen [C6] Section 5.1). The choice of distance in the proof of Theorem II.2 is obviously flexible. All that we really need is (i–ii). The proof shows that the same statement as in Theorem II.2 holds when $B_a^c(\rho)$ is replaced by an arbitrary open subset of $\mathfrak{M}_1(\Gamma)$. In fact, it holds for much more general sets, but again, we will postpone such topological considerations to Chapter III (compare with Comment (9) in Section I.4).

EXERCISE II.6 *Compute* $I_\rho(\nu)$ *when* ρ *is the uniform distribution on* Γ *and* ν *is the uniform distribution on a subset of* Γ.

II.2 The pair empirical measure

The empirical measure records one value from $X_1, X_2, \ldots$ at each instant of time. It is possible to expand on this by recording two successive values at each instant of time. This will turn out to be very useful in Chapter IV, where we will drop the assumption that the sequence be i.i.d. and will consider Markov sequences.

More precisely, let us introduce the *pair empirical measure*

$$L_n^2 = \frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i, X_{i-1})},$$

with the convention that periodic boundary conditions be used, i.e., $X_{n+1} = X_1$. The random measure L_n^2 belongs to the set $\mathfrak{M}_1(\Gamma \times \Gamma)$. In fact, because of the periodic boundary conditions, it belongs to the subset

$$\widetilde{\mathfrak{M}}_1(\Gamma \times \Gamma) = \Big\{\nu = (\nu_{st}) \in \mathfrak{M}_1(\Gamma \times \Gamma)\colon\ \sum_t \nu_{st} = \sum_t \nu_{ts}\ \forall s\Big\},$$

i.e., those probability measures on $\Gamma \times \Gamma$ whose marginals coincide. As before, $\widetilde{\mathfrak{M}}_1(\Gamma \times \Gamma)$ turns into a Polish space with the total variation distance

$$d(\mu, \nu) = \frac{1}{2} \sum_{s,t} |\mu_{st} - \nu_{st}|.$$

It follows from Birkhoff's Ergodic Theorem (Petersen [C6] Section 2.2) that

$$d(L_n^2, \rho \times \rho) \underset{n \to \infty}{\longrightarrow} 0 \qquad \mathbb{P}\text{–a.s.}$$

EXERCISE II.7 *Show that the last statement can also be derived from the* SLLN.

The analogue of Theorem II.2 describing the large deviations of L_n^2 away from $\rho \times \rho$ reads:

THEOREM II.8 *Let (X_i) be i.i.d. random variables satisfying* (II.1). *Let $L_n^2 = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i,X_{i+1})}$ with periodic boundary conditions. Then, for all $a > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(L_n^2 \in B_a^c(\rho\times\rho)\big) = -\inf_{\nu\in B_a^c(\rho\times\rho)} I_\rho^2(\nu),$$

where $B_a(\rho\times\rho) = \{\nu \in \widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma)\colon d(\nu,\rho\times\rho) \le a\}$ and

$$I_\rho^2(\nu) = \sum_{s,t} \nu_{st} \log\Big(\frac{\nu_{st}}{\bar{\nu}_s \rho_t}\Big), \tag{II.9}$$

with $\bar{\nu}_s = \sum_t \nu_{st}$.

PROOF. The proof is similar in spirit to that of Theorem II.2, but the combinatorics is more involved. This is due to the fact that the pairs in the definition of L_n^2 are "interlocked". The following argument is an expansion of Ellis [A5] Section I.5.

Let

$$K_n = \Big\{k = (k_{st}) \in \mathbb{N}_0^{2r}\colon \sum_{s,t} k_{st} = n, \sum_t k_{st} = \sum_t k_{ts}\ \forall s\Big\},$$

and note that $\frac{1}{n}K_n \subset \widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma)$ for all $n\in\mathbb{N}$. Let us write $\bar{k}_s = \sum_t k_{st}$. Then, for all $k \in K_n$,

$$\mathbb{P}\Big(L_n^2(s,t) = \frac{k_{st}}{n}\ \forall s,t\Big)$$

will be of the form $\prod_s \rho_s^{\bar{k}_s}$ times a combinatorial factor accounting for all the possible arrangements of $X_1,\dots,X_n$ that give rise to $k = (k_{st})$. For the evaluation of this combinatorial factor we have to make a small excursion into graph theory.

Let us mark each occurence of (s,t) in $X_1,\dots,X_n$ by drawing an arrow from s to t. In this way we obtain an oriented graph $G(k)$, having Γ as its set of vertices and the arrows as its set of oriented edges (see Fig. 4). The periodic boundary conditions ensure that for each vertex s the number of ingoing arrows (i.e., $\sum_t k_{ts}$) equals the number of outgoing arrows (i.e., $\sum_t k_{st}$). Some vertices of Γ may be without arrows, but this will not matter for the argument below. The total number of arrows is n.

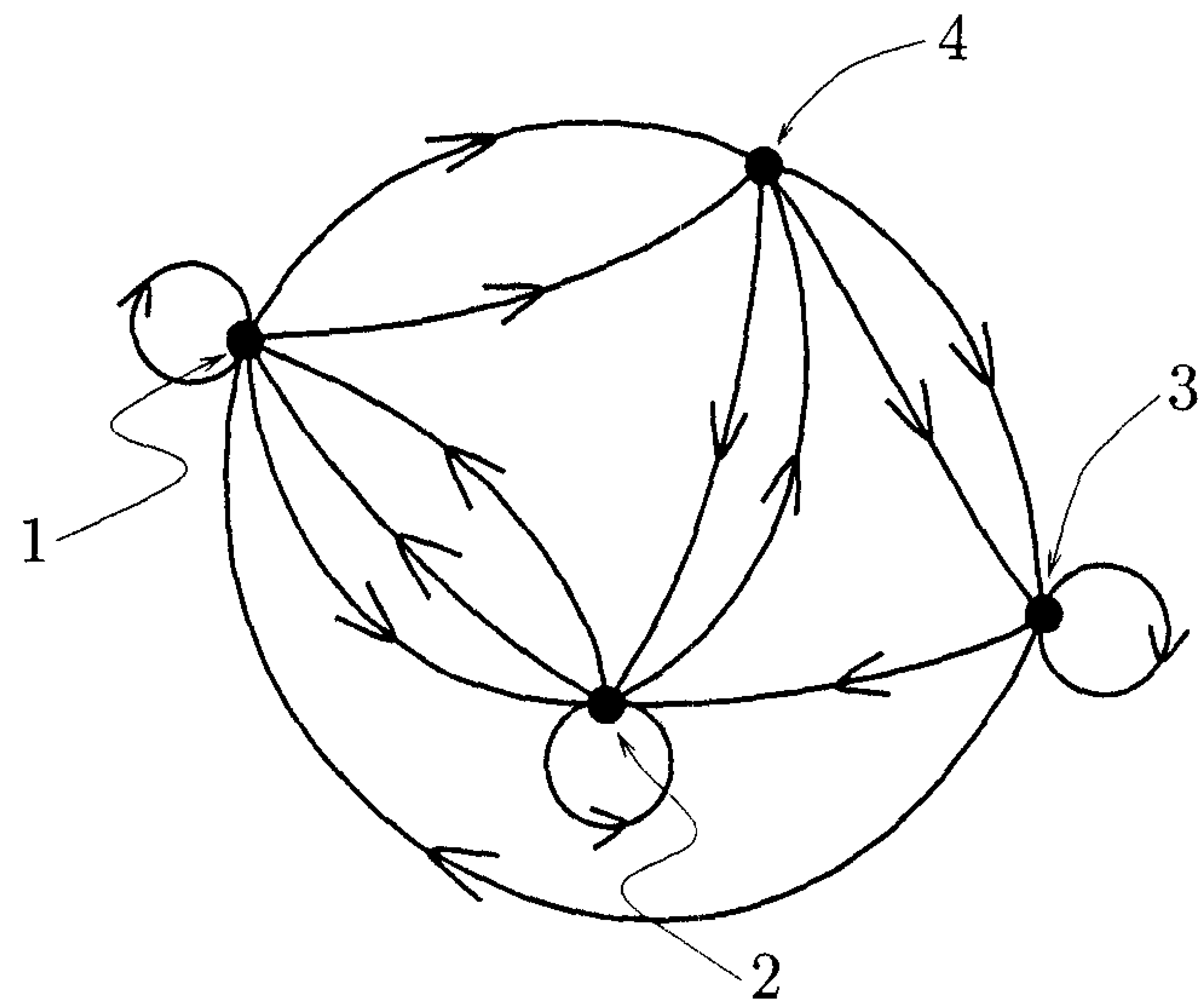

FIG. 4. The graph $G(k)$ for $r = 4$ and $k = \begin{pmatrix} 1&1&0&2\\ 2&1&0&1\\ 1&1&1&0\\ 0&1&2&0 \end{pmatrix}$

We now have the following characterization:

$$\mathbb{P}\Big(L_n^2(s,t)=\frac{k_{st}}{n}\ \forall s,t\Big)=\sharp\big(G(k)\big)\,\frac{\mathcal{E}\big(G(k)\big)}{\prod\limits_{s,t}k_{st}!}\prod_s \rho_s^{\bar{k}_s}.$$

Here, $\mathcal{E}\big(G(k)\big)$ denotes the number of *Euler circuits* on $G(k)$ (i.e., the number of looped paths respecting the arrows and making use of each arrow precisely once), the factors $k_{st}!$ compensate for distinguishing between the different arrows from s to t, while the factor $\sharp\big(G(k)\big)$ counts the number of cyclic shifts of $X_1,\dots,X_n$ that are distinct. This last factor obviously satisfies

$$1\le \sharp\big(G(k)\big)\le n.$$

The following lemma gives us control over the RHS in the previous display.

LEMMA II.10

$$\prod_{s:\,\bar{k}_s>0}(\bar{k}_s-1)!\le \mathcal{E}\big(G(k)\big)\le \prod_{s:\,\bar{k}_s>0}\bar{k}_s!$$

PROOF. It is easy to see that the condition $\sum_t k_{ts}=\sum_t k_{st}\ \forall s$ ensures that there is at least one Euler circuit. Pick any Euler circuit $\mathcal{C}$. Let us go round $\mathcal{C}$ and, for each vertex, assign the color red to the outgoing arrow that is used last in $\mathcal{C}$. If after that we permute the non-colored arrows, of which vertex s has $\bar{k}_s-1$, then we again get an Euler circuit. Indeed, the circuit can never get stuck after the permutation, because the red arrows serve as "escape route": once the circuit has used a red arrow it can never return to the vertex from which this red arrow goes out. Clearly, all Euler circuits obtained by such permutations are distinct. This gives us the lower bound.

The upper bound is obtained as follows. We may attempt to build an Euler circuit by picking from the arrows as we go along. In this way we may or may not end up generating an Euler circuit: we will get stuck at the starting vertex after we have used up all the outgoing arrows there, even though we may not yet have used up all the other arrows. Clearly, we cannot make more than $\bar{k}_s!$ different decisions where to go from vertex s. This gives us the upper bound. □

Now we can finish the proof of Theorem II.8. With the help of the preceding observations we find that

$$\mathbb{P}\Big(L_n^2(s,t)=\frac{k_{st}}{n}\ \forall s,t\Big)=e^{O(\log n)}\,\frac{\prod\limits_s \bar{k}_s!}{\prod\limits_{s,t}k_{st}!}\prod_s \rho_s^{\bar{k}_s}\quad \text{uniformly on } K_n,$$

where we use that the gap between the bounds in Lemma II.10 is $1\le\prod_{s:\,\bar{k}_s>0}\bar{k}_s\le n^r$.

The rest of the proof is analogous to that of Theorem II.2. For $k\in K_n$, let $\nu_n(k)=\frac{1}{n}k\in\mathfrak{M}_1(\Gamma\times\Gamma)$. Then, clearly,

$$Q_n(a)\le\mathbb{P}\big(L_n^2\in B_a^c(\rho\times\rho)\big)\le|K_n|Q_n(a),$$

with

$$Q_n(a)=\max_{k\in K_n:\,\nu_n(k)\in B_a^c(\rho\times\rho)}\mathbb{P}\Big(L_n^2(s,t)=\frac{k_{st}}{n}\ \forall s,t\Big).$$

Via Stirling's formula and the observation that $|K_n| = O(n^{r^2-1}) = e^{O(\log n)}$ we get

$$\begin{aligned}\tfrac{1}{n}\log \mathbb{P}\big(L_n^2 \in B_a^c(\rho\times\rho)\big) &= O\Big(\tfrac{\log n}{n}\Big) + \tfrac{1}{n}\log Q_n(a)\\ &= O\Big(\tfrac{\log n}{n}\Big) - \min_{k\in K_n:\ \nu_n(k)\in B_a^c(\rho\times\rho)} I_\rho^2\big(\nu_n(k)\big),\end{aligned}$$

which gives us the result after letting $n \to \infty$ and using the analogues of (i) and (ii) at the end of the proof of Theorem II.2. □

Equation (II.9) says that $I_\rho^2(\nu) = H(\nu|\bar\nu\times\rho)$, the *relative entropy of ν with respect to $\bar\nu\times\rho$*. It has the following properties:

LEMMA II.11 *Assume* (II.1). *Then:*
(i) I_ρ^2 *is finite, continuous and strictly convex on* $\widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma)$, *except along line segments* $\{\alpha\nu + (1-\alpha)\nu' \colon \alpha \in [0,1]\}$ *between any ν and ν' satisfying* $\nu_{st}/\bar\nu_s = \nu'_{st}/\bar\nu'_s$ $\forall s,t$. *Along such line segments I_ρ^2 is affine, i.e.,* $I_\rho^2(\alpha\nu+(1-\alpha)\nu') = \alpha I_\rho^2(\nu) + (1-\alpha)I_\rho^2(\nu')$ $\forall\alpha\in[0,1]$.
(ii) $I_\rho^2(\nu) \geq 0$ *with equality if and only if* $\nu = \rho\times\rho$.

EXERCISE II.12 *Prove Lemma* II.11. *Explain where the affine part of I_ρ^2 comes from.*

Comparing I_ρ^2 in Theorem II.8 with I_ρ in Theorem II.2, we note that $\bar\nu_s\rho_t$ appears in the denominator instead of $\rho_s\rho_t$. This comes from the fact that we are recording $(X_1,X_2),(X_2,X_3),\ldots$ rather than $(X_1,X_2),(X_3,X_4),\ldots$ Two alternative ways to write I_ρ^2 are:

$$\begin{aligned}I_\rho^2(\nu) &= I_\rho(\bar\nu) + H(\nu|\bar\nu\times\bar\nu)\\ &= \sum_s \bar\nu_s H(\nu[s]|\rho) \quad \text{with } \nu[s]\in\mathfrak{M}_1(\Gamma) \text{ defined by } \nu[s]_t = \nu_{st}/\bar\nu_s,\end{aligned}$$

which elucidates the effect of the interlocking of pairs. Thus, we have a "two layer structure": a large deviation of L_n^2 arises from a large deviation of L_n combined with a large deviation of L_n^2 given the marginal L_n.

II.3 A toy application of Sanov for pairs

We demonstrate the power of Theorem II.8 by the following example, which plays the same illustrative role as the example in Section I.5. A more serious application will be described in Chapter IX.

Let $\Gamma = \{1,2\}$ and $\rho = (1-p,p)$ with $p\in(0,1)$, i.e., $\mathbb{P}(X_1 = 1) = 1-p$ and $\mathbb{P}(X_1 = 2) = p$. Let $Y_1, Y_2\ldots$ be defined as follows: put $X_0 = Y_0 = 1$ and, for $i = 0,1,2,\ldots$, let

$$Y_{i+1} = \begin{cases} 2Y_i & \text{if } X_{i+1}\neq X_i\\ Y_i & \text{if } X_{i+1} = X_i.\end{cases}$$

The question is how $\mathbb{E}\,Y_n$ behaves for large n.

In view of the example in Section I.5, it is clear that the naive guess

$$\mathbb{E}\,Y_n = \mathbb{E}\,2^{\#\{0\leq i<n:\ X_{i+1}\neq X_i\}} \simeq 2^{\mathbb{E}\,\#\{0\leq i<n:\ X_{i+1}\neq X_i\}} = 2^{n2p(1-p)}$$

is illusory, since it does not appropriately account for the contribution of rare events. For a precise evaluation, we write

$$\mathbb{E}\,Y_n = O(1)\,\mathbb{E}\,2^{n[L_n^2(1,2)+L_n^2(2,1)]},$$

where the factor $O(1)$ arises because L_n^2 is defined with periodic boundary conditions. With the help of Theorem II.8 and (I.2) we find that

$$\lim_{n\to\infty}\frac{1}{n}\log \mathbb{E}\, Y_n = \sup_{\nu\in\widetilde{\mathfrak{M}}_1(\{1,2\}\times\{1,2\})}\big[(\nu_{12}+\nu_{21})\log 2 - I_\rho^2(\nu)\big].$$

EXERCISE II.13 *(i) Calculate* $\mathbb{E}\, Y_n$ *for* $p=\frac{1}{2}$ *and* $n\in\mathbb{N}$. *Hint: Note that flips* $0\leftrightarrow 1$ *along the sequence* $X_1,\ldots,X_n$ *occur independently when* $p=\frac{1}{2}$.
(ii) Check that the answer fits with the variational expression in the last display.

EXERCISE II.14 *Let* $\nu^*(p)=(\nu_{11}(p),\nu_{12}(p),\nu_{21}(p),\nu_{22}(p))$ *be the maximizer of the variational expression. Show that* $\nu^*(p)\neq((1-p)^2,p(1-p),p(1-p),p^2)$.

II.4 A contraction principle

It is intuitively clear that the large deviations of the empirical average $\frac{1}{n}S_n$ (Cramér's Theorem) can be derived from the large deviations of the empirical measure L_n (Sanov's Theorem). Namely, for the set-up in (II.1) we have the relation

$$\frac{1}{n}S_n=\sum_{s=1}^{r} sL_n(s).$$

The precise link is given by the following theorem taken from Ellis [A5] Section VIII.3.

THEOREM II.15 *Let* (X_i) *be i.i.d. random variables satisfying* (II.1). *For* $\nu\in\mathfrak{M}_1(\Gamma)$, *let* $m_\nu=\sum_s s\nu_s$. *Then, for all* $a>0$,

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big(\frac{1}{n}S_n\in B_a^{\mathrm{c}}(m_\rho)\Big)=-\inf_{z\in B_a^{\mathrm{c}}(m_\rho)} I(z),$$

where $B_a(m_\rho)=\{z\in\mathbb{R}\colon\ |z-m_\rho|\le a\}$ *and*

$$I(z)=\inf_{\nu\in\mathfrak{M}_1(\Gamma)\colon\, m_\nu=z} I_\rho(\nu). \tag{II.16}$$

PROOF. First note that $\{\frac{1}{n}S_n\in B_a^{\mathrm{c}}(m_\rho)\}$ is the same as $\{L_n\in\widehat{B}_a^{\mathrm{c}}(\rho)\}$ with

$$\widehat{B}_a(\rho)=\big\{\nu\in\mathfrak{M}_1(\Gamma)\colon\ |m_\nu-m_\rho|\le a\big\}.$$

Therefore, by Theorem II.2 with $B_a^{\mathrm{c}}(\rho)$ replaced by $\widehat{B}_a^{\mathrm{c}}(\rho)$, we have

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big(\frac{1}{n}S_n\in B_a^{\mathrm{c}}(m_\rho)\Big)=-\inf_{\nu\in\widehat{B}_a^{\mathrm{c}}(\rho)} I_\rho(\nu).$$

Indeed, $\widehat{B}_a^{\mathrm{c}}(\rho)$ is an open subset of $\mathfrak{M}_1(\Gamma)$ (w.r.t. the total variation distance) and so the remark made at the end of Section II.1 applies here. The claim now follows from the observation that

$$\inf_{\nu\in\widehat{B}_a^{\mathrm{c}}(\rho)} I_\rho(\nu)=\inf_{z\in B_a^{\mathrm{c}}(m_\rho)}\ \inf_{\nu\in\mathfrak{M}_1(\Gamma)\colon\, m_\nu=z} I_\rho(\nu),$$

where the second infimum equals $I(z)$ in (II.16). □

Note that the heart of Theorem II.15 lies in the observation that

$$\frac{1}{n}S_n=z \qquad\Longleftrightarrow\qquad L_n\in\{\nu\in M(\Gamma)\colon\ m_\nu=z\}.$$

This link enables us to go from the empirical measure back to the empirical average. For obvious reasons Theorem II.15 is called a *contraction principle*. The formula linking I_ρ to I shows that large deviations are always done "in the least unlikely of all the unlikely ways", as was already observed at the end of Section I.4.

Theorem II.15 can be given the following pictorial interpretation (see Fig. 5). The sample space Γ^n carrying $X_1, \ldots, X_n$ can be partitioned into boxes, arranged in columns and rows, with L_n approximately constant in each box and $m_{L_n} = \frac{1}{n}S_n$ approximately constant in each column. Apart from the box corresponding to "typical" sequences $X_1, \ldots, X_n$ for which $L_n \approx \rho$ and $\frac{1}{n}S_n \approx m_\rho$ (corresponding to the shaded box in Fig. 5), all boxes have an exponentially small probability. The evaluation of $\mathbb{P}(|\frac{1}{n}S_n - m_\rho| > a)$ requires summing over the probabilities of all the columns with $|m_{L_n} - m_\rho| > a$. If the number of boxes in these columns does not grow exponentially fast with n (which is indeed the case when Γ is finite), then according to (I.2) the sum will be dominated by the box with the largest probability. This box is where $L_n \approx \nu^*$, with ν^* the minimizer of I_ρ corresponding to the value $z^* = m_{\nu^*}$ that minimizes I over the complement of the interval $[m_\rho - a, m_\rho + a]$.

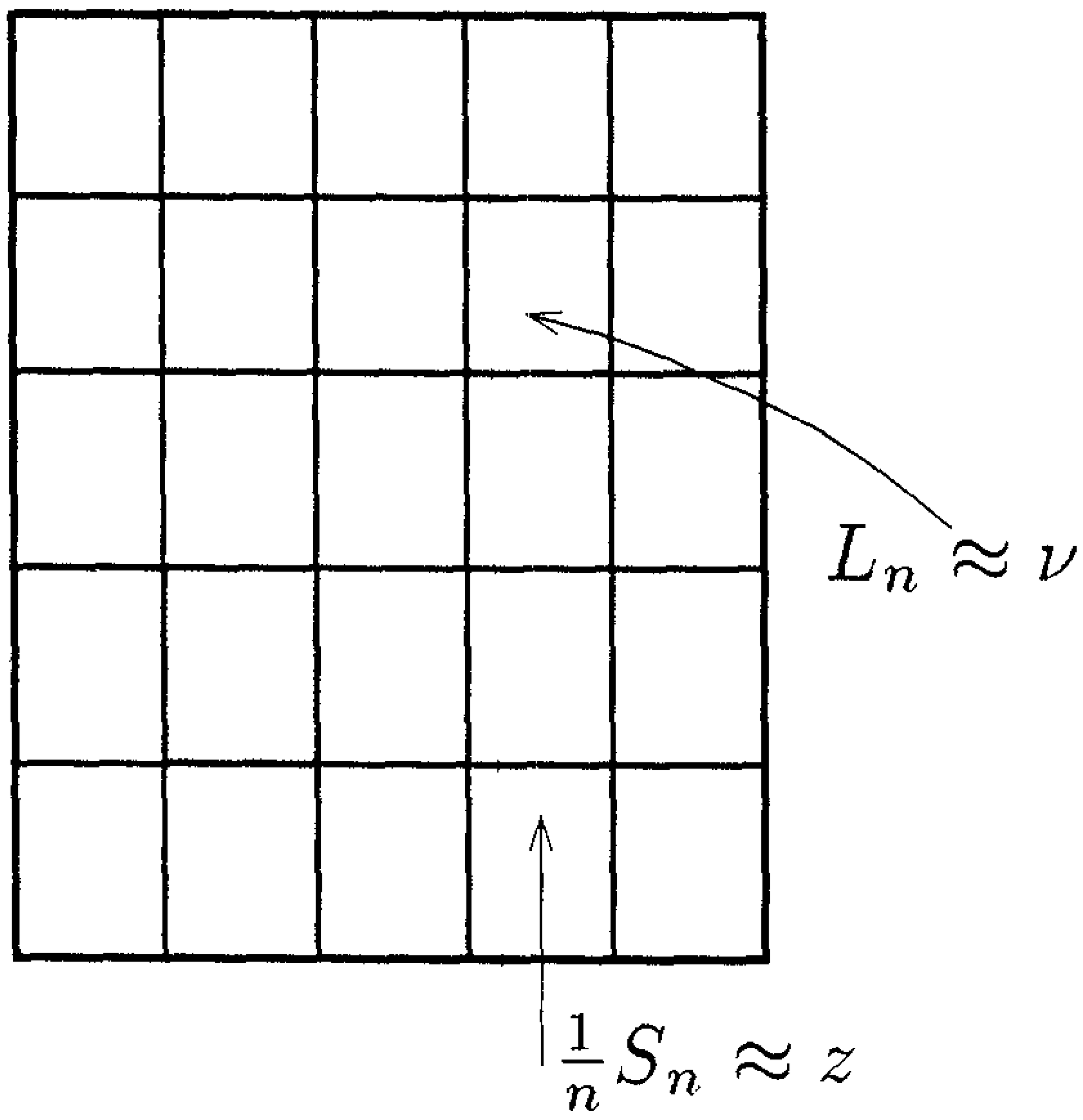

FIG. 5. Pictorial interpretation of the contraction principle

EXERCISE II.17 *Show, by using the method of Lagrange multipliers, that $I(z)$ of Theorem I.4 and $I(z)$ of Theorem II.15 coincide.*

There is an analogous contraction principle linking I_ρ^2 to I_ρ. We will not spell out the details. A general contraction principle will be described in Section III.5.

II.5 The empirical process

In the same way as we have counted pairs, we can also count N-words for any integer $N \geq 2$. For that we define the *N-word empirical measure*

$$L_n^N = \frac{1}{n}\sum_{i=1}^{n} \delta_{(X_i,\ldots,X_{i+N-1})},$$

where for convenience we again use periodic boundary conditions $X_{n+m} = X_m$ ($m = 1, \ldots, N-1$). The random measure L_n^N is an element of the set

$$\widetilde{\mathfrak{M}}_1(\Gamma^N) = \Big\{\nu = (\nu_{s_1,\ldots,s_N}) \in \mathfrak{M}_1(\Gamma^N): \sum_{s_N} \nu_{s_1,\ldots,s_{N-1},s_N} = \sum_{s_N} \nu_{s_N,s_1,\ldots,s_{N-1}} \ \forall s_1,\ldots,s_{N-1}\Big\},$$

which turns into a Polish space with the total variation distance

$$d(\mu,\nu) = \frac{1}{2} \sum_{s_1,\ldots,s_N} |\mu_{s_1,\ldots,s_N} - \nu_{s_1,\ldots,s_N}|.$$

In analogy with Theorem II.8, we have the following large deviation result for L_n^N:

THEOREM II.18 *Let (X_i) be i.i.d. random variables satisfying* (II.1). *Let $L_n^N = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i,\ldots,X_{i+N-1})}$ with periodic boundary conditions. Then, for all $a > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(L_n^N \in B_a^c(\rho^N)\big) = - \inf_{\nu\in B_a^c(\rho^N)} I_\rho^N(\nu),$$

where $B_a(\rho^N) = \{\nu \in \widetilde{\mathfrak{M}}_1(\Gamma^N) \colon d(\nu,\rho^N) \le a\}$ and

$$I_\rho^N(\nu) = \sum_{s_1,\ldots,s_N} \nu_{s_1,\ldots,s_N} \log\Big(\frac{\nu_{s_1,\ldots,s_N}}{\bar\nu_{s_1,\ldots,s_{N-1}}\rho_{s_N}}\Big), \tag{II.19}$$

with $\bar\nu_{s_1,\ldots,s_{N-1}} = \sum_{s_N} \nu_{s_1,\ldots,s_{N-1},s_N}$.

PROOF. The proof is a straightforward extension of the proof of Theorem II.8 in Section II.2 and amounts to counting Euler circuits on a graph where the arrows label N-words. We leave this for the reader to verify. □

Equation (II.19) says that $I_\rho^N(\nu) = H(\nu|\bar\nu \times \rho)$. The choice of distance is again flexible, and the same result holds when $B_a^c(\rho^N)$ is replaced by an arbitrary open set. The analogue of Theorem II.11 reads:

LEMMA II.20 *Assume* (II.1). *Then:*
(i) *I_ρ^N is finite, continuous and strictly convex on $\widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma)$, except along line segments $\{\alpha\nu+(1-\alpha)\nu' \colon \alpha \in [0,1]\}$ between any ν and ν' satisfying $\nu_{s_1,\ldots,s_N}/\bar\nu_{s_1,\ldots,s_{N-1}} = \nu'_{s_1,\ldots,s_N}/\bar\nu'_{s_1,\ldots,s_{N-1}}$ $\forall s_1,\ldots,s_N$. Along such line segments I_ρ^N is affine, i.e., $I_\rho^N(\alpha\nu + (1-\alpha)\nu') = \alpha I_\rho^N(\nu) + (1-\alpha) I_\rho^N(\nu')$ $\forall \alpha \in [0,1]$.*
(ii) *$I_\rho^N(\nu) \ge 0$ with equality if and only if $\nu = \rho^N$.*

PROOF. Same as in Exercise II.12. □

Similarly as in Section II.4, we have a contraction principle from L_n^N to L_n^{N-1}, etc.

EXERCISE II.21 *Show that $\bar\nu \in \widetilde{\mathfrak{M}}_1(\Gamma^{N-1})$.*

The N-word empirical measure tells us about the frequencies at which words of length N occur along the sequence $X_1,\ldots,X_n$. With a leap of imagination, we can push this situation to the extreme where the words that we are counting have length n themselves, i.e.,

$$L_n^n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(X_i,\dots,X_{i+n-1})},$$

with periodic boundary conditions $X_{n+m} = X_m$ ($m = 1,\dots,n-1$). In order to study the large deviation behavior of L_n^n as $n \to \infty$, it is useful to extend $X_1,\dots,X_n$ periodically by putting

$$\mathbb{X}^{(n)} = \underbrace{X_1,\dots,X_n}, \underbrace{X_1,\dots,X_n}, \underbrace{X_1,\dots,X_n}, \dots$$

and defining

$$R_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\sigma^i(\mathbb{X}^{(n)})},$$

with σ the left-shift on $\Gamma^{\mathbb{N}}$ (i.e., $(\sigma x)_i = x_{i+1}$ for $i \in \mathbb{N}$ and $x \in \Gamma^N$). Clearly, R_n is a random element of the set of σ-invariant probability measures on $\Gamma^{\mathbb{N}}$:

$$\widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}}) = \{\nu \in \mathfrak{M}_1(\Gamma^{\mathbb{N}}) \colon \nu \circ \sigma^{-1} = \nu\}.$$

Since the latter can be viewed as the set of stationary random processes, R_n is called the *empirical process.* By Birkhoff's Ergodic Theorem, we have

$$R_n \underset{n\to\infty}{\longrightarrow} \rho^{\mathbb{N}} \text{ weakly} \qquad \mathbb{P}\text{–a.s.},$$

where weakly stands for convergence on cylinder sets.

For $N \in \mathbb{N}$, let d_N denote the total variation distance on $\mathfrak{M}_1(\Gamma^N)$. For $\mu,\nu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}})$, define

$$d(\mu,\nu) = \sum_{N\in\mathbb{N}} 2^{-N} d_N(\pi_N\mu, \pi_N\nu),$$

where $\pi_N \colon \Gamma^{\mathbb{N}} \to \Gamma^N$ is the projection that chops off all but the first N coordinates (i.e., $\pi_N x = (x_1,\dots,x_N)$ for $x \in \Gamma^{\mathbb{N}}$) and $\pi_N\mu = \mu \circ \pi_N^{-1}$. Clearly, d turns $\widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}})$ into a Polish space.

EXERCISE II.22 *Show that if $\mu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}})$, then $\pi_N\mu \in \widetilde{\mathfrak{M}}_1(\Gamma^N)$ for all $N \in \mathbb{N}$.*

The following theorem captures the large deviations of R_n away from $\rho^{\mathbb{N}}$. (See Varadhan [A8] Section 8, Ellis [A5] Chapter IX, Dembo and Zeitouni [A2] Section 6.5.3.)

THEOREM II.23 *Let (X_i) be an i.i.d. sequence satisfying* (II.1). *For $a > 0$, let $B_a(\rho^{\mathbb{N}}) = \{\nu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}}) \colon d(\nu,\rho^{\mathbb{N}}) \le a\}$, and define*

$$J(a) = \inf_{\nu\in B_a^c(\rho^{\mathbb{N}})} I_\rho^\infty(\nu),$$

with

$$I_\rho^\infty(\nu) = \sup_{N\ge 2} I_\rho^N(\pi_N\nu) = \sup_{N\ge 2} H(\pi_N\nu|\pi_{N-1}\nu \times \rho). \tag{II.24}$$

Then:
(a) $\liminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}\big(R_n \in B_a^c(\rho^{\mathbb{N}})\big) \ge -J(a)$ *for all $a > 0$.*
(b) $\limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}\big(R_n \in B_a^c(\rho^{\mathbb{N}})\big) \le -J(a-)$ *for all $a > 0$.*
Here, $J(a-) = \lim_{\delta\downarrow 0} J(a-\delta)$.

PROOF. The claim will follow from the large deviation behavior of L_n^N as $n \to \infty$ with N fixed, together with an argument showing that we can pass to the limit $N \to \infty$ afterwards. Along the way we will need the following three properties:

1. $N \mapsto I_\rho^N(\pi_N \nu)$ is non-decreasing for every $\nu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}})$.
2. $a \mapsto J(a)$ is non-decreasing and right-continuous.
3. $\inf_{\nu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}}):\ \pi_M \nu = \mu_M} I_\rho^\infty(\nu) = I_\rho^M(\mu_M)$ for all $M \in \mathbb{N}$ and $\mu_M \in \widetilde{\mathfrak{M}}_1(\Gamma^M)$.

EXERCISE II.25 *Prove properties 1 and 2.*

The proof of property 3 (which is a contraction principle) is deferred to the end of the proof.

(a) Lower bound:

The idea is to use the fact that the distance d is "myopic to far away coordinates". Since $\pi_N R_n = L_n^N$ for $n \geq N \geq 1$, we have

$$\mathbb{P}\big(d(R_n, \rho^{\mathbb{N}}) > a\big) \geq \mathbb{P}\Big(\sum_{N=1}^{M} 2^{-N} d_N(L_n^N, \rho^N) > a\Big) \qquad \forall n \geq M \geq 1.$$

By applying the large deviation result for L_n^M in Theorem II.18, we get (recall the remark made below Lemma II.18)

$$\begin{aligned}
&\lim_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\Big(\sum_{N=1}^{M} 2^{-N} d_N(L_n^N, \rho^N) > a\Big) \\
&= - \inf_{\nu_M \in \widetilde{\mathfrak{M}}_1(\Gamma^M):\ \sum_{N=1}^{M} 2^{-N} d_N(\pi_N \nu_M, \rho^N) > a} I_\rho^M(\nu_M) \\
&= - \inf_{\nu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}}):\ \sum_{N=1}^{M} 2^{-N} d_N(\pi_N \nu, \rho^N) > a} I_\rho^\infty(\nu) \qquad \forall M \geq 1,
\end{aligned}$$

where the last equality follows from property 3. Hence we find

$$\begin{aligned}
&\liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big(d(R_n, \rho^{\mathbb{N}}) > a\big) \\
&\geq - \inf_{\nu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}}):\ d(\nu, \rho^{\mathbb{N}}) > a + 2^{-M}} I_\rho^\infty(\nu) = -J(a + 2^{-M}) \qquad \forall M \geq 1.
\end{aligned}$$

Now pass to the limit $M \to \infty$ and use property 2.

(b) Upper bound:

Simply note that

$$\mathbb{P}\big(d(R_n, \rho^{\mathbb{N}}) > a\big) \leq \mathbb{P}\Big(\sum_{N=1}^{M} 2^{-N} d_N(L_n^N, \rho^N) > a - 2^{-M}\Big) \quad \forall n \geq M \geq 1$$

and repeat the argument, to get

$$\begin{aligned}
&\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big(d(R_n, \rho^{\mathbb{N}}) > a\big) \\
&\leq - \inf_{\nu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}}):\ d(\nu, \rho^{\mathbb{N}}) > a - 2^{-M}} I_\rho^\infty(\nu) = -J(a - 2^{-M}) \qquad \forall M \geq 1.
\end{aligned}$$

Pass to the limit $M \to \infty$.

Proof of property 3:

Fix $M \in \mathbb{N}$ and $\mu_M \in \widetilde{\mathfrak{M}}_1(\Gamma^M)$. We begin by constructing a sequence of probability measures

$$\mu^N \in \widetilde{\mathfrak{M}}_1(\Gamma^N), \qquad N = M, M+1, M+2, \ldots,$$

with $\mu^M = \mu_M$, such that

$$\begin{array}{lll} \text{(i)} & \pi_N \mu^{N+1} = \mu^N & \forall N \geq M. \\ \text{(ii)} & I_\rho^{N+1}(\mu^{N+1}) = I_\rho^N(\mu^N) & \forall N \geq M. \end{array} \tag{II.26}$$

This sequence is defined by the recursive scheme

$$\mu^{N+1}_{s_1,s_2,\ldots,s_N,s_{N+1}} = \frac{\mu^N_{s_1,s_2,\ldots,s_N}\mu^N_{s_2,\ldots,s_N,s_{N+1}}}{\mu^{N-1}_{s_2,\ldots,s_N}}, \qquad N \geq M, \tag{II.27}$$

where we put $\mu^{M-1} = \bar{\mu}^M$. Identity (II.26)(i) is easily checked via induction on N. To get (II.26)(ii), first note that for all $N \in \mathbb{N}$ and all $\nu_N \in \widehat{\mathfrak{M}}_1(\Gamma^N)$ we may write (II.19) as

$$I_\rho^N(\nu_N) = -\sum_s (\pi_1\nu_N)_s \log \rho_s - h(\nu_N) + h(\pi_{N-1}\nu_N), \tag{II.28}$$

with

$$h(\nu_N) = -\sum_{s_1,\ldots,s_N} \nu_{s_1,\ldots,s_N} \log \nu_{s_1,\ldots,s_N}$$

the entropy of ν_N. Next note that (II.27) yields

$$h(\mu^{N+1}) = 2h(\mu^N) - h(\mu^{N-1}), \qquad N \geq M.$$

Combining this with (II.28), we find

$$I_\rho^{N+1}(\mu^{N+1}) - I_\rho^N(\mu^N) = -h(\mu^{N+1}) + 2h(\mu^N) - h(\mu^{N-1}) = 0 \qquad \forall N \geq M.$$

We proceed as follows. By (II.26)(i), the sequence $(\mu^N)_{N\geq M}$ is a consistent family of probability measures, and so by the Kolmogorov Extension Theorem there exists a $\mu^\infty \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}})$ such that $\pi_N\mu^\infty = \mu^N$ for all $N \geq M$. Moreover, by property 1, (II.24) and (II.26)(ii), we have

$$I_\rho^\infty(\mu^\infty) = \sup_{N\geq M} I_\rho^N(\pi_N\mu^\infty) = \sup_{N\geq M} I_\rho^N(\mu^N) = I_\rho^M(\mu^M).$$

Since $\mu^M = \mu_M$, it follows that

$$\inf_{\nu\in\widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}}):\ \pi_M\nu=\mu_M} I_\rho^\infty(\nu) \leq I_\rho^M(\mu_M).$$

However, the reverse inequality trivially holds by property 1. □

It is not a priori obvious when and where J is continuous. Consequently, the bounds in Theorem II.23 do not give us equality. This is the price we have to pay for wanting to deal with infinite words, because we were forced to do an approximation argument. In Chapter III we will come to appreciate this situation better.

The rate function I_ρ^∞ in (II.24) has the following properties:

LEMMA II.29 *Assume* (II.1).

(i) I_ρ^∞ *is finite, lower semi-continuous and affine on* $\widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}})$.

(ii) $I_\rho^\infty(\nu) \geq 0$ *with equality if and only if* $\nu = \rho^{\mathbb{N}}$.

EXERCISE II.30 *Prove Lemma* II.29.

II.6 Comments

(1) An alternative formula for I_ρ^∞ reads

$$I_\rho^\infty(\nu) = \lim_{N\to\infty} \frac{1}{N} H(\pi_N\nu|\rho^N). \tag{II.31}$$

Indeed, this easily follows from (II.24), property 1 and the recursion relation

$$H(\pi_N\nu|\rho^N) = H(\pi_N\nu|\pi_{N-1}\nu \times \rho) + H(\pi_{N-1}\nu|\rho^{N-1}), \qquad N \geq 2,$$

since iteration of the latter yields

$$H(\pi_N\nu|\rho^N) = \sum_{M=1}^{N} H(\pi_M\nu|\pi_{M-1}\nu \times \rho) = \sum_{M=1}^{N} I_\rho^M(\pi_M\nu), \qquad N \geq 1.$$

Thus, $I_\rho^\infty(\nu)$ is the relative entropy of ν with respect to $\rho^{\mathbb{N}}$ per coordinate. This is why $I_\rho^\infty(\nu)$ is called the *specific relative entropy* of ν with respect to $\rho^{\mathbb{N}}$.

(2) The fact that I_ρ^∞ is affine is a natural consequence of the fact that we are dealing with an infinite-dimensional situation. Recall that we already saw affineness sneaking into Lemma II.11 when we had words of length 2. The situation can be understood as follows. It follows from (II.31) that

$$I_\rho^\infty(\nu) = -\sum_s (\pi_1\nu)_s \log \rho_s - \widehat{h}(\nu), \tag{II.32}$$

with

$$\widehat{h}(\nu) = \lim_{N\to\infty} \frac{1}{N} h(\pi_N\nu) \tag{II.33}$$

the *specific entropy* of ν. Every σ-invariant probability measure ν on $\Gamma^{\mathbb{N}}$ can be uniquely extended to a σ-invariant probability measure ν on $\Gamma^{\mathbb{Z}}$. In terms of this extension, we have

$$\widehat{h}(\nu) = \mathbb{E}\left(h\big(\pi_{\{0\}}\nu(\mathcal{F}_{-\mathbb{N}})\big)\right). \tag{II.34}$$

Here, given $\nu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{Z}})$, we write $\pi_{\{0\}}\nu(\mathcal{F}_{-\mathbb{N}}) \in \mathfrak{M}_1(\Gamma)$ to denote the conditional probability law of the 0-th coordinate given the sigma-field $\mathcal{F}_{-\mathbb{N}}$ generated by the coordinates labelled with the past $-\mathbb{N}$, the expectation is w.r.t. this sigma-field, and h denotes entropy (see Petersen [C6] Chapter 5). Thus, $\widehat{h}(\nu)$ is the *average entropy of a single coordinate conditioned on its past.*

EXERCISE II.35 *Derive* (II.34) *from* (II.33).

Now, the key observation is that $\nu \mapsto \widehat{h}(\nu)$ is affine between any pair of *ergodic* elements of $\widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{Z}})$. Indeed, if ν_1, ν_2 with $\nu_1 \neq \nu_2$ are ergodic, then

$$\widehat{h}\big(\alpha\nu_1 + (1-\alpha)\nu_2\big) = \alpha\widehat{h}(\nu_1) + (1-\alpha)\widehat{h}(\nu_2) \qquad \forall\alpha \in [0,1],$$

because the past determines the ergodic component (as is evident from looking at ergodic averages over the past). It easily follows that this statement extends to *any* pair of elements of $\widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{Z}})$, because $\widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{Z}})$ is a Choquet simplex, i.e., all σ-invariant probability measures on $\Gamma^{\mathbb{Z}}$ can be written as convex combinations of ergodic probability measures (see Georgii [C3] Section 14.1).

(3) The representation in (II.32) can be traced back to the Shannon-McMillan-Breiman Theorem in ergodic theory, for which we refer the reader to Petersen [C6] Section 6.2.

(4) The large deviation property in Theorem II.23 is very powerful, because it implies all the earlier large deviation properties via the contraction principle, in the spirit of the argument in Section II.4. A drawback, however, is that the rate function I_ρ^∞ only comes implicitly as a supremum, which makes it somewhat hard to work with. For further background on the properties and the interpretation of I_ρ^∞, we refer the reader to Ellis [A5] Chapter IX.

(5) The proof of Theorem II.23 is an example of a proof via a so-called "projective limit": large deviations in an infinite-dimensional setting are derived by "lifting up" large deviations in finite dimensions. The projective limit approach, which in its general form is due to Dawson and Gärtner [D14], is a very powerful tool. Large deviation results in finite dimensions can be derived with the help of relatively simple combinatorics (as in Sections II.1 and II.2), while such arguments fail in an infinite-dimensional setting, or are cumbersome to carry through. See Dembo and Zeitouni [A2] Section 4.6 for more details.

II.7 Extension to countable state space

Before moving on to the general theory in Chapter III, we indicate how the condition of a finite state space assumed in (II.1) can be relaxed with the help of an approximation argument. This extension is obviously very important for applications. We focus on the empirical measure L_n, though the argument applies equally well to L_n^N or R_n.

Let ρ be a probability measure on $\mathbb{N}$ such that $\rho_s > 0$ for all $s \in \mathbb{N}$. Let (X_i) be i.i.d. with marginal law ρ, and consider the empirical measure $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, which is a random element of $\mathfrak{M}_1(\mathbb{N})$. On this space we again use the total variation distance $d(\mu,\nu) = \frac{1}{2}\sum_{s\in\mathbb{N}} |\mu_s - \nu_s|$. The following result generalizes Theorem II.2.

THEOREM II.36 *For $a > 0$, let $B_a(\rho) = \{\nu \in \mathfrak{M}_1(\mathbb{N})\colon d(\nu,\rho) \le a\}$, and define*

$$J(a) = \inf_{\nu \in B_a^c(\rho)} I_\rho(\nu),$$

with

$$I_\rho(\nu) = \sum_{s\in\mathbb{N}} \nu_s \log\left(\frac{\nu_s}{\rho_s}\right). \tag{II.37}$$

Then:
(a) $\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(L_n \in B_a^c(\rho)) \ge -J(a)$ *for all* $a > 0$.
(b) $\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(L_n \in B_a^c(\rho)) \le -J(a-)$ *for all* $a > 0$.

PROOF. Let $\pi_N\colon \mathbb{N} \to \{1,\ldots,N\}$ be the truncation defined by $\pi_N(s) = s \wedge N$, and write $\pi_N\nu = \nu \circ \pi_N^{-1}$. We will need the following three properties:

1. $0 \le d(\nu,\rho) - d(\pi_N\nu, \pi_N\rho) \le \sum_{s=N}^\infty \rho_s$ uniformly in $\nu \in \mathfrak{M}_1(\mathbb{N})$.
2. $a \mapsto J(a)$ is non-decreasing and right-continuous.
3. The sum defining $I_\rho(\nu)$ is convergent in $[0,\infty]$. The map $N \mapsto I_{\pi_N\rho}(\pi_N\nu)$ is non-decreasing for all $\nu \in \mathfrak{M}_1(\mathbb{N})$ and has limit $I_\rho(\nu)$.

EXERCISE II.38 *Prove properties 1, 2 and 3.*

(a) Lower bound:

By property 1, we have

$$\mathbb{P}\big(d(L_n,\rho) > a\big) \geq \mathbb{P}\Big(d(\pi_N L_n, \pi_N \rho) > a\Big) \qquad \forall N \geq 1.$$

Since $\pi_N L_n = \frac{1}{n}\sum_{i=1}^n \delta_{\pi_N(X_i)}$, we can apply Theorem II.2 to obtain

$$\lim_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\Big(d(\pi_N L_n, \pi_N\rho) > a\Big)$$

$$= - \inf_{\nu\in\mathfrak{M}_1(\mathbb{N}):\ d(\pi_N\nu,\pi_N\rho)>a} I_{\pi_N\rho}(\pi_N\nu) \qquad \forall N \geq 1.$$

By property 3, we have $I_{\pi_N\rho}(\pi_N\nu) \leq I_\rho(\nu)$ for all $N \geq 1$. Moreover, from property 1 we know that for all $\delta > 0$ there exists an $n_0 = n_0(\delta)$ such that

$$\big\{\nu \in \mathfrak{M}_1(\mathbb{N})\colon d(\pi_N\nu, \pi_N\rho) > a\big\} \supset \big\{\nu \in \mathfrak{M}_1(\mathbb{N})\colon d(\nu,\rho) > a+\delta\big\} \qquad \forall n \geq n_0.$$

Therefore

$$\inf_{\nu\in\mathfrak{M}_1(\mathbb{N}):\ d(\pi_N\nu,\pi_N\rho)>a} I_{\pi_N\rho}(\pi_N\nu) \leq \inf_{\nu\in\mathfrak{M}_1(\mathbb{N}):\ d(\nu,\rho)>a+\delta} I_\rho(\nu) = J(a+\delta),$$

and so

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(d(L_n,\rho) > a\big) \geq -J(a+\delta).$$

Let $\delta \downarrow 0$ and use property 2.

(b) Upper bound:

Fix $\delta > 0$. By property 1, there exists an $n_0 = n_0(\delta)$ such that

$$\mathbb{P}\big(d(L_n,\rho) > a\big) \leq \mathbb{P}\Big(d(\pi_N L_n, \pi_N\rho) > a - \delta\Big) \qquad \forall n \geq n_0.$$

Again using Theorem II.2, we have

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\Big(d(\pi_N L_n, \pi_N\rho) > a - \delta\Big)$$

$$\leq - \inf_{\nu\in\mathfrak{M}_1(\mathbb{N}):\ d(\pi_N\nu,\pi_N\rho)>a-\delta} I_{\pi_N\rho}(\pi_N\nu) \qquad \forall N \geq 1.$$

Next we show that

$$\inf_{\nu\in\mathfrak{M}_1(\mathbb{N}):\ d(\pi_N\nu,\pi_N\rho)>a-\delta} I_{\pi_N\rho}(\pi_N\nu) \geq J(a-\delta) \qquad \forall N \geq 1.$$

Indeed, fix $N \geq 1$ and pick any $\nu \in \mathfrak{M}_1(\mathbb{N})$ with $d(\pi_N\nu, \pi_N\rho) > a - \delta$. Set $\nu_{[N]} = \sum_{s=N}^\infty \nu_s$ and $\rho_{[N]} = \sum_{s=N}^\infty \rho_s$, and define $\widetilde{\nu} \in \mathfrak{M}_1(\mathbb{N})$ by

$$\widetilde{\nu}_s = \begin{cases} \nu_s & \text{if } s < N, \\ \frac{\nu_{[N]}}{\rho_{[N]}}\rho_s & \text{if } s \geq N. \end{cases}$$

Then an easy computation gives

$$\begin{aligned} \pi_N\widetilde{\nu} &= \pi_N\nu, \\ I_\rho(\widetilde{\nu}) &= I_{\pi_N\rho}(\pi_N\nu). \end{aligned}$$

Moreover, $d(\widetilde{\nu},\rho) = d(\pi_N\widetilde{\nu}, \pi_N\rho)$ and therefore $I_{\pi_N\rho}(\pi_N\nu) \geq J(a-\delta)$. Since ν was arbitrary, this proves the claim. Thus, we have now arrived at

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(d(L_n,\rho) > a\Big) \leq -J(a-\delta).$$

Let $\delta \downarrow 0$. □

Again, it is not a priori clear when and where J is continuous. (The trouble is that $B_a^c(\rho)$ may consist of disconnected components and that these components may vanish as a increases through certain particular values.) We have:

LEMMA II.39 *Let* $\Gamma = \mathbb{N}$.
(i) I_ρ *is lower semi-continuous and strictly convex on* $\mathfrak{M}_1(\mathbb{N})$.
(ii) I_ρ *has compact level sets.*
(iii) $I_\rho(\nu) \geq 0$ *with equality if and only if* $\nu = \rho$.

EXERCISE II.40 *Prove Lemma II.39.*

Thus, the rate function I_ρ on $\mathfrak{M}_1(\mathbb{N})$ looses some of the nice properties its counterpart for finite state space has (compare with Lemma II.4). In particular, we loose the everywhere finiteness and continuity. In Chapter III we will come to appreciate why.

The above limiting argument can be generalized much further. For instance, Theorem II.36 can be extended to $\mathfrak{M}_1(\mathbb{R})$, in which case the entropy function becomes

$$I_\rho(\nu) = H(\nu|\rho) = \begin{cases} \int_{\mathbb{R}} d\nu \log \frac{d\nu}{d\rho} & \text{if } \nu \ll \rho, \\ \infty & \text{otherwise,} \end{cases} \tag{II.41}$$

where $\nu \ll \rho$ means that ν is absolutely continuous w.r.t. ρ. An extension of this type will be needed in Chapter X. For more background we refer the reader to Deuschel and Stroock [A3] Section 3.2, Dembo and Zeitouni [A2] Section 6.2.

CHAPTER III

GENERAL THEORY

In this chapter we leave the i.i.d. setting and present a general theory unifying and extending the results that were derived in the preceding chapters. This theory was first formulated in the right degree of abstraction by Varadhan [D61]. In Section III.1 we begin with the basic definitions. In Section III.2 we comment on the significance of these definitions.

III.1 The large deviation principle (LDP)

Let $\mathcal{X}$ be a Polish space with distance $d\colon \mathcal{X} \times \mathcal{X} \to [0, \infty)$. Recall the following definition.

DEFINITION III.1 *$f\colon \mathcal{X} \to [-\infty, \infty]$ is* **lower semi-continuous** *if it satisfies any of the following equivalent properties (see Fig. 6):*
(i) $\liminf_{n\to\infty} f(x_n) \geq f(x)$ *for all* (x_n), x *such that* $x_n \to x$ *in* $\mathcal{X}$.
(ii) $\lim_{\epsilon\downarrow 0} \inf_{y\in B_\epsilon(x)} f(y) = f(x)$ *with* $B_\epsilon(x) = \{y \in \mathcal{X}\colon\ d(x,y) < \epsilon\}$.
(iii) f *has closed level sets, i.e.,* $f^{-1}([-\infty, c]) = \{x \in \mathcal{X}\colon\ f(x) \leq c\}$ *is closed for all* $c \in \mathbb{R}$.

EXERCISE III.2 *Prove the equivalence of (i), (ii) and (iii) in Definition* III.1.

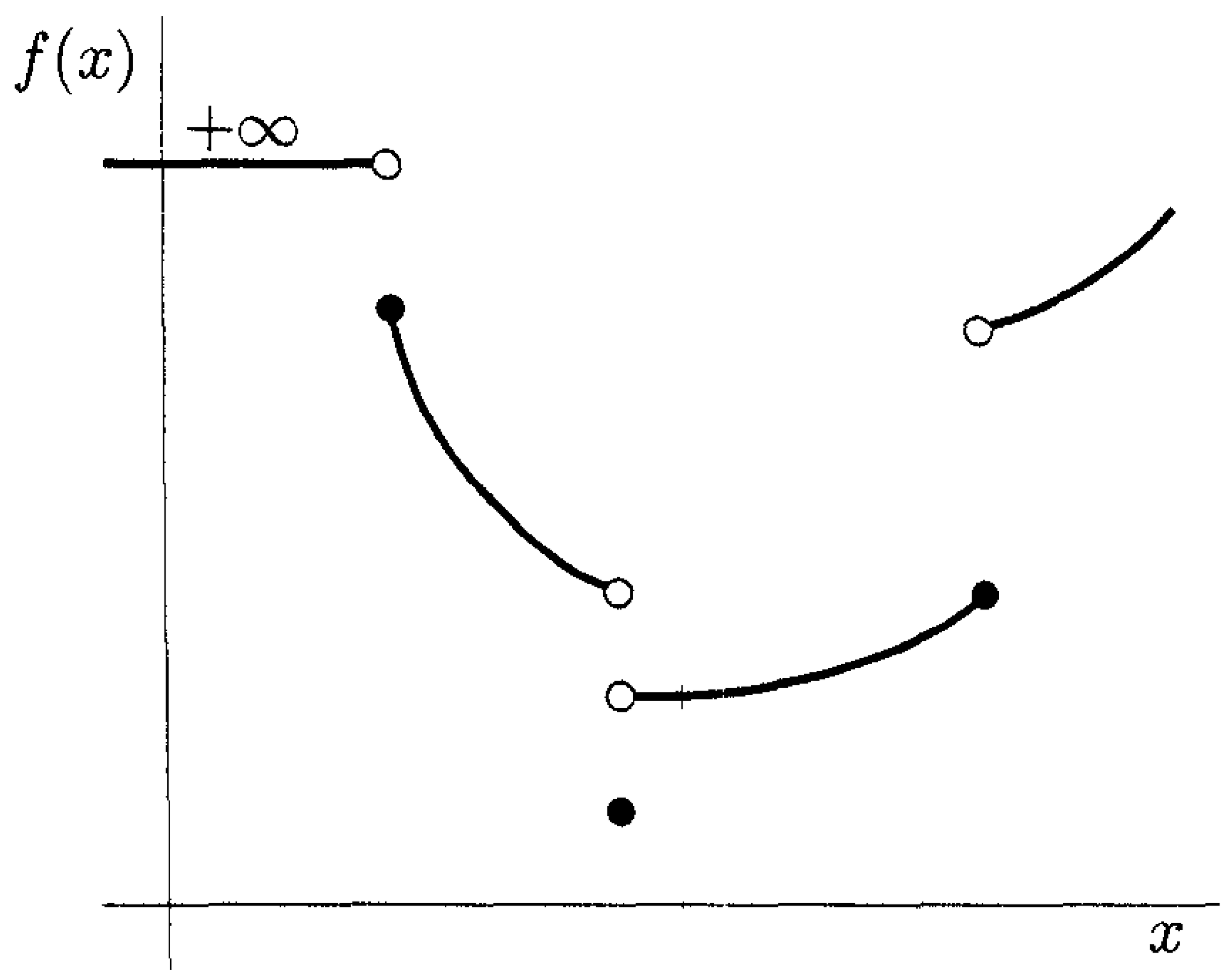

FIG. 6. A lower semi-continuous function

We will frequently need the following important fact:

LEMMA III.3 *A lower semi-continuous function attains a minimum on every non-empty compact set.*

EXERCISE III.4 *Prove Lemma* III.3.

Here are the key definitions of large deviation theory.

DEFINITION III.5 *The function* $I\colon \mathcal{X} \to [0,\infty]$ *is called a* **rate function** *if*
(D1) $I \not\equiv \infty$.
(D2) *I is lower semi-continuous.*
(D3) *I has compact level sets.*

DEFINITION III.6 *A sequence of probability measures* (P_n) *on* $\mathcal{X}$ *is said to satisfy the* **large deviation principle** (LDP) *with rate n and with rate function I if*
(D1′) *I is a rate function in the sense of Definition* III.5.
(D2′) $\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \le -I(C) \quad \forall C \subset \mathcal{X}$ *closed.*
(D3′) $\liminf_{n\to\infty} \frac{1}{n} \log P_n(O) \ge -I(O) \quad \forall O \subset \mathcal{X}$ *open.*
Here the bounds are in terms of the set function defined by

$$I(S) = \inf_{x\in S} I(x), \qquad S \subset \mathcal{X}. \tag{III.7}$$

The goal of large deviation theory is to build up an arsenal of theorems based on these two definitions. Some of these theorems will be described in the rest of this chapter. In order to apply them to concrete situations, one must of course verify that the triple $(\mathcal{X}, (P_n), I)$ one is working with satisfies the LDP.

In what follows we write, for $S \subset \mathcal{X}$,

$$\begin{aligned} \mathrm{cl}(S) &= \text{the closure of } S \\ \mathrm{int}(S) &= \text{the interior of } S. \end{aligned}$$

The abstract setting of Definitions III.5 and III.6 raises the natural question of uniqueness of the rate function. This question is settled by the following theorem taken from Ellis [A5] Section II.3.

THEOREM III.8 *Let* (P_n) *satisfy the* LDP. *Then the associated rate function I is unique.*

PROOF. Let I and J be two rate functions for (P_n). We show that $I(x) = J(x)$ for all $x \in \mathcal{X}$. Fix $x \in \mathcal{X}$ and consider the sequence of open balls $B_N = B_{1/N}(x)$ of radius $1/N$, $N \in \mathbb{N}$. Then, by Definition III.6(D2′–D3′),

$$\begin{aligned} -I(x) &\le -I(B_{N+1}) \le \liminf_{n\to\infty} \tfrac{1}{n} \log P_n(B_{N+1}) \\ &\le \limsup_{n\to\infty} \tfrac{1}{n} \log P_n\big(\mathrm{cl}(B_{N+1})\big) \le -J\big(\mathrm{cl}(B_{N+1})\big) \le -J(B_N), \end{aligned}$$

where the first inequality holds because $x \in B_{N+1}$ and the last inequality holds because $B_N \supset \mathrm{cl}(B_{N+1})$. Let $N \to \infty$ and use the lower semi-continuity of J by Definition III.5(D2), which implies that $\lim_{N\to\infty} J(B_N) = J(x)$ by Definition III.1(ii). Then we get $I(x) \ge J(x)$. The opposite inequality follows from symmetry. □

EXERCISE III.9 *For each of the following examples determine whether the* LDP *is satisfied with rate n and, if so, with what rate function:*
1. $\mathcal{X} = \mathbb{R}$, P_n *uniform on* $[-n, n]$.
2. $\mathcal{X} = \mathbb{R}$, P_n *uniform on* $[-\frac{1}{n}, \frac{1}{n}]$.
3. $\mathcal{X} = [-1, 1]$, P_n *uniform on* $[-1, 1]$.

EXERCISE III.10 *(Suggested by G. O'Brien.) Let* Z_n *be a single random variable with a binomial distribution with parameters n and* p_n. *Let* $P_n(\cdot) = P(Z_n/np_n \in \cdot\,)$. *Show that if* $\lim_{n\to\infty} p_n = 0$ *and* $\lim_{n\to\infty} np_n = \infty$, *then* (P_n) *satisfies the* LDP *on* $\mathbb{R}$ *with rate* np_n *and with rate function given by* $I(z) = z\log z - z + 1$, $z \ge 0$, *and* $I(z) = \infty$, $z < 0$. *Does the answer ring a bell? (Recall Exercise I.11.)*

III.2 Comments

The following comments help to explain the definitions in Section III.1:

(1) In Definition III.6 it is crucial to make a difference between open and closed sets. Naively, one might try to replace (D2′) and (D3′) by the stronger requirement that

$$\lim_{n\to\infty} \frac{1}{n} \log P_n(S) = -I(S) \quad \forall S \subset \mathcal{X} \text{ Borel.} \tag{III.11}$$

However, this would be far too restrictive: many examples that satisfy (D2′) and (D3′) do not satisfy (III.11). For instance, P_n might be non-atomic for all n. In that case $P_n(\{x\}) = 0$ for all n and all $x \in \mathcal{X}$, so by picking $S = \{x\}$ we would find that (III.11) could only be true with $I \equiv \infty$, which is excluded by (D1). Still, we will see that (D2′) and (D3′) are enough to build up a theory because the two inequalities can be manipulated together.

(2) We say that a set $S \subset \mathcal{X}$ is **I-continuous** if

$$I\big(\text{int}\,(S)\big) = I\big(\text{cl}(S)\big).$$

Clearly, the LDP implies that (III.11) holds for all I-continuous sets. In many examples this is a large class. For instance, if I is continuous, then all S satisfying $S \subset \text{cl}\big(\text{int}\,(S)\big)$ are I-continuous, which includes all open sets.

EXERCISE III.12 *Identify the I-continuous sets for the rate function in Theorem I.3.*

(3) The role of open and closed sets in the LDP is similar to their role in weak convergence of probability measures: (P_n) is said to converge weakly to P if

(D2″) $\limsup_{n\to\infty} P_n(C) \leq P(C) \quad \forall C \subset \mathcal{X}$ closed.
(D3″) $\liminf_{n\to\infty} P_n(O) \geq P(O) \quad \forall O \subset \mathcal{X}$ open.

One can therefore view (D2′) and (D3′) in Definition III.6 as analogues of weak convergence on an exponential scale.

(4) Since (D2″) and (D3″) are equivalent to

$$\int_{\mathcal{X}} F(x) P_n(dx) \underset{n\to\infty}{\longrightarrow} \int_{\mathcal{X}} F(x) P(dx) \qquad \forall F \in C_b(\mathcal{X}),$$

with $C_b(\mathcal{X})$ the space of bounded continuous functions on $\mathcal{X}$, it is intuitively clear that the LDP is ideally suited for handling convergence of integrals of exponential functionals. This intuition will be worked out in Section III.3. The analogy of the LDP with weak convergence has been explored in detail by O'Brien and Vervaat [D46] and O'Brien [D45]. See also the monograph by Dupuis and Ellis [A4], which is centered around this analogy.

(5) In Definition III.5 it is a matter of tradition to include (D2), even though (D3) implies (D2) by the equivalence stated in Definition III.1. The role of (D3) is to guarantee that the family (P_n) is **exponentially tight**, i.e.,

$$\forall M < \infty\ \exists K_M \subset \mathcal{X} \text{ compact:} \quad \limsup_{n\to\infty} \frac{1}{n} \log P_n(\mathcal{X} \setminus K_M) \leq -M$$

(see Dembo and Zeitouni [A2] Exercise 4.1.10). This property is the analogue of tightness in weak convergence. In Section III.6 we will describe a version of the LDP in which (D3) is relaxed.

(6) The LDP implies that

$$\inf_{x\in\mathcal{X}} I(x) = I(\mathcal{X}) = 0,$$

because $P_n(\mathcal{X}) = 1$ for all n and $\mathcal{X}$ is closed. Moreover, by Lemma III.3 in combination with Definition III.5, there is an $x \in \mathcal{X}$ such that $I(x) = 0$. In many examples this zero is unique and corresponds to an underlying SLLN, but there are examples where it is not unique (see e.g. Chapter VII).

(7) It is possible to set up Definitions III.5 and III.6 in the framework of an arbitrary topological space. We will, however, not insist on this degree of generality and refer the reader to Dembo and Zeitouni [A2] Section 1.2. Without the structure of a Polish space the theory tends to become more cumbersome and many results in the Polish space setting fail to carry over. Conversely, as more structure is added to $\mathcal{X}$, stronger results can be obtained. See e.g. Section III.7.

III.3 Varadhan's Lemma

We are now ready to formulate the first important general theorem of large deviation theory, which is due to Varadhan [D61]. This theorem is a far-reaching generalization of the Laplace method that was used in Sections I.5 and II.3. Though the result comes as a theorem, it is commonly referred to as "Varadhan's Lemma" in the literature. It will be used frequently in Part B.

THEOREM III.13 **(Varadhan's Lemma)** *Let (P_n) satisfy the* LDP *on $\mathcal{X}$ with rate n and with rate function I. Let $F\colon \mathcal{X} \to \mathbb{R}$ be a continuous function that is bounded from above. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} e^{nF(x)} P_n(dx) = \sup_{x\in\mathcal{X}} \big[F(x) - I(x)\big]. \tag{III.14}$$

PROOF. Let

$$J_n(S) = \int_S e^{nF(x)} P_n(dx), \qquad S \subset \mathcal{X} \text{ Borel},$$

and put

$$a = \sup_{x\in\mathcal{X}} F(x), \quad b = \sup_{x\in\mathcal{X}} \big[F(x) - I(x)\big].$$

Note that $-\infty < b \le a < \infty$, because $I \ge 0$ and F is continuous and bounded from above. We proceed by proving upper and lower bounds. We will repeatedly make use of (I.2).

Upper bound:

We slice the space $\mathcal{X}$ according to the values of F. Let $C = F^{-1}([b,a])$, and for $N \in \mathbb{N}$ define the sets

$$C_j^N = F^{-1}\big([c_{j-1}^N, c_j^N]\big), \qquad j = 1, \ldots, N,$$

where $c_j^N = b + \frac{j}{N}(a-b)$ for $j = 0, 1, \ldots, N$. It is clear that

$$C = \bigcup_{j=1}^{N} C_j^N.$$

Now, F being continuous, all C_j^N are closed. Hence it follows from the LDP (see Definition III.6(D2')) that

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C_j^N) \le -I(C_j^N) \quad \forall j.$$

If we take into account that $F(x) \le c_j^N$ on C_j^N, then from (I.2) we obtain the inequality

$$\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) \le \max_{1\le j\le N} \left[c_j^N - I(C_j^N)\right].$$

This can be further developed by inserting the inequality $c_j^N \le \inf_{x\in C_j^N} F(x) + \frac{1}{N}(a-b)$, to get

$$\begin{aligned}
\limsup_{n\to\infty} \tfrac{1}{n} \log J_n(C) &\le \max_{1\le j\le N} \Big\{ \inf_{x\in C_j^N} F(x) - \inf_{x\in C_j^N} I(x) \Big\} + \tfrac{1}{N}(a-b) \\
&\le \max_{1\le j\le N} \sup_{x\in C_j^N} [F(x) - I(x)] + \tfrac{1}{N}(a-b) \\
&= \sup_{x\in C} \big[F(x) - I(x)\big] + \tfrac{1}{N}(a-b) \\
&\le b + \tfrac{1}{N}(a-b).
\end{aligned}$$

Letting $N\to\infty$, we get $\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) \le b$. To extend this upper bound to $J_n(\mathcal{X})$, we make use of the trivial estimate $J_n(\mathcal{X}\setminus C) \le e^{nb}$. Via one more application of (I.2) this leads to

$$\limsup_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) \le b.$$

Lower bound:

Pick $x \in \mathcal{X}$ and $\epsilon > 0$ arbitrary. Then the set

$$O_{x,\epsilon} = \{y \in \mathcal{X} \colon\ F(y) > F(x) - \epsilon\}$$

is an open neighborhood of x by the continuity of F. It follows from the LDP (see Definition III.6(D3')) that

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(O_{x,\epsilon}) \ge -I(O_{x,\epsilon}).$$

Since $I(O_{x,\epsilon}) \le I(x)$, this estimate gives us

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n(O_{x,\epsilon}) \ge F(x) - \epsilon - I(x).$$

Now use that $J_n(\mathcal{X}) \ge J_n(O_{x,\epsilon})$, let $\epsilon \downarrow 0$ and afterwards take the supremum over $x \in \mathcal{X}$, to find

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) \ge b.$$

□

Bryc [D8] has proved an *inverse* of Varadhan's Lemma, which reads as follows. Let

$$\Lambda_n(F) = \frac{1}{n} \log \int_{\mathcal{X}} e^{nF(x)} P_n(dx), \qquad F \in C_b(\mathcal{X}).$$

If (P_n) is exponentially tight and $\lim_{n\to\infty} \Lambda_n(F) = \Lambda(F) \in \mathbb{R}$ exists for all $F \in C_b(\mathcal{X})$, then (P_n) satisfies the LDP with rate n and with rate function I given by

$$I(x) = \sup_{F\in C_b(\mathcal{X})} [F(x) - \Lambda(F)], \qquad x \in \mathcal{X}. \tag{III.15}$$

This result is conceptually interesting because it fortifies the earlier observed analogy of the LDP with weak convergence (see Comments (3) and (4) in Section III.2). However, it is also of practical use, since the existence of $\Lambda(F)$ need only to be verified for a sufficiently rich subclass of $C_b(\mathcal{X})$, which is why this result can sometimes serve as a route to establishing the LDP. (See O'Brien and Vervaat [D46] for an extended version of Bryc's Theorem.)

Equation (III.15) is the inverse of the relation

$$\Lambda(F) = \sup_{x\in\mathcal{X}}[F(x) - I(x)], \qquad F \in C_b(\mathcal{X}), \tag{III.16}$$

appearing in (III.14). It is tempting to compare this duality with the one mentioned in Comment (6) in Section I.4. However, the rate function I in the present general context need *not* be convex.

The continuity and boundedness requirements on F in Theorem III.10 can be weakened a little. We refer the reader to the literature (see e.g. Deuschel and Stroock [A3] Section 2.1, Dembo and Zeitouni [A2] Section 4.3).

III.4 The LDP for integrals of exponential functionals

In this section we give an alternative version of Varadhan's Lemma, one that allows us to *generate one* LDP *from another via tilting* (see Ellis [A5] Section II.7). We will have occasion to use this version in Chapters IV and X.

THEOREM III.17 **(Tilted LDP)** *Let (P_n) satisfy the* LDP *on $\mathcal{X}$ with rate n and with rate function I. Let $F\colon \mathcal{X} \to \mathbb{R}$ be a continuous function that is bounded from above. Define*

$$J_n(S) = \int_S e^{nF(x)} P_n(dx), \qquad S \subset \mathcal{X} \text{ Borel.}$$

Then the sequence (P_n^F) of probability measures defined by

$$P_n^F(S) = \frac{J_n(S)}{J_n(\mathcal{X})}, \qquad S \subset \mathcal{X} \text{ Borel,}$$

satisfies the LDP *on $\mathcal{X}$ with rate n and with rate function*

$$I^F(x) = \sup_{y\in\mathcal{X}}\big[F(y) - I(y)\big] - \big[F(x) - I(x)\big]. \tag{III.18}$$

PROOF. Since we know from Theorem III.13 that the asymptotics of $J_n(\mathcal{X})$ corresponds exactly to the term $\sup_{y\in\mathcal{X}}\big[F(y) - I(y)\big]$, it suffices to show that

$$\limsup_{n\to\infty} \tfrac{1}{n}\log J_n(C) \;\le\; b(C) \quad \forall C \subset \mathcal{X} \text{ closed,}$$

$$\liminf_{n\to\infty} \tfrac{1}{n}\log J_n(O) \;\ge\; b(O) \quad \forall O \subset \mathcal{X} \text{ open,}$$

with

$$b(S) = \sup_{x\in S}[F(x) - I(x)], \qquad S \subset \mathcal{X}.$$

The proof of these two inequalities follows the same line of argument as in the proof of Theorem III.13. This is left for the reader to verify. □

EXERCISE III.19 *Check that I^{F} is a rate function in the sense of Definition* III.5.

III.5 The Contraction Principle

In this section we formulate a theorem that enables us to *generate one* LDP *from another via contraction.* This theorem, besides being conceptually important, will turn out to be very useful later on (see e.g. Sections IV.2 and IV.4). The contraction principle encountered in Section II.4 is a special case. For a broader perpective and for alternative versions, the reader is referred to Dembo and Zeitouni [A2] Section 4.2.

THEOREM III.20 **(Contraction Principle)** *Let (P_n) be a sequence of probability measures on a Polish space $\mathcal{X}$ that satisfies the* LDP *with rate n and with rate function I. Let*

$$\begin{cases} \mathcal{Y} & \text{be a Polish space,} \\ T\colon \mathcal{X} \to \mathcal{Y} & \text{a continuous map,} \\ Q_n = P_n \circ T^{-1} & \text{an image probability measure.} \end{cases}$$

Then (Q_n) satisfies the LDP *on $\mathcal{Y}$ with rate n and with rate function J given by*

$$J(y) = \inf_{x \in \mathcal{X}\colon\, T(x)=y} I(x), \tag{III.21}$$

with the convention $\inf_{\emptyset} I = \infty$.

PROOF. Since T is continuous, T^{-1} maps open sets into open sets and closed sets into closed sets. Since (P_n) satisfies Definition III.6(D2′–D3′) with rate function I, it follows that (Q_n) does so too with rate function J. Indeed, pick $C \subset \mathcal{Y}$ closed and write

$$\begin{aligned} \limsup_{n\to\infty} \tfrac{1}{n} \log Q_n(C) &= \limsup_{n\to\infty} \tfrac{1}{n} \log P_n\big(T^{-1}(C)\big) \leq -I\big(T^{-1}(C)\big) \\ &= -\inf_{x\in T^{-1}(C)} I(x) = -\inf_{y\in C} \inf_{x\in T^{-1}(\{y\})} I(x) = -\inf_{y\in C} J(y) = -J(C). \end{aligned}$$

A similar argument works for $O \subset \mathcal{Y}$ open. Hence it remains to prove that J is a rate function in the sense of Definition III.5.

Clearly, $\mathcal{D}_I = \{x \in \mathcal{X}\colon I(x) < \infty\} \neq \emptyset$ implies $\mathcal{D}_J = \{y \in \mathcal{Y}\colon J(y) < \infty\} \neq \emptyset$, proving that Definition III.5(D1) carries over. Since I has compact level sets, and since a continuous image of a compact set is again compact, also J has compact level sets, proving that Definition III.5(D3) carries over. Hence J is a rate function (recall Comment (5) in Section III.2) and the proof is complete. □

The contraction principle encountered in Section II.4 is a special case of Theorem III.20 with:

$$\mathcal{X} = \mathfrak{M}_1(\Gamma), \qquad \mathcal{Y} = \mathbb{R}, \qquad T(\nu) = \textstyle\sum_s s\nu_s,$$

$$P_n(\cdot) = \mathbb{P}(L_n \in \cdot), \qquad Q_n(\cdot) = \mathbb{P}(\tfrac{1}{n} S_n \in \cdot).$$

III.6 The weak LDP

There is a version of the LDP in which some of the conditions in Definitions III.5 and III.6 are relaxed. We will need this version in Section III.7.

DEFINITION III.22 *The function* $I\colon \mathcal{X} \to [0,\infty]$ *is called a* **weak rate function** *if*
(D1) $I \not\equiv \infty$.
(D2) *I is lower semi-continuous.*
(D3*) *I has closed level sets.*

DEFINITION III.23 *A sequence of probability measures* (P_n) *on* $\mathcal{X}$ *is said to satisfy the* **weak large deviation principle** *(weak* LDP*) with rate n and with weak rate function I if*
(D1′*) *I is a weak rate function in the sense of Definition* III.22.
(D2′*) $\limsup_{n\to\infty} \frac{1}{n} \log P_n(K) \le -I(K) \quad \forall K \subset \mathcal{X}$ *compact.*
(D3′) $\liminf_{n\to\infty} \frac{1}{n} \log P_n(O) \ge -I(O) \quad \forall O \subset \mathcal{X}$ *open.*

Thus, the compact level sets and the upper bound for non-compact closed sets are sacrificed. In the weak LDP the exponential tightness of (P_n) is missing (recall Comment (5) in Section III.2). In fact:

The weak LDP *and exponential tightness together imply the* LDP.

(See Dembo and Zeitouni [A2] Section 1.2.)

REMARKS III.24
1. By Definition III.1, (D2) and (D3*) are equivalent. It is again a matter of tradition to include both (compare with Comment (5) in Section III.2).
2. O'Brien [D45] has shown that also for the weak LDP the associated rate function is unique (recall Theorem III.8). Of course, the proof of Theorem III.8 carries over when $\mathcal{X}$ is locally compact.
3. Neither Varadhan's Lemma nor the Contraction Principle in general applies with only the weak LDP. This is evident from the proofs, because the inverse under a continuous map of a compact set need not be compact.

The last observation shows that the weak LDP is "limping". In practice, however, there are many concrete situations where one only has a weak LDP but manages to use it after doing some extra work. We will encounter examples in Chapters VII and VIII.

The nomenclature in the literature varies. Some authors reserve the names "good rate function" and "full LDP" for Definitions III.5 and III.6, and simply use "rate function" and "LDP" for Definitions III.22 and III.23.

III.7 Convexity

In Chapters I and II various rate functions were calculated explicitly. It is an important fact that all these rate functions are *convex*. It should be noted, however, that there is no a priori reason why rate functions should be convex. In fact, many are not and it suffices to glance at Theorem III.17 to see why. In this section we discuss how convexity arises in a general context when we are dealing with *additive functionals* of i.i.d. sequences. The argument below follows Bahadur and Zabell [D3]. See also Stroock [A7] Section 3.

Throughout this section we assume that the Polish space $\mathcal{X}$ has the following structural properties:

$$
\begin{array}{ll}
(1) & \mathcal{X} \text{ is a convex subset of a linear space.} \\
(2) & \mathcal{X} \text{ is locally convex (in the topology induced by } d). \\
(3) & \lim_{\beta\to\alpha} d\big(\alpha x+(1-\alpha)y, \beta x+(1-\beta)y\big)=0 \qquad \forall x,y\in\mathcal{X}\ \forall\alpha\in(0,1). \\
(4) & \text{In } \mathcal{X}\text{, the closed convex hull of a compact set is compact.}
\end{array}
\tag{III.25}
$$

REMARKS III.26
1. Local convexity means that every open set is the union of open convex sets.
2. The convex hull of a set is the intersection of all the convex sets containing it.
3. Property (III.25)(4) holds when d satisfies a certain convexity condition. The reader is referred to Deuschel and Stroock [A3] Section 3.1, Dembo and Zeitouni [A2] Section 6.1.

We are interested in establishing the LDP for the empirical average of an i.i.d. sequence (Y_i) on a probability space $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P})$, where $\mathcal{B}(\mathcal{X})$ is the Borel sigma-field on $\mathcal{X}$.

THEOREM III.27 *Let (Y_i) be i.i.d. $\mathcal{X}$-valued random variables, with $\mathcal{X}$ a Polish space satisfying* (III.25). *For $S\subset\mathcal{X}$ Borel, let $P_n(S)=\mathbb{P}(\frac{1}{n}\sum_{i=1}^n Y_i\in S)$.*
(a) *For every $A\subset\mathcal{X}$ open and convex*

$$\lim_{n\to\infty}\frac{1}{n}\log P_n(A) = -\widehat{I}(A)$$

exists and is finite if and only if $P_n(A)>0$ for some n.
(b) *The function $I\colon\ \mathcal{X}\to[0,\infty]$ defined by*

$$I(x)=\sup_{\substack{A\ni x\\ A \text{ open convex}}} \widehat{I}(A), \qquad x\in\mathcal{X},$$

is a weak rate function.
(c) *I is convex.*

PROOF. (a) If A is convex, then

$$
\begin{aligned}
P_{n+m}(A) &= \mathbb{P}\left(\tfrac{n}{n+m}\Big[\tfrac{1}{n}\sum_{i=1}^{n}Y_i\Big]+\tfrac{m}{n+m}\Big[\tfrac{1}{m}\sum_{i=n+1}^{n+m}Y_i\Big]\in A\right)\\
&\geq \mathbb{P}\Big(\tfrac{1}{n}\sum_{i=1}^{n}Y_i\in A,\ \tfrac{1}{m}\sum_{i=n+1}^{n+m}Y_i\in A\Big)\\
&= P_n(A)P_m(A) \qquad \forall m,n\in\mathbb{N}
\end{aligned}
$$

(here we use assumption (III.25)(1)). Consequently, the sequence (a_n) with $a_n = -\log P_n(A)$ is subadditive, i.e., $a_{n+m}\leq a_n+a_m$ $\forall m,n\in\mathbb{N}$. If $P_n(A)=0$ for all n, then the claim is trivially true with $\widehat{I}(A)=\infty$. Below we will show that if A is open and convex, then

$$P_n(A)>0 \text{ for some } n \qquad \Longrightarrow \qquad P_n(A)>0 \text{ for } n \text{ large enough.} \tag{III.28}$$

This fact allows us to apply the standard limit theorem for subadditive sequences:

LEMMA III.29 *If (a_n) is a non-negative subadditive sequence with $a_n < \infty$ for n large enough, then*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\in\mathbb{N}} \frac{a_n}{n} \in [0,\infty).$$

PROOF. See e.g. Madras and Slade [C5] Lemma 1.2.2. □

When applied to $a_n = -\log P_n(A)$, Lemma III.29 in combination with (III.28) gives the claim in part (a) of the theorem. The limit $\lim_{n\to\infty} \frac{a_n}{n}$ is what we have called $\widehat{I}(A)$.

The proof of (III.28) uses a perturbation argument showing that if A is open and convex, then

$$\text{for } n \text{ large enough:} \qquad P_n(A) > 0 \Longrightarrow P_{n+1}(A) > 0.$$

This argument goes as follows. Suppose that $P_n(A) > 0$. Then, because $\mathcal{X}$ is Polish, there exists a compact set $K \subset A$ such that $P_n(K) > 0$. For $\epsilon > 0$, let $O_\epsilon = \cup_{x\in K} B_\epsilon(x)$, with $B_\epsilon(x)$ the open ball of radius ϵ around x. For n large enough (depending on ϵ, K) we have $O_\epsilon \supset \frac{n}{n+1}K$, and so

$$\mathbb{P}\Big(\frac{1}{n+1}\sum_{i=1}^{n} Y_i \in O_\epsilon\Big) \geq \mathbb{P}\Big(\frac{1}{n+1}\sum_{i=1}^{n} Y_i \in \frac{n}{n+1}K\Big) = P_n(K) > 0.$$

Moreover, for n large enough (depending on ϵ) we have $\mathbb{P}(\frac{1}{n+1}Y_{n+1} \in B_\epsilon(0)) > 0$. Therefore

$$\begin{aligned} P_{n+1}(O_{2\epsilon}) &= \mathbb{P}\Big(\tfrac{1}{n+1}\sum_{i=1}^{n+1} Y_i \in O_{2\epsilon}\Big) \\ &\geq \mathbb{P}\Big(\tfrac{1}{n+1}\sum_{i=1}^{n} Y_i \in O_\epsilon,\ \tfrac{1}{n+1}Y_{n+1} \in B_\epsilon(0)\Big) \\ &= \mathbb{P}\Big(\tfrac{1}{n+1}\sum_{i=1}^{n} Y_i \in O_\epsilon\Big)\mathbb{P}\Big(\tfrac{1}{n+1}Y_{n+1} \in B_\epsilon(0)\Big) > 0. \end{aligned}$$

However, for ϵ small enough we have $O_{2\epsilon} \subset A$. Hence, we get $P_{n+1}(A) > 0$. This completes the proof of (III.28).

(b) We must show that I satisfies the conditions in Definition III.22. Obviously, $I \geq 0$. Pick $0 \leq c < \infty$. If $I(x) > c$ for some $x \in \mathcal{X}$, then $\widehat{I}(A) > c$ for some open convex $A \ni x$ (by the definition of I). But $\inf_{y\in A} I(y) \geq \widehat{I}(A)$, and therefore $I(y) > c$ for y in some open convex neighborhood of x. Consequently, the set $\{x \in \mathcal{X}\colon I(x) > c\}$ is open, i.e., I has closed level sets. By the equivalence in Definition III.1, I is therefore lower semi-continuous. Thus, it remains to show that $I \not\equiv \infty$. This goes as follows.

Because $\mathcal{X}$ is Polish, there exists a compact set K such that $\mathbb{P}(Y_1 \in K) = P_1(K) > 0$. Let $\bar{K}$ be the closed convex hull of K. Then, by assumption (III.25)(4), $\bar{K}$ is compact. The convexity of $\bar{K}$ yields $P_n(\bar{K}) \geq P_1(\bar{K})^n$, and hence

$$\liminf_{n\to\infty} \frac{1}{n}\log P_n(\bar{K}) \geq -C(\bar{K}) \qquad \text{with} \qquad C(\bar{K}) = -\log P_1(\bar{K}) < \infty.$$

For all $\epsilon > 0$, the open covering $\cup_{x\in \bar{K}} B_\epsilon(x)$ of $\bar{K}$ has a finite subcovering $\cup_{i=1}^N B_\epsilon(x_i)$ with $N = N(\epsilon, K)$. Using (I.2) and part (a) of the theorem, we have

$$\begin{aligned}\limsup_{n\to\infty} \tfrac{1}{n} \log P_n\Big(\bigcup_{i=1}^N B_\epsilon(x_i)\Big) &\le \limsup_{n\to\infty} \tfrac{1}{n} \log \sum_{i=1}^N P_n(B_\epsilon(x_i)) \\ &= -\min_{i=1,\dots,N} \widehat{I}(B_\epsilon(x_i)).\end{aligned}$$

Therefore we get

$$\min_{i=1,\dots,N} \widehat{I}(B_\epsilon(x_i)) \le C(\bar{K}).$$

Consequently, for all $\epsilon > 0$ there exists a $x_\epsilon \in \bar{K}$ such that $\widehat{I}(B_\epsilon(x_\epsilon)) \le C(\bar{K})$. The sequence $(x_\epsilon)_{\epsilon>0}$ has a convergent subsequence with limit $\bar{x} \in \bar{K}$. For every $\delta > 0$, along this subsequence we eventually have $B_\epsilon(x_\epsilon) \subset B_\delta(\bar{x})$, and so we conclude that

$$\widehat{I}(B_\delta(\bar{x})) \le C(\bar{K}) \qquad \forall \delta > 0.$$

But $I(\bar{x}) = \sup_{\delta>0} \widehat{I}(B_\delta(\bar{x}))$, by assumption (III.25)(2) and the definition of I, and so $I(\bar{x}) \le C(\bar{K})$. Thus, we indeed have $I \not\equiv \infty$.
(c) Pick $x_1, x_2 \in \mathcal{X}$ such that $x_1 \ne x_2$. Pick open convex sets $A_1 \ni x_1$, $A_2 \ni x_2$, and put $x_{1,2} = \frac{1}{2}(x_1 + x_2)$ and $A_{1,2} = \frac{1}{2}(A_1 + A_2) \ni x_{1,2}$ (see Fig. 7; the dots are the midpoints of the line pieces connecting the corners of A_1 and A_2).

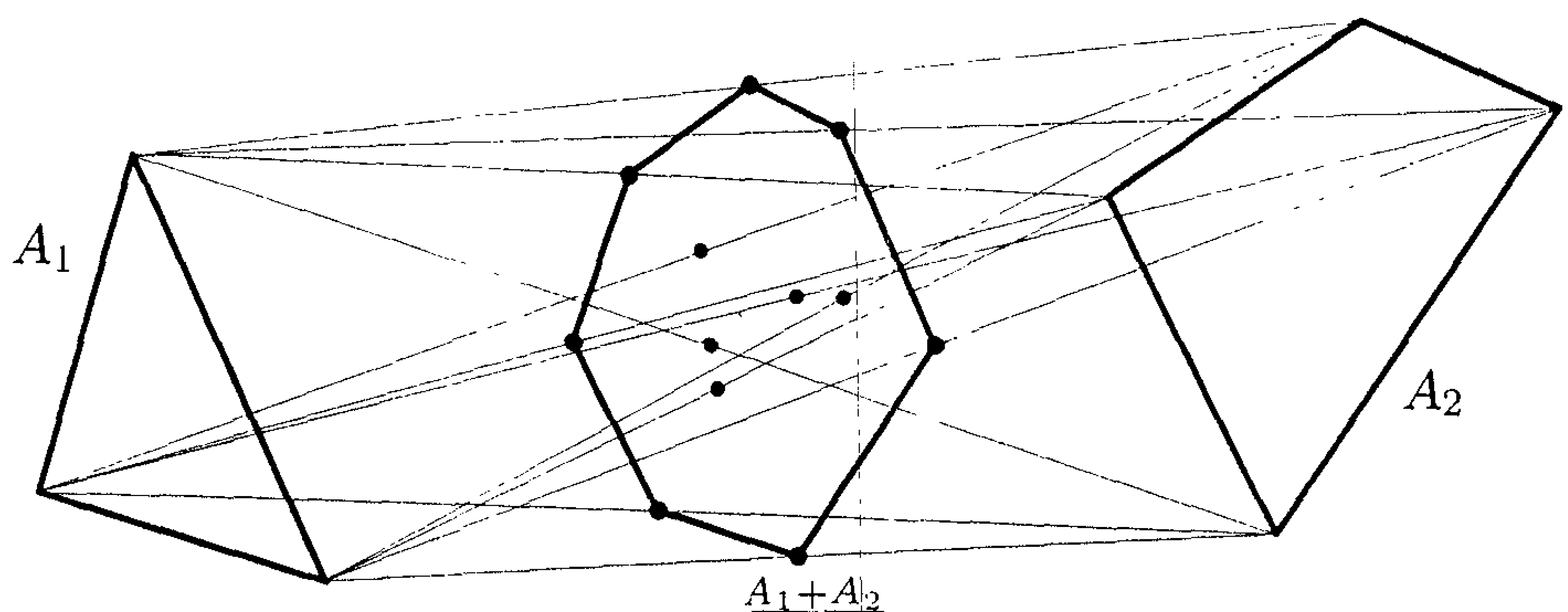

FIG. 7. An example of $(A_1, A_2) \mapsto \frac{1}{2}(A_1 + A_2)$

Clearly, $A_{1,2}$ is an open convex set, and

$$\begin{aligned}P_{2n}(A_{1,2}) &= \mathbb{P}\Big(\tfrac{1}{2n}\sum_{i=1}^{2n} Y_i \in A_{1,2}\Big) \\ &\ge \mathbb{P}\Big(\tfrac{1}{n}\sum_{i=1}^{n} Y_i \in A_1,\ \tfrac{1}{n}\sum_{i=n+1}^{2n} Y_i \in A_2\Big) \\ &= P_n(A_1)P_n(A_2).\end{aligned}$$

Therefore, by part (a) of the theorem,

$$\widehat{I}(A_{1,2}) \le \frac{1}{2}\big[\widehat{I}(A_1) + \widehat{I}(A_2)\big].$$

But $\widehat{I}(A_1) \leq I(x_1)$ and $\widehat{I}(A_2) \leq I(x_2)$. Hence $\widehat{I}(A_{1,2}) \leq \frac{1}{2}[I(x_1) + I(x_2)]$. Now take the supremum over A_1, A_2 and use that, by assumption (III.25)(2) and the definition of I,

$$\sup_{A_1,A_2 \text{ open convex}:\ A_1 \ni x_1, A_2 \ni x_2} \widehat{I}(A_{1,2}) = I(x_{1,2}),$$

to get

$$I(x_{1,2}) \leq \frac{1}{2}\big[I(x_1) + I(x_2)\big].$$

By iteration of this inequality we find that

$$I\big(\alpha x + (1-\alpha)y\big) \leq \alpha I(x) + (1-\alpha)I(y) \qquad \forall \text{ dyadic } \alpha \in (0,1).$$

However, from assumption (III.25)(3) and the lower semi-continuity of I it easily follows that the last display holds for all $\alpha \in (0,1)$. □

Theorem III.27 makes no claim as to validity of the LDP for (P_n). However, the above argument can be easily turned into a proof of the weak LDP.

THEOREM III.30 **(General Cramér Theorem)** *The sequence of probability measures (P_n) in Theorem* III.27 *satisfies the weak* LDP *with rate n and with weak rate function I.*

PROOF. We need to verify the conditions in Definition III.23.

Lower bound:

Let $O \subset \mathcal{X}$ be open. Pick $x \in O$. Because of (III.25)(2), there exists an A open convex such that $O \supset A \ni x$. Hence, with the help of Theorem III.27(a), we have

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(O) \geq -\widehat{I}(A) \geq -I(x).$$

Optimize over x to get the claim.

Upper bound:

For $\delta > 0$ and $x \in \mathcal{X}$, define

$$I^\delta(x) = \min\left\{I(x) - \delta, \frac{1}{\delta}\right\}.$$

By Theorem III.27(a–b), for every $x \in \mathcal{X}$ there exists an $A_x^\delta \ni x$ open convex such that

$$\lim_{n\to\infty} \frac{1}{n} \log P_n(A_x^\delta) = -\widehat{I}(A_x^\delta) \leq -I^\delta(x).$$

Let $K \subset \mathcal{X}$ be compact. Then the open covering $\cup_{x\in K} A_x^\delta$ of K has a finite subcovering $\cup_{i=1,\ldots,N} A_{x_i}^\delta$ with $N = N(\delta, K)$. It therefore follows with the help of (I.2) that

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(K) \leq \max_{i=1,\ldots,N} [-I^\delta(x_i)] \leq -\inf_{x\in K} I^\delta(x).$$

Now let $\delta \downarrow 0$. Then the RHS tends to $-\inf_{x\in K} I(x) = -I(K)$. □

Theorems III.27 and III.30 explain why convex rate functions are natural for additive functionals of i.i.d. sequences. They also give us an extension of Cramér's Theorem and Sanov's Theorem to a rather general context, albeit in the weak LDP form. There is no statement identifying I explicitly. However, with more additional structure on $\mathcal{X}$ this identification can be achieved too: for instance, when

$\mathcal{X}$ is a topological vector space the rate function becomes a generalized Legendre transform. See Dembo and Zeitouni [A2] Section 4.5.

REMARK III.31 The proofs of Theorems III.27 and III.30 carry over when the sequence (Y_i) is stationary and superadditive. See e.g. Grimmett [D33]. It also carries over when (Y_i) has strong mixing properties (see e.g. Dembo and Zeitouni [A2] Exercise 6.1.18). Thus, we are not at all tied down to the i.i.d. context. Some generalizations will be discussed in Chapter V.

We close this section with the following observation relating to the Contraction Principle:

THEOREM III.32 (**Convexity Preservation**) *Assume* (III.25)(1). *Then convexity of the rate function is preserved under contraction by linear transformations.*

PROOF. Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces satisfying (III.25)(1). Let (P_n) be a sequence of probability measures on $\mathcal{X}$ satisfying the LDP with rate n and with rate function I. Let $T\colon \mathcal{X} \to \mathcal{Y}$ be a linear map and let $Q_n = P_n \circ T^{-1}$. Then we know from Theorem III.20 that (Q_n) satisfies the LDP with rate n and with rate function J given by (III.21). We must show that J inherits convexity from I.

Pick $y_1, y_2 \in \mathcal{Y}$ and $\alpha \in [0,1]$. Then

$$\begin{aligned} J(\alpha y_1 + (1-\alpha)y_2) &= \inf_{x\in\mathcal{X}:\ T(x)=\alpha y_1+(1-\alpha)y_2} I(x) \\ &= \inf_{x_1,x_2\in\mathcal{X}:\ T(\alpha x_1+(1-\alpha)x_2)=\alpha y_1+(1-\alpha)y_2} I(\alpha x_1 + (1-\alpha)x_2) \\ &\le \inf_{x_1,x_2\in\mathcal{X}:\ T(x_1)=y_1,\, T(x_2)=y_2} I(\alpha x_1 + (1-\alpha)x_2), \end{aligned}$$

where the linearity of T is used to get the inequality. If I is convex on $\mathcal{X}$, then

$$I(\alpha x_1 + (1-\alpha)x_2) \le \alpha I(x_1) + (1-\alpha)I(x_2).$$

Substitution of this inequality allows us to split the infimum and to obtain

$$J(\alpha y_1 + (1-\alpha)y_2) \le \alpha J(y_1) + (1-\alpha)J(y_2).$$

Since y_1, y_2, α are arbitrary, J is convex on $\mathcal{Y}$. □

EXERCISE III.33 *Show that if I is strictly convex on its domain, then so is J.*

III.8 Relation to earlier results

Before closing this chapter, we backtrack and make the link with the examples described in Chapters I and II:

THEOREM III.34 *In each of the following examples the* LDP *holds with rate* n *and with rate function as computed in Chapters* I *and* II*:*

(a) $\mathcal{X} = \mathbb{R}$, $P_n(\cdot) = \mathbb{P}(\frac{1}{n}S_n \in \cdot)$, $I = \mathcal{L}_{\log\varphi}$ *($\mathcal{L}$ = Legendre transform).*
(b) $\mathcal{X} = \mathfrak{M}_1(\Gamma)$, $P_n(\cdot) = \mathbb{P}(L_n \in \cdot)$, $I = I_\rho$.
(c) $\mathcal{X} = \widetilde{\mathfrak{M}}_1(\Gamma^N)$, $P_n(\cdot) = \mathbb{P}(L_n^N \in \cdot)$, $I = I_\rho^N$.
(d) $\mathcal{X} = \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}})$, $P_n(\cdot) = \mathbb{P}(R_n \in \cdot)$, $I = I_\rho^\infty$.
(e) $\mathcal{X} = \mathfrak{M}_1(\mathbb{N})$, $P_n(\cdot) = \mathbb{P}(L_n \in \cdot)$, $I = I_\rho$.
(The distance on $\mathcal{X}$ is chosen appropriately.)

PROOF. In Chapters I and II we only proved large deviations for sets that are closed halflines, respectively, complements of closed balls. Therefore a little work is needed to get the result for arbitrary open or closed sets, but this is straightforward.
(a) It is immediate from Theorem I.4 and the convexity of I that the lower bound in the LDP holds for open balls and the upper bound in the LDP holds for closed balls (see also Exercise I.23). Hence, the weak LDP can be deduced via the standard approximation of open sets and compact sets by open and closed balls, respectively, in the spirit of the proof of Theorem III.30. To get the LDP, all we need to do is prove exponential tightness (see Section III.6). But this follows easily because I has compact level sets (recall Comment (5) in Section III.2). Lemma I.14 shows that I is a rate function.
(b) Analogous to (a) via Theorem II.2. Hence the weak LDP holds by the standard approximation. Since $\mathfrak{M}_1(\Gamma)$ is compact, we get the LDP for free. Lemma II.4 shows that I_ρ is a rate function.
(c) Analogous to (b) via Theorem II.8. Lemma II.20 shows that I_ρ^N is a rate function.
(d) The proof of Theorem II.23 shows that the lower bound in the LDP holds for open balls and the upper bound in the LDP holds for closed balls. Hence the argument is again analogous to (b). Note that $\Gamma^{\mathbb{N}}$ is compact in the product topology (by Tychonoff's Theorem). Consequently, $\widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}})$ is compact in the topology induced by the "myopic" total variation distance used in Theorem II.23. Lemma II.29 shows that I_ρ^∞ is a rate function.
(e) Analogous to (d) via Theorem II.36. Lemma II.39 shows that I_ρ is a rate function. □

Note that examples (a–e) satisfy (III.25). Also note that all the contractions encountered in Chapter II are linear.

CHAPTER IV

LARGE DEVIATIONS FOR MARKOV SEQUENCES

In Chapters I and II we studied large deviations for i.i.d. sequences. This led us to describe a general theory of large deviations in Chapter III. In this chapter we turn to Markov chains and show how some of the theorems in Chapters I and II transform to this context. The results to be described go back to Donsker and Varadhan [D17].

IV.1 Radon-Nikodym formula

We restrict ourselves to the following situation:

$$\begin{aligned} &X_i \in \Gamma = \{1, \ldots, r\} \subset \mathbb{N}, \\ &X_1, X_2, \ldots \text{ is Markov with transition matrix } P = (P_{st})_{s,t \in \Gamma}, \\ &P_{st} > 0 \ \forall s, t \in \Gamma. \end{aligned} \qquad \text{(IV.1)}$$

This condition is the analogue of condition (II.1) for i.i.d. sequences. The stationary distribution $\pi = (\pi_s)$ of the Markov chain is unique and satisfies

$$\pi_s > 0 \qquad \forall s \in \Gamma.$$

We pick as initial distribution

$$\mathbb{P}(X_1 = s) = \pi_s \qquad \forall s \in \Gamma,$$

so the Markov chain is stationary. The law of (X_i) is denoted by $\mathbb{P}^X$.

As we will see in a moment, it is convenient to introduce auxiliary i.i.d. Γ-valued random variables with distribution π. We denote these random variables by $Y_1, Y_2, \ldots$, and they obviously satisfy condition (II.1). The law of (Y_i) is denoted by $\mathbb{P}^Y$.

Let us recall the pair empirical measure

$$L_n^2 = \frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i, X_{i+1})},$$

with periodic boundary conditions $X_{n+1} = X_1$. Our key observation now is that the probability of a given realization of the Markov chain is a functional of L_n^2, modulo boundary terms. Namely,

$$\begin{aligned} \mathbb{P}^X[x_1, \ldots, x_n] &= \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \\ &= \pi_{x_1} P_{x_1, x_2} \times \cdots \times P_{x_{n-1}, x_n} \\ &= \frac{\pi_{x_1}}{P_{x_n, x_1}} e^{\sum_{i=1}^{n} \log P_{x_i, x_{i+1}}} \\ &= \frac{\pi_{x_1}}{P_{x_n, x_1}} e^{n \sum_{s,t} L_n^2[x_1, \ldots, x_n](s,t) \log P_{st}}, \end{aligned}$$

where $L_n^2[x_1,\dots,x_n]$ is the pair empirical measure associated with $x_1,\dots,x_n$. A prefactor arises, which however is bounded away from 0 and ∞ and therefore is negligible on an exponential scale. Similarly, we have

$$\mathbb{P}^Y[x_1,\dots,x_n] = \prod_{i=1}^n \pi_{x_i} = e^{n\sum_{s,t} L_n^2[x_1,\dots,x_n](s,t)\log \pi_t}.$$

Combining the last two displays, we obtain

$$\frac{d\,\mathbb{P}^X}{d\,\mathbb{P}^Y}[\,\cdot\,] = O(1)e^{nF(L_n^2[\,\cdot\,])}, \tag{IV.2}$$

with

$$F(\nu) = \sum_{s,t} \nu_{st}\log\Big(\frac{P_{st}}{\pi_t}\Big), \qquad \nu \in \widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma).$$

The Radon-Nikodym formula in (IV.2) allows us to transform questions about (X_i) into questions about (Y_i). Note that F is bounded and continuous on $\widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma)$.

IV.2 The LDP for discrete-time Markov chains

With the help of the Radon-Nikodym formula in (IV.2) and the version of Varadhan's Lemma given in Section III.4, we can deduce the large deviation properties for Markov chains from those for i.i.d. sequences. We will do this for the pair empirical measure, because this quantity is best adapted to the "pair dependence" in Markov chains, as is evident from (IV.2).

THEOREM IV.3 *Let (X_i) be a Markov chain satisfying* (IV.1). *Then the family (P_n^X) defined by*

$$P_n^X(\,\cdot\,) = \mathbb{P}^X(L_n^2 \in \,\cdot\,)$$

satisfies the LDP *on $\widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma)$ with rate n and with rate function*

$$I_P^2(\nu) = \sum_{s,t}\nu_{st}\log\Big(\frac{\nu_{st}}{\bar\nu_s P_{st}}\Big), \tag{IV.4}$$

where $\bar\nu_s = \sum_t \nu_{st}$.

PROOF. It follows from (IV.2) that, for every Borel set $S\subset\widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma)$,

$$\frac1n\log P_n^X(S) = \frac1n\log\int_S \mathbb{P}_n^X(L_n^2\in d\nu) = O\Big(\frac1n\Big) + \frac1n\log\int_S e^{nF(\nu)}\mathbb{P}_n^Y(L_n^2\in d\nu).$$

We now observe two facts. Firstly, we know from Theorem II.8 that (P_n^Y) defined by $P_n^Y(\cdot) = \mathbb{P}^Y(L_n^2\in\,\cdot\,)$ satisfies the LDP on $\widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma)$ with rate n and with rate function

$$I_\pi^2(\nu) = \sum_{s,t}\nu_{st}\log\Big(\frac{\nu_{st}}{\bar\nu_s\pi_t}\Big).$$

Secondly, the integral in the RHS of the previous display is exactly of the form appearing in Theorem III.17, because F is bounded and continuous. Hence, (P_n^X) satisfies the LDP on $\widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma)$ with rate n and with rate function

$$I_P^2(\nu) = I_\pi^2(\nu) - F(\nu) = \sum_{s,t}\nu_{st}\log\Big(\frac{\nu_{st}}{\bar\nu_s P_{st}}\Big).$$

□

Equation (IV.4) says that $I_P^2(\nu) = H(\nu|\bar{\nu} \otimes P)$, the relative entropy of ν with respect to $\bar{\nu} \otimes P$, defined by $(\bar{\nu} \otimes P)_{st} = \bar{\nu}_s P_{st}$. It has properties reminiscent of the i.i.d. case:

LEMMA IV.5 *Assume* (IV.1).
(i) I_P^2 *is finite, continuous and strictly convex on* $\widetilde{\mathfrak{M}}_1(\Gamma \times \Gamma)$, *except along line segments* $\{\alpha\nu + (1-\alpha)\nu' \colon \alpha \in [0,1]\}$ *between any* ν *and* ν' *satisfying* $\nu_{st}/\bar{\nu}_s = \nu'_{st}/\bar{\nu}'_s$ $\forall s,t$. *Along such line segments* I_P^2 *is affine.*
(ii) $I_P^2(\nu) \geq 0$ *with equality if and only if* $\nu = \pi \otimes P$.

PROOF. (i) This property is an immediate consequence of Lemma II.11 and the relation $I_P^2(\nu) = I_\pi^2(\nu) - F(\nu)$. Note that $\nu \mapsto F(\nu)$ is linear.
(ii) Use Jensen's inequality, which tells us that $I_P^2(\nu) = 0$ if and only if

$$\nu_{st} = \bar{\nu}_s P_{st} \qquad \forall s,t.$$

Sum over s and use the condition that ν has equal marginals, to obtain

$$\bar{\nu}_t = \sum_s \nu_{st} = \sum_s \bar{\nu}_s P_{st} \qquad \forall t.$$

So $\bar{\nu}$ is a stationary distribution for P and therefore is equal to π. □

A corollary of Theorem IV.3 is the following LDP for the empirical measure L_n.

THEOREM IV.6 *Assume* (IV.1). *Let* $P_n^X(\cdot) = \mathbb{P}^X(L_n \in \cdot)$. *Then* (P_n^X) *satisfies the* LDP *on* $\mathfrak{M}_1(\Gamma)$ *with rate* n *and with rate function*

$$I_P(\mu) = \inf_{\nu \in \widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma)\colon\, \bar{\nu}=\mu} I_P^2(\nu).$$

PROOF. This follows from Theorem IV.3 by applying the Contraction Principle in Theorem III.20. □

For later purposes we transform the formula for I_P into a different one.

THEOREM IV.7 *Assume* (IV.1). *Then*

$$I_P(\mu) = \sup_{u>0} \Big[-\sum_{s=1}^{r} \mu_s \log\Big(\frac{(Pu)_s}{u_s}\Big)\Big], \qquad \text{(IV.8)}$$

where the supremum runs over all $u\colon \Gamma \to (0,\infty)$, *and* $(Pu)_s = \sum_t P_{st}u_t$.

PROOF. Fix $\mu \in \mathfrak{M}_1(\Gamma)$. Abbreviate

$$i_\mu(u) = -\sum_s \mu_s \log\left(\frac{(Pu)_s}{u_s}\right).$$

Suppose that $\mu > 0$, i.e., $\mu_s > 0\ \forall s$. Then $u \mapsto i_\mu(u)$ is continuous and differentiable on the set $\{u > 0\colon \sup_s u_s = 1\}$ and tends to ∞ when any of the components of u tends to 0. Hence, the supremum in (IV.8) is attained at $u = u^*$ solving

$$0 = -\sum_s \mu_s \Big[\frac{P_{st}}{(Pu^*)_s} - \frac{\delta_{st}}{u_t^*}\Big] \qquad \forall t,$$

i.e.,

$$\mu_t = \sum_s \mu_s Q^{u^*}_{st} \qquad \forall t \qquad \text{with} \qquad Q^{u^*}_{st} = \frac{P_{st}u^*_t}{(Pu^*)_s} > 0.$$

In other words, the maximizer u^* is such that μ is the stationary distribution of the stochastic matrix Q^{u^*}.

EXERCISE IV.9 *Show that the maximizer u^* is unique.*

Let ν^* be defined by

$$\nu^*_{st} = \mu_s Q^{u^*}_{st}.$$

Since $\nu, \nu^* \in \widetilde{\mathfrak{M}}_1(\Gamma \times \Gamma)$ with $\bar{\nu} = \bar{\nu}^* = \mu$, an easy computation gives

$$I^2_P(\nu) = I^2_P(\nu^*) + H(\nu|\nu^*) = i_\mu(u^*) + H(\nu|\nu^*),$$

where we use that $\nu^*_{st}/\bar{\nu}^*_s P_{st} = u^*_t/(Pu^*)_s$. The infimum of the last term over the set $\{\nu \in \widetilde{\mathfrak{M}}_1(\Gamma \times \Gamma) \colon \bar{\nu} = \mu\}$ is zero. This completes the proof for $\mu > 0$. If the latter condition fails, then the statement in the theorem is still true: restrict the sums to the support of μ. □

LEMMA IV.10 *Under* (IV.1), I_P *in* (IV.8) *has all the properties in Lemma* II.4.

EXERCISE IV.11 *Prove Lemma* IV.10.

EXERCISE IV.12 *Compute I_P for $P = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$ with $p \in (0,1)$.*

IV.3 Comments

Here are some additional facts concerning the results in Section IV.2.

(1) Since the empirical measure L_n keeps track of how often the Markov chain visits different points of the state space Γ, it is sometimes referred to as the *occupation time measure*. In general, the supremum in Theorem IV.7 identifying the rate function $I_P(\mu)$ cannot be evaluated explicitly, not even when P is symmetric. This is because, unlike pairs, singletons are "not natural for Markov chains". In contrast, for i.i.d. sequences we found that $I_\rho(\mu)$, the corresponding rate function for L_n, can be evaluated explicitly as the relative entropy of μ with respect to ρ (see Theorem II.2). Only after we go "one level lower" to the empirical average $\frac{1}{n}S_n$, we end up with a rate function, the Legendre transform, that can in general not be simplified further (see Theorem I.4).

(2) It should be clear that the LDP's in Theorems IV.3 and IV.6 are closely linked with Perron-Frobenius theory for finite positive matrices (see Seneta [C8] Chapter 1). The earliest results establishing this link are due to Miller [D44]. See also Section V.4.

(3) In the proof of Theorem IV.3 the trick enabling us to use the LDP for i.i.d. random variables is based on lifting the transition matrix to the exponential (which is the analogue of the Cramér transform used in Section I.3). Since L^2_n is a functional of L^N_n for all $N \geq 3$, the same Radon-Nikodym argument allows us to deduce the LDP for the latter quantities as well. We then find that the corresponding rate

function reads (compare with Theorem II.18)

$$\begin{aligned} I_P^N(\nu) &= H(\nu|\bar\nu\otimes P) \\ &= \sum_{s_1,\dots,s_N} \nu_{s_1,\dots,s_N}\log\left(\frac{\nu_{s_1,\dots,s_N}}{\bar\nu_{s_1,\dots,s_{N-1}}P_{s_{N-1}s_N}}\right), \qquad \nu\in\widetilde{\mathfrak{M}}_1(\Gamma^N). \end{aligned}$$

Similarly, the LDP for the quantity R_n can be deduced via the same type of reasoning and the rate function is (compare with Theorem II.23)

$$I_P^\infty(\nu) = \sup_{N\geq 2} I_P^N(\pi_N\nu) = \sup_{N\geq 2} H\left(\pi_N\nu|\pi_{N-1}\nu\otimes P\right), \qquad \nu\in\widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}}).$$

See also Deuschel and Stroock [A3] Section 4.4, Dembo and Zeitouni [A2] Section 6.5.

(4) If P fails to satisfy the positivity condition in (IV.1) but is irreducible, then Theorem IV.3 still applies when $\Gamma\times\Gamma$ is replaced by $\{(s,t)\in\Gamma\times\Gamma\colon P_{st}>0\}$. The proof is easily adapted. It is also not relevant that the Markov chain starts in π.

(5) Generalizations of Theorems IV.3, IV.6 and IV.7 away from (IV.1) are possible under certain conditions, provided of course we also extend our result in Theorem II.8 away from (II.1). For instance, in view of Theorem II.36 our results immediately carry over to $\Gamma=\mathbb{N}$ when the transition matrix P on $\mathbb{N}\times\mathbb{N}$ is such that P_{st}/π_t is bounded away from 0 and ∞ (see the end of Section IV.1). Such Markov chains are called "uniformly ergodic". For further information we refer the reader to Donsker and Varadhan [D17], Iscoe, Ney and Nummelin [D40], Ellis [D19], Ellis and Wyner [D20], de Acosta [D1], Bolthausen and Schmock [D7].

IV.4 The LDP for continuous-time Markov chains

In this section we move away from the discrete-time setting adopted so far in order to consider the generalization of Sanov's theorem to continuous-time Markov chains. We will see that if the Markov chain is *reversible*, then a nice formula for the rate function comes out. The results to be described go back to Donsker and Varadhan [D17]. An application will be given in Chapter VIII.

Let

$$\begin{array}{l}(X_t)_{t\geq 0} \text{ be a } \Gamma\text{-valued continuous-time Markov chain} \\ \text{with an irreducible generator } G=(G_{ij})_{i,j\in\Gamma}.\end{array} \tag{IV.13}$$

Here Γ is the finite set that was used in the discrete-time setting (and we switch from indices s,t to i,j to avoid a clash with the time index). Define the empirical measure (or occupation time measure)

$$L_t = \frac{1}{t}\int_0^t \delta_{X_s}\,ds,$$

which is a random element of $\mathfrak{M}_1(\Gamma)$. Recall that $\mathfrak{M}_1(\Gamma)$ is a Polish space with the total variation distance $d(\mu,\nu)=\frac{1}{2}\sum_s|\mu_s-\nu_s|$. In what follows, the obvious analogue of Definition III.6 for a continuous-time index applies.

THEOREM IV.14 *Assume* (IV.13). *Let* $P_t(\cdot)=\mathbb{P}(L_t\in\cdot)$. *Then:*
(a) (P_t) *satisfies the* LDP *on* $\mathfrak{M}_1(\Gamma)$ *with rate t and with rate function*

$$\widetilde{I}_G(\nu) = \sup_{u>0}\left[-\sum_{i=1}^{r}\nu_i\frac{(Gu)_i}{u_i}\right], \tag{IV.15}$$

where the supremum runs over all $u\colon \Gamma \to (0,\infty)$, *and* $(Gu)_i = \sum_j G_{ij}u_j$.
(b) *If G is symmetric, then the supremum can be evaluated explicitly:*

$$\widetilde{I}_G(\nu) = -\sum_{i,j=1}^{r} \sqrt{\nu_i}G_{ij}\sqrt{\nu_j} = \langle \sqrt{\nu}, (-G)\sqrt{\nu}\,\rangle. \tag{IV.16}$$

PROOF. The proof uses an approximation argument in which the claim is deduced from Theorems IV.6 and IV.7 by chopping up time into small intervals and shrinking these intervals to zero afterwards.

(a) Fix $\delta > 0$. Define

$$L_t^\delta = \lfloor t/\delta \rfloor^{-1} \sum_{k=1}^{\lfloor t/\delta \rfloor} \delta_{X_{k\delta}}.$$

Note that this is the empirical measure after $\lfloor t/\delta \rfloor$ steps of the Γ-valued discrete-time Markov chain with strictly positive transition matrix

$$P^\delta = e^{\delta G}.$$

Let $P_t^\delta(\cdot) = \mathbb{P}(L_t^\delta \in \cdot)$. Then, using Theorems IV.6 and IV.7, we have that (P_t^δ) satisfies the LDP on $\mathfrak{M}_1(\Gamma)$ with rate t and with rate function

$$\frac{1}{\delta} I_{P^\delta}(\nu) = \frac{1}{\delta}\sup_{u>0}\Big[-\sum_{i=1}^{r}\nu_i \log\Big(\frac{(P^\delta u)_i}{u_i}\Big)\Big].$$

Next, L_t^δ is a good approximation of L_t for small δ because

$$d(L_t, L_t^\delta) \le \lfloor t/\delta \rfloor^{-1} \times \text{the number of jumps by the Markov chain in } [0,t].$$

Indeed, the latter number is stochastically dominated by $N(ct)$, a Poisson random variable with mean ct, where

$$c = \sup_i \sum_{j\neq i} G_{ij} < \infty.$$

Hence, for $\epsilon > \delta c$ we can estimate

$$\limsup_{t\to\infty} \frac{1}{t}\log \mathbb{P}\big(d(L_t, L_t^\delta) \ge \epsilon\big) \le \limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}\big(N(ct) \ge \epsilon\lfloor t/\delta\rfloor\big) = -cI\Big(\frac{\epsilon}{\delta c}\Big)$$

with

$$I(z) = z\log z - z + 1$$

(recall Exercise I.11). From this it follows that

$$\lim_{\delta\downarrow 0}\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}\big(d(L_t, L_t^\delta) \ge \epsilon\big) = -\infty \qquad \forall \epsilon > 0, \tag{IV.17}$$

showing that the approximation is sharp in the limit as $\delta \downarrow 0$. It is now possible to deduce that (P_t) satisfies the LDP on $\mathfrak{M}_1(\Gamma)$ with rate t and with rate function $\widetilde{I}_G$ defined by

$$\lim_{\delta\downarrow 0}\frac{1}{\delta} I_{P^\delta}(\nu) = \widetilde{I}_G(\nu), \qquad \nu \in \mathfrak{M}_1(\Gamma). \tag{IV.18}$$

The argument proceeds in three steps. The proof that $\widetilde{I}_G$ is a rate function is deferred to Exercise IV.23.

Convergence:

Since $P^0 = \mathbb{1}$ and $\frac{d}{d\delta}P^\delta = GP^\delta$, we have

$$\begin{aligned}
\tfrac{1}{\delta}I_{P^\delta}(\nu) &= \sup_{u>0}\Big[-\sum_i \nu_i\, \tfrac{1}{\delta}\int_0^\delta d\epsilon\, \tfrac{(GP^\epsilon u)_i}{(P^\epsilon u)_i}\Big] \\
&\le \tfrac{1}{\delta}\int_0^\delta d\epsilon \sup_{u>0}\Big[-\sum_i \nu_i \tfrac{(GP^\epsilon u)_i}{(P^\epsilon u)_i}\Big] \\
&\le \tfrac{1}{\delta}\int_0^\delta d\epsilon\, \widetilde{I}_G(\nu) = \widetilde{I}_G(\nu)
\end{aligned}$$

and so

$$\limsup_{\delta\downarrow 0} \tfrac{1}{\delta} I_{P^\delta}(\nu) \le \widetilde{I}_G(\nu).$$

On the other hand, for all $u > 0$ there exists a $c(u) < \infty$ such that for all $\delta > 0$

$$\begin{aligned}
-\tfrac{1}{\delta}\log\Big(\tfrac{(P^\delta u)_i}{u_i}\Big) &= -\tfrac{1}{\delta}\log\Big(1 + \tfrac{1}{u_i}([P^\delta - \mathbb{1}]u)_i\Big) \\
&\ge -\tfrac{1}{u_i}(\tfrac{1}{\delta}[P^\delta - \mathbb{1}]u)_i \ge -\tfrac{(Gu)_i}{u_i} - c(u)\delta.
\end{aligned}$$

Multiplying by ν_i, summing over i and letting $\delta \downarrow 0$, we find that

$$\liminf_{\delta\downarrow 0} \tfrac{1}{\delta} I_{P^\delta}(\nu) \ge \liminf_{\delta\downarrow 0}\Big[-\sum_i \nu_i \tfrac{1}{\delta}\log\Big(\tfrac{(P^\delta u)_i}{u_i}\Big)\Big] \ge -\sum_i \nu_i \tfrac{(Gu)_i}{u_i} \text{ for all } u > 0.$$

Take the supremum over $u > 0$ to obtain

$$\liminf_{\delta\downarrow 0} \tfrac{1}{\delta} I_{P^\delta}(\nu) \ge \widetilde{I}_G(\nu).$$

Hence we have proved (IV.18). We proceed to prove the bounds in the LDP.

Lower bound:

Let $O \subset \mathfrak{M}_1(\Gamma)$ be open. For any $\epsilon > 0$ we have

$$\mathbb{P}(L_t \in O) \ge \mathbb{P}(L_t^\delta \in O_\epsilon) - \mathbb{P}\big(d(L_t, L_t^\delta) \ge \epsilon\big),$$

where $O_\epsilon = O \cap \{x\colon d(x, \partial O) > \epsilon\}$ is the ϵ-interior of O. By (IV.17), the last term decays at a rate that tends to infinity as $\delta \downarrow 0$ for every $\epsilon > 0$. Hence we have

$$\liminf_{t\to\infty} \frac{1}{t}\log \mathbb{P}(L_t \in O) \ge -\liminf_{\epsilon\downarrow 0}\ \liminf_{\delta\downarrow 0} \frac{1}{\delta}\Big[I_{P^\delta}(O_\epsilon)\Big] \ge -\liminf_{\epsilon\downarrow 0} \widetilde{I}_G(O_\epsilon),$$

where in the last inequality we use that $\frac{1}{\delta}I_{P^\delta} \le \widetilde{I}_G$ (as established above). The RHS is equal to $-\widetilde{I}_G(O)$ because every $x \in O$ falls in O_ϵ for ϵ small enough. Thus,

$$\liminf_{t\to\infty} \frac{1}{t}\log \mathbb{P}(L_t \in O) \ge -\widetilde{I}_G(O).$$

Upper bound:

Since $\mathfrak{M}_1(\Gamma)$ is compact, every closed set is compact. Let $K \subset \mathfrak{M}_1(\Gamma)$ be compact. For any $\epsilon > 0$ we have

$$\mathbb{P}(L_t \in K) \le \mathbb{P}(L_t^\delta \in K_\epsilon) + \mathbb{P}\big(d(L_t, L_t^\delta) \ge \epsilon\big),$$

where $K_\epsilon = K \cup \{x\colon d(x, \partial K) \le \epsilon\}$ is the ϵ-exterior of K. Using (IV.17) once more, we get

$$\limsup_{t\to\infty} \frac{1}{t}\log \mathbb{P}(L_t \in K) \le -\limsup_{\epsilon\downarrow 0}\ \limsup_{\delta\downarrow 0}\Big[\frac{1}{\delta} I_{P^\delta}(K_\epsilon)\Big]. \qquad \text{(IV.19)}$$

Next, we show that

$$\limsup_{\delta\downarrow 0} \frac{1}{\delta} I_{P^\delta}(K_\epsilon) \geq \widetilde{I}_G(K_\epsilon) \qquad \forall \epsilon > 0, \tag{IV.20}$$

which goes as follows. Since K_ϵ is compact and $\frac{1}{\delta}I_{P^\delta}$ is continuous (Lemma IV.10), there exists a $\nu_\epsilon^\delta \in K_\epsilon$ such that $\frac{1}{\delta}I_{P^\delta}(K_\epsilon) = \frac{1}{\delta}I_{P^\delta}(\nu_\epsilon^\delta)$. The sequence $(\nu_\epsilon^\delta)_{\delta>0}$ has a convergent subsequence with a limit $\nu_\epsilon \in K_\epsilon$. Along this subsequence we have

$$\lim_{\delta\downarrow 0} \frac{1}{\delta} I_{P^\delta}(\nu_\epsilon^\delta) \geq \widetilde{I}_G(\nu_\epsilon),$$

by the same argument as for the lower estimate in (IV.18). This implies (IV.20) because $\widetilde{I}_G(\nu_\epsilon) \geq \widetilde{I}_G(K_\epsilon)$. Finally, we show that

$$\limsup_{\epsilon\downarrow 0} \widetilde{I}_G(K_\epsilon) \geq \widetilde{I}_G(K), \tag{IV.21}$$

which goes as follows. Since K_ϵ is compact and $\widetilde{I}_G$ is continuous (see Lemma IV.22 below), there exists a $\tilde{\nu}_\epsilon \in K_\epsilon$ such that $\widetilde{I}_G(K_\epsilon) = \widetilde{I}_G(\tilde{\nu}_\epsilon)$. The sequence $(\tilde{\nu}_\epsilon)_{\epsilon>0}$ has a convergent subsequence with a limit $\tilde{\nu} \in K$. Along this subsequence we have

$$\limsup_{\epsilon\downarrow 0} \widetilde{I}_G(\tilde{\nu}_\epsilon) = \widetilde{I}_G(\tilde{\nu}).$$

This implies (IV.21) because $\widetilde{I}_G(\tilde{\nu}) \geq \widetilde{I}_G(K)$. Combining (IV.19), (IV.20) and (IV.21), we arrive at

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}(L_t \in K) \leq -\widetilde{I}_G(K).$$

(b) Split

$$\widetilde{I}_G(\nu) = -\sum_{i,j} \sqrt{\nu_i} G_{ij} \sqrt{\nu_j} - \inf_{u>0} \sum_{i,j} G_{ij} \left(\nu_i \frac{u_j}{u_i} - \sqrt{\nu_i}\sqrt{\nu_j} \right).$$

If G is symmetric, then the last term may be rewritten as

$$-\inf_{u>0} \frac{1}{2} \sum_{i\neq j} G_{ij} \left(\sqrt{\frac{\nu_i u_j}{u_i}} - \sqrt{\frac{\nu_j u_i}{u_j}} \right)^2,$$

which is obviously zero because $G_{ij} > 0$ for all $i \neq j$. □

LEMMA IV.22 *Under* (IV.13), $\widetilde{I}_G$ *in* (IV.15) *has all the properties in Lemma* II.4.

EXERCISE IV.23 *Prove Lemma* IV.22.

EXERCISE IV.24 *Show that if the Markov chain is reversible, i.e.,* $\pi_i G_{ij} = \pi_j G_{ji}\ \forall i,j$ *with* π *the stationary distribution, then*

$$\widetilde{I}_G(\nu) = -\sum_{i,j} \pi_i \sqrt{\tfrac{\nu_i}{\pi_i}} G_{ij} \sqrt{\tfrac{\nu_j}{\pi_j}} = \left\langle \sqrt{\tfrac{\nu}{\pi}}, (-G)\sqrt{\tfrac{\nu}{\pi}} \right\rangle_\pi.$$

What is particularly noteworthy about Theorem IV.14 is that $\widetilde{I}_G$ takes on a nice form in the reversible case. This is, however, specific for the continuous-time setting, since I_P in Theorem IV.6 does *not* simplify in the reversible case. The function

$$f \mapsto \langle f, (-G)f \rangle_\pi$$

is referred to in the literature as the *Dirichlet form* associated with $-G$ (see Fukushima [C2] Chapter 1). This is a prominent object in Markov chain theory, for instance, because it plays a key role in rate of convergence estimates to equilibrium.

The Contraction Principle gives us the LDP for the empirical average

$$\frac{1}{t}S_t = \frac{1}{t}\int_0^t X_s ds.$$

EXERCISE IV.25 *Compute the rate function for* $\frac{1}{t}S_t$ *when* $G_{ij} = 1$ *for all* $i, j \in \Gamma$ *with* $i \neq j$.

It is further possible to extend Theorem IV.14 to a countable or continuous state space, for instance, $\Gamma = \mathbb{N}$ or $\Gamma = \mathbb{R}$ in the spirit of Section II.7, provided the generator G is suitably restricted. See the references in Comment (5) in Section IV.3.

CHAPTER V

LARGE DEVIATIONS FOR DEPENDENT SEQUENCES

In this chapter we describe a generalization of Cramér's Theorem, due to Gärtner [D24] and Ellis [D18], for random sequences that have a form of moderate dependence. The random variables take values in $\mathbb{R}^d$. We will see that, as a consequence of this theorem, many of the results in Chapters I, II and IV can be placed under a single heading. The main result to be described will be used in Chapter VII.

V.1 Preliminaries

We are given a sequence (Z_n) of random variables on a probability space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P})$, where $\mathcal{B}(\mathbb{R}^d)$ is the Borel sigma-field on $\mathbb{R}^d$, with moment generating functions

$$\varphi_n(t) = \mathbb{E}\, e^{\langle t, Z_n \rangle}, \qquad t \in \mathbb{R}^d, n \in \mathbb{N},$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product. We are aiming at an LDP for the family (P_n) given by $P_n(\cdot) = \mathbb{P}(Z_n \in \cdot)$. Throughout what follows we make the following assumptions:

$$\begin{aligned} &(1) \quad \lim_{n\to\infty} \tfrac{1}{n} \log \varphi_n(nt) = \Lambda(t) \in [-\infty, \infty] \text{ exists.} \\ &(2) \quad 0 \in \operatorname{int}(\mathcal{D}_\Lambda), \text{ with } \mathcal{D}_\Lambda = \{t \in \mathbb{R}^d \colon \Lambda(t) < \infty\}. \end{aligned} \qquad \text{(V.1)}$$

We need the following definition.

DEFINITION V.2 *Let Λ^* denote the Legendre transform of Λ, i.e.,*

$$\Lambda^*(x) = \sup_{t \in \mathbb{R}^d} [\langle x, t \rangle - \Lambda(t)], \qquad x \in \mathbb{R}^d. \qquad \text{(V.3)}$$

A point $x \in \mathbb{R}^d$ is called exposed for Λ^* if there exists a point $t \in \mathbb{R}^d$ such that

$$\Lambda^*(y) - \Lambda^*(x) > \langle y - x, t \rangle \qquad \forall y \neq x.$$

Such t is called (the normal to) an exposing hyperplane for x.

Obviously, $x \mapsto \Lambda^*(x)$ is convex. At exposed points it is strictly convex in all directions (see Fig. 8).

Since Λ^* turns out to play the role of the rate function in the Gärtner-Ellis Theorem, to be formulated in Section V.2, we begin by observing the following properties.

LEMMA V.4 *Assume* (V.1).
(i) Λ *is convex and* $\Lambda > -\infty$ *everywhere.*
(ii) Λ^* *is a rate function in the sense of Definition* III.5 *and is convex.*

PROOF. (i) Since $\log \varphi_n$ is convex for all n, so is Λ. Since $\Lambda(0) = 0$, convexity in combination with (V.1)(2) implies that $\Lambda > -\infty$ everywhere.
(ii) Trivially, $\Lambda^*(x) \geq -\Lambda(0) = 0$ for all $x \in \mathbb{R}^d$. Moreover, Λ^* is lower semi-continuous and convex because it is the supremum of linear functions. By (V.1)(2), there exists a $\delta > 0$ such that $B_{2\delta}(0) \subset \text{int}(\mathcal{D}_\Lambda)$. Since Λ is convex, it is continuous on $\text{int}(\mathcal{D}_\Lambda)$, and so we have $\sup_{t \in B_\delta(0)} \Lambda(t) = c < \infty$. Hence ($|\cdot|$ is Euclidean distance)

$$\Lambda^*(x) \geq \sup_{t \in B_\delta(0)} [\langle x, t \rangle - \Lambda(t)] \geq \delta |x| - c.$$

Consequently, Λ^* has bounded level sets. However, it is lower semi-continuous, and hence its level sets are compact. To prove that $\Lambda^* \not\equiv \infty$, simply note that because of $\Lambda(0) = 0$, (V.1)(2) and the convexity of Λ, we know that $\Lambda(t) \geq \langle x_0, t \rangle$ for some $x_0 \in \mathbb{R}^d$ and all $t \in \mathbb{R}^d$. Hence $\Lambda^*(x_0) = 0$. □

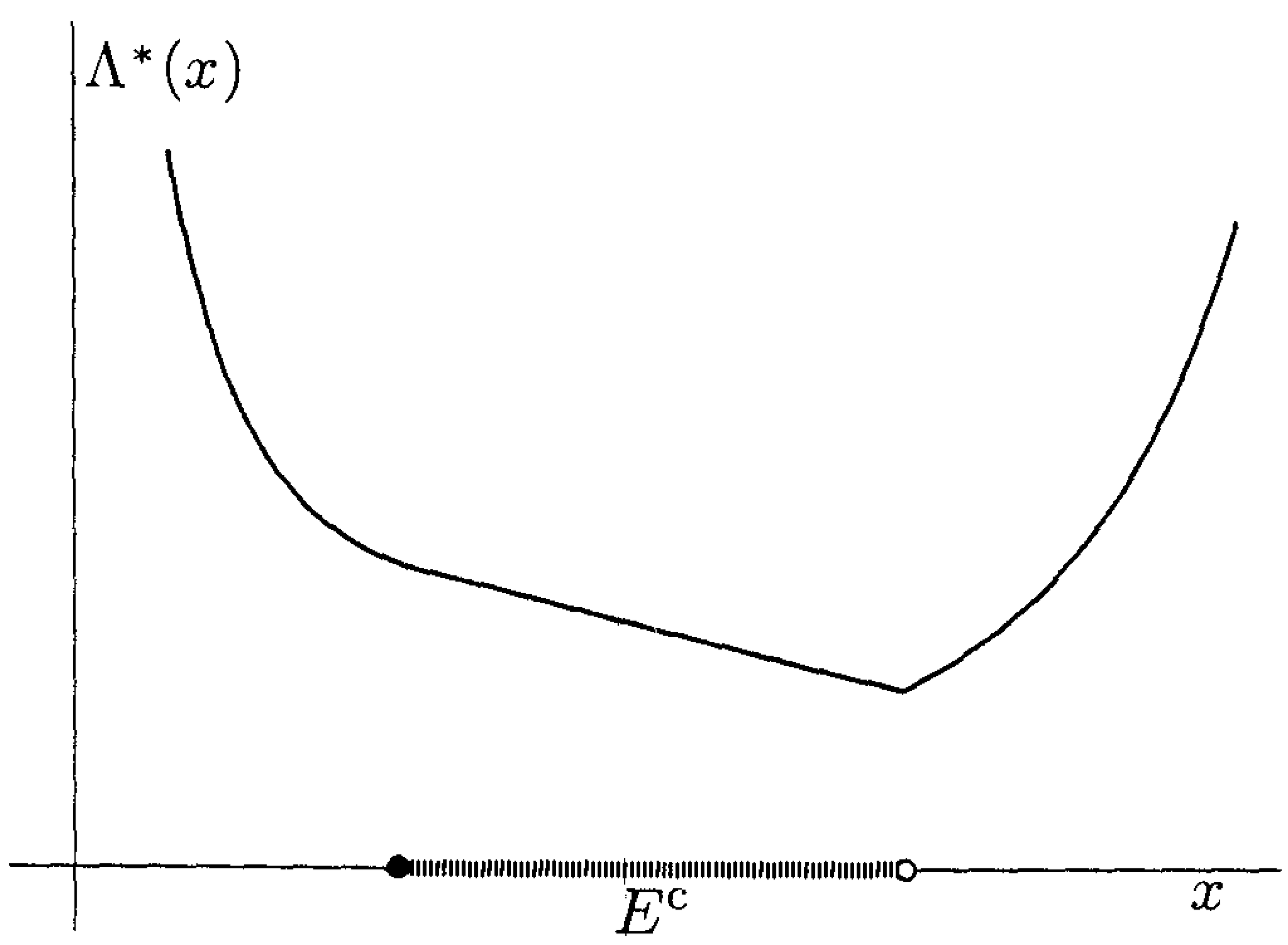

FIG. 8. An example of a set E of exposed points

V.2 The Gärtner-Ellis Theorem

With these preliminaries we are now ready to formulate the main result. Write

$$\Lambda^*(S) = \inf_{x \in S} \Lambda^*(x), \qquad S \subset \mathbb{R}^d. \tag{V.5}$$

THEOREM V.6 *Assume* (V.1). *Let* $P_n(\cdot) = \mathbb{P}(Z_n \in \cdot)$. *Then*
(a) $\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \leq -\Lambda^*(C) \quad \forall C \subset \mathbb{R}^d$ *closed.*
(b) $\liminf_{n\to\infty} \frac{1}{n} \log P_n(O) \geq -\Lambda^*(O \cap E) \quad \forall O \subset \mathbb{R}^d$ *open,*
where $E = E(\Lambda, \Lambda^*)$ *is the set of exposed points of* Λ^* *whose exposing hyperplane belongs to* $\text{int}(\mathcal{D}_\Lambda)$.
(c) *Suppose, in addition, that* Λ *satisfies:*

(1) Λ *is lower semi-continuous on* $\mathbb{R}^d$.
(2) Λ *is differentiable on* $\text{int}(\mathcal{D}_\Lambda)$.
(3) *Either* $\mathcal{D}_\Lambda = \mathbb{R}^d$ *or* Λ *is steep at* $\partial\mathcal{D}_\Lambda$, *i.e.,* $\lim_{t\to\partial\mathcal{D}_\Lambda:\, t\in\mathcal{D}_\Lambda} |\nabla\Lambda(t)| = \infty$.

Then $O \cap E$ *may be replaced by* O *in the* RHS *of* (ii). *Consequently,* (P_n) *satisfies the* LDP *on* $\mathbb{R}^d$ *with rate n and with rate function* Λ^*.

PROOF. The proof is analogous to that of Cramér's Theorem in Section I.8. However, there are a number of points that are a bit more delicate.

(a) Upper bound:

We first prove the claim for compact sets. Pick $\delta > 0$ arbitrary. For $x \in \mathbb{R}^d$, define

$$\Lambda_\delta^*(x) = \min\{\Lambda^*(x) - \delta, \tfrac{1}{\delta}\}.$$

For every $x \in \mathbb{R}^d$ there exists a $t_x \in \mathbb{R}^d$ such that

$$\langle x, t_x\rangle - \Lambda(t_x) \geq \Lambda_\delta^*(x).$$

Moreover, for every $x \in \mathbb{R}^d$ there exists a neighborhood A_x of x such that

$$\inf_{y \in A_x} \langle y - x, t_x\rangle \geq -\delta.$$

By the exponential Chebyshev inequality we therefore have

$$\begin{aligned} P_n(A_x) &= \mathbb{P}(Z_n \in A_x) \leq \mathbb{P}\Big(\langle Z_n - x, t_x\rangle \geq -\delta\Big) \\ &\leq e^{\delta n}\,\mathbb{E}\big(e^{n\langle Z_n - x, t_x\rangle}\big) = e^{\delta n}\varphi_n(nt_x)e^{-n\langle x, t_x\rangle}. \end{aligned}$$

Let $K \subset \mathbb{R}^d$ be compact. Then the covering $\cup_{x\in K} A_x$ of K has a finite subcovering $\cup_{i=1,\dots,N} A_{x_i}$. Hence, by (I.2),

$$\begin{aligned} \tfrac{1}{n}\log P_n(K) &\leq \tfrac{1}{n}\log[N \max_{i=1,\dots,N} P_n(A_{x_i})] \\ &\leq \tfrac{1}{n}\log N + \delta - \min_{i=1,\dots,N}\Big[\langle x_i, t_{x_i}\rangle - \tfrac{1}{n}\log\varphi_n(nt_{x_i})\Big]. \end{aligned}$$

Let $n \to \infty$ to obtain

$$\begin{aligned} \limsup_{n\to\infty} \tfrac{1}{n}\log P_n(K) &\leq \delta - \min_{i=1,\dots,N}\Big[\langle x_i, t_{x_i}\rangle - \Lambda(t_{x_i})\Big] \\ &\leq \delta - \min_{i=1,\dots,N}\Lambda_\delta^*(x_i) \leq \delta - \Lambda_\delta^*(K). \end{aligned}$$

Let $\delta \downarrow 0$ to get

$$\limsup_{n\to\infty} \frac{1}{n}\log P_n(K) \leq -\Lambda^*(K).$$

The extension from compact sets to closed sets amounts to showing the exponential tightness of (P_n) (recall Comment (5) in Section III.2). Indeed, if C is closed, then $C \cap [-N, N]^d$ is compact for all $N > 0$, and by the result in the last display in combination with (I.2) we have

$$\limsup_{n\to\infty} \frac{1}{n}\log P_n(C) \leq \max\big\{-\Lambda^*\big(C \cap [-N,N]^d\big), -M_N\big\} \qquad \forall N > 0$$

with

$$-M_N = \limsup_{n\to\infty} \frac{1}{n}\log P_n\big(\mathbb{R}^d \setminus [-N,N]^d\big).$$

If we manage to prove that $\lim_{N\to\infty} M_N = \infty$, then we will get the claim because

$$\lim_{N\to\infty} \Lambda^*\big(C \cap [-N,N]^d\big) = \Lambda^*(C).$$

Let $e^1, \dots, e^d$ denote the unit vectors in $\mathbb{R}^d$. Since $0 \in \text{int}(\mathcal{D}_\Lambda)$ by (V.1)(2), there exists a $\delta > 0$ such that $\Lambda(-\delta e^i) < \infty$ and $\Lambda(\delta e^i) < \infty$ for all i. By the exponential Chebyshev inequality we have, for all i,

$$\begin{aligned} \mathbb{P}(Z_n^i \leq -N) &\leq e^{-n\delta N}\varphi_n(-n\delta e^i) \\ \mathbb{P}(Z_n^i \geq N) &\leq e^{-n\delta N}\varphi_n(n\delta e^i). \end{aligned}$$

Hence, by (I.2),

$$\begin{aligned} -M_N &\leq \limsup_{n\to\infty} \tfrac{1}{n}\log P_n\Big(\exists i \in \{1,\dots,d\}\colon Z_n^i \notin [-N,N]\Big) \\ &\leq -\delta N + \max_{i=1,\dots,d} \max\{\Lambda(-\delta e_i), \Lambda(\delta e_i)\}, \end{aligned}$$

which proves the exponential tightness and completes the proof for closed sets.

(b) <u>*Lower bound*</u>:

Let $B_\epsilon(x)$ denote the open ball of radius ϵ around x. It suffices to prove the following:

$$\lim_{\epsilon\downarrow 0}\liminf_{n\to\infty} \frac{1}{n}\log P_n(B_\epsilon(x)) \geq -\Lambda^*(x) \quad \forall x \in E.$$

Indeed, for any open set O we have

$$P_n(O) \geq P_n(B_\epsilon(x)) \qquad \forall x \in O \cap E \;\forall \epsilon \leq \epsilon_0(x),$$

and so the claim will follow after letting $n \to \infty$, $\epsilon \downarrow 0$ and optimizing over $x \in O \cap E$.

Fix $x \in E$, and let $\tau \in \text{int}\,(\mathcal{D}_\Lambda)$ be an exposing hyperplane for x. Then $\varphi_n(n\tau) < \infty$ for n large enough, and so we can define a tilted probability measure $\widehat{P}_n$ by putting

$$\frac{d\widehat{P}_n}{dP_n}(y) = \frac{1}{\varphi_n(n\tau)}e^{n\langle y,\tau\rangle}, \qquad y \in \mathbb{R}^d.$$

(This is the analogue of the Cramér transform in Section I.3.) Next, we compute

$$\begin{aligned} \tfrac{1}{n}\log P_n(B_\epsilon(x)) &= \tfrac{1}{n}\log \textstyle\int_{B_\epsilon(x)} P_n(dy) \\ &= \tfrac{1}{n}\log\varphi_n(n\tau) + \tfrac{1}{n}\log\textstyle\int_{B_\epsilon(x)} e^{-n\langle y,\tau\rangle}\widehat{P}_n(dy) \\ &\geq \tfrac{1}{n}\log\varphi_n(n\tau) - \langle x,\tau\rangle - \epsilon|\tau| + \tfrac{1}{n}\log\widehat{P}_n(B_\epsilon(x)), \end{aligned}$$

where the last inequality uses that $|y - x| \leq \epsilon$ for $y \in B_\epsilon(x)$. Hence we arrive at

$$\lim_{\epsilon\downarrow 0}\liminf_{n\to\infty}\frac{1}{n}\log P_n(B_\epsilon(x)) \geq [\Lambda(\tau) - \langle x,\tau\rangle] + \lim_{\epsilon\downarrow 0}\liminf_{n\to\infty}\frac{1}{n}\log\widehat{P}_n(B_\epsilon(x)).$$

Since the first term is $\geq -\Lambda^*(x)$, it remains to show that

$$\lim_{\epsilon\downarrow 0}\liminf_{n\to\infty}\frac{1}{n}\log\widehat{P}_n(B_\epsilon(x)) = 0. \tag{V.7}$$

In Section I.3 we saw that for the i.i.d. case (V.7) could be easily proved with the help of Lemmas I.8–I.10 (which rely on the CLT). Now, however, we need a different argument. Let $\widehat{\varphi}_n$ denote the moment generating function associated with $\widehat{P}_n$. Then we have the relation

$$\widehat{\varphi}_n(nt) = \frac{\varphi_n(n(t+\tau))}{\varphi_n(n\tau)}.$$

Hence $(\widehat{\varphi}_n)$ satisfies (V.1)(1), with

$$\lim_{n\to\infty}\frac{1}{n}\log\widehat{\varphi}_n(nt)=\widehat{\Lambda}(t)=\Lambda(t+\tau)-\Lambda(\tau),\qquad t\in\mathbb{R}^d,$$

and $\widehat{\Lambda}$ satisfies (V.1)(2). Moreover, we have the relation

$$\widehat{\Lambda}^*(x)=\sup_{t\in\mathbb{R}^d}[\langle x,t\rangle-\widehat{\Lambda}(t)]=\Lambda^*(x)-\langle x,\tau\rangle+\Lambda(\tau),$$

and, by Lemma V.4(ii), $\widehat{\Lambda}^*$ is a rate function in the sense of Definition III.5. We can therefore apply part (a) of the theorem to $(\widehat{P}_n)$ and obtain

$$\limsup_{n\to\infty}\frac{1}{n}\log\widehat{P}_n\big(\mathbb{R}^d\setminus B_\epsilon(x)\big)\le-\widehat{\Lambda}^*\big(\mathbb{R}^d\setminus B_\epsilon(x)\big)\qquad\forall\epsilon>0.$$

However, by Definition III.5(D3) and Lemma III.3,

$$\widehat{\Lambda}^*\big(\mathbb{R}^d\setminus B_\epsilon(x)\big)=\widehat{\Lambda}^*(x_0)\text{ for some }x_0\neq x.$$

Since $x\in E$ and since $\tau\in\operatorname{int}(\mathcal{D}_\Lambda)$ is an exposing hyperplane for x, we have

$$\begin{aligned}\widehat{\Lambda}^*(x_0)&=\Lambda^*(x_0)-\langle x_0,\tau\rangle+\Lambda(\tau)\\&\ge[\Lambda^*(x_0)-\langle x_0,\tau\rangle]-[\Lambda^*(x)-\langle x,\tau\rangle]>0.\end{aligned}$$

Hence we find that

$$\limsup_{n\to\infty}\frac{1}{n}\log\widehat{P}_n\big(\mathbb{R}^d\setminus B_\epsilon(x)\big)<0\qquad\forall\epsilon>0.$$

Since $\widehat{P}_n(\mathbb{R}^d)=1$ for all n, this implies (V.7) and completes the proof for open sets.

(c) *Removal of E from the lower bound*:

Let $A\subset\mathbb{R}^d$ be non-empty and convex. Then the *relative interior* of A is defined as

$$\operatorname{rint}(A)=\Big\{x\in A\colon\forall y\in A\ \exists\epsilon>0\text{ such that }x-\epsilon(y-x)\in A\Big\}.$$

This is the interior "as viewed from A with disregard for $\mathbb{R}^d$". In particular, $\operatorname{rint}(A)\supset\operatorname{int}(A)$. Note that if $A=\{x\}$, then $\operatorname{int}(A)=\emptyset$ while $\operatorname{rint}(A)=A$.

We need to borrow the following fact from convex analysis.

LEMMA V.8 *If Λ satisfies properties* (1–3) *in Theorem* V.6(c), *then* $E\supset\operatorname{rint}(\mathcal{D}_{\Lambda^*})$, *where* $\mathcal{D}_{\Lambda^*}=\{x\in\mathbb{R}^d\colon\Lambda^*(x)<\infty\}$.

PROOF. See Rockafellar [C7], Corollary 26.4.1. □

REMARKS V.9
1. $\mathcal{D}_{\Lambda^*}$ is a convex set, because Λ^* is a convex function by Lemma V.4(ii).
2. $\operatorname{rint}(\mathcal{D}_{\Lambda^*})\neq\emptyset$, because $\mathcal{D}_{\Lambda^*}\neq\emptyset$ by Lemma V.4(ii).

We can now finish the proof. By Lemma V.8, to show that $\Lambda^*(O\cap E)=\Lambda^*(O)$ it suffices to show that, for any open set $O\subset\mathbb{R}^d$,

$$\Lambda^*\big(O\cap\operatorname{rint}(\mathcal{D}_{\Lambda^*})\big)\le\Lambda^*(O).$$

This being trivially true when $O \cap \mathcal{D}_{\Lambda^*} = \emptyset$ (because both sides are infinite), we may assume that $O \cap \mathcal{D}_{\Lambda^*} \neq \emptyset$. Pick $y \in O \cap \mathcal{D}_{\Lambda^*}$ and $z \in \operatorname{rint}(\mathcal{D}_{\Lambda^*})$. Then, for all $\delta > 0$ sufficiently small, $\delta z + (1-\delta)y \in O \cap \operatorname{rint}(\mathcal{D}_{\Lambda^*})$. It now follows that

$$\Lambda^*\big(O \cap \operatorname{rint}(\mathcal{D}_{\Lambda^*})\big) \le \lim_{\delta \downarrow 0} \Lambda^*\big(\delta z + (1-\delta)y\big) \le \Lambda^*(y),$$

where the last inequality uses that Λ^* is continuous on $\operatorname{rint}(\mathcal{D}_{\Lambda^*})$ and convex on $\mathbb{R}^d$. Taking the infimum over $y \in O \cap \mathcal{D}_{\Lambda^*}$, we get the claim. □

Fig. 9 shows two examples where $E \not\supset \operatorname{rint}(\mathcal{D}_{\Lambda^*})$:

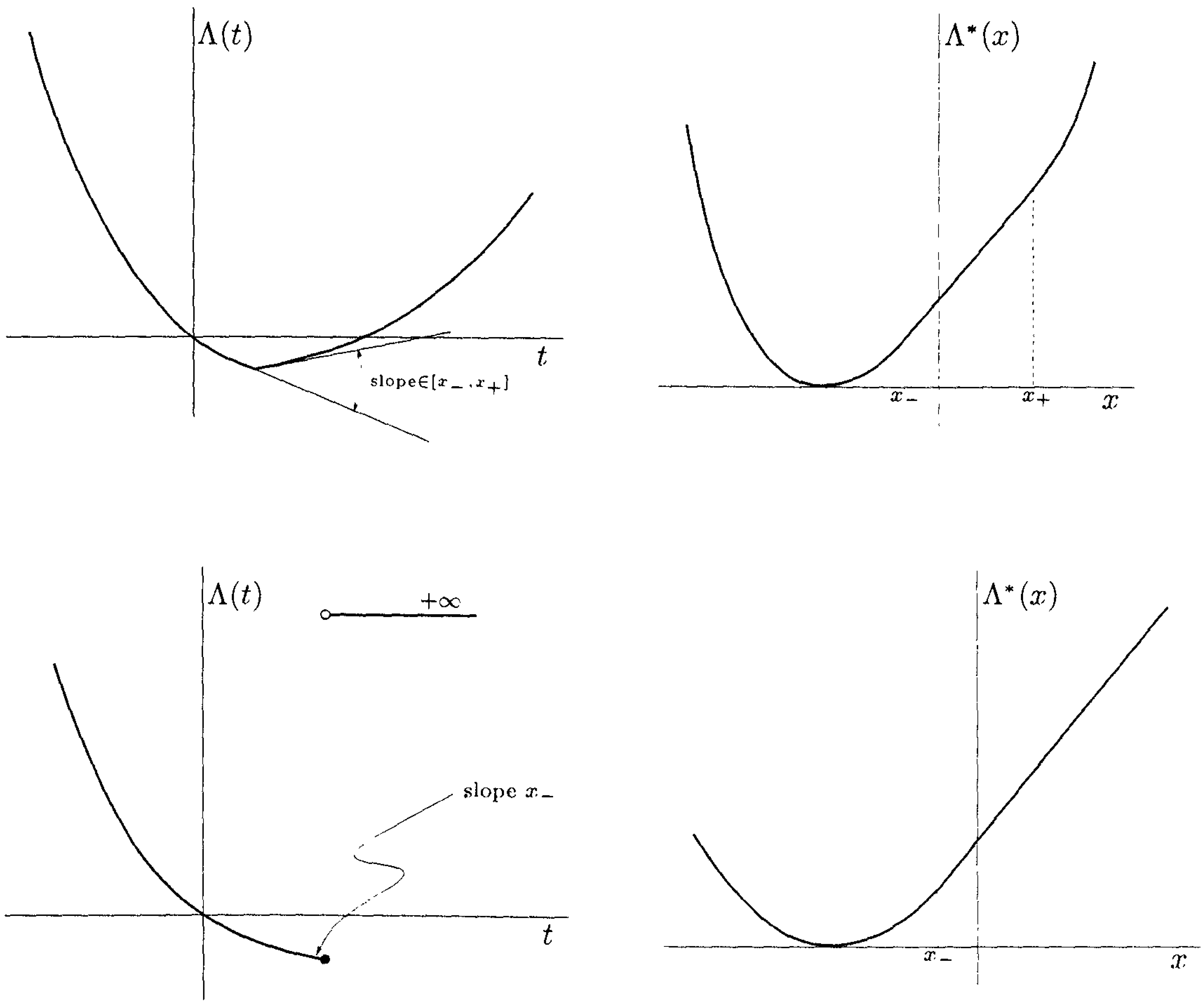

FIG. 9. Necessity of differentiability and steepness in Lemma V.8

The following exercise further elucidates Lemma V.8.

EXERCISE V.10 *Let* $Z_n = (0, Y_n) \in \mathbb{R}^2$ *with* $Y_n = \frac{1}{n}\sum_{i=1}^n X_i$ *and* (X_i) *i.i.d. standard normal random variables on* $\mathbb{R}$. *Show that* $\operatorname{int}(\mathcal{D}_{\Lambda^*}) = \emptyset$, *while* $\operatorname{rint}(\mathcal{D}_{\Lambda^*})$ *is the vertical axis in* $\mathbb{R}^2$.

V.3 Comments

Here are some comments on what has happened in Section V.2.

(1) The Gärtner-Ellis Theorem can be viewed as the analogue on an exponential scale of the route to the CLT via characteristic functions. Namely, just as the CLT can be deduced from scaling of the characteristic function (e.g. for i.i.d. random

variables) the Gärtner-Ellis Theorem shows that large deviations can be deduced from scaling of the cumulant generating function.

EXERCISE V.11 *Prove the* LDP *in Exercise* III.10 *with the help of Theorem* V.6.

(2) If

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

with (X_i) a stationary random sequence, then we may interpret (V.1) as a kind of moderate dependence assumption on (X_i). Consider, for instance, the extreme opposite where $X_i \equiv Y$. Then $\varphi_n \equiv \varphi_Y$, the moment generating function of Y, and it is easily seen that (V.1) always fails, unless Y is degenerate.

EXERCISE V.12 *Let $(X_i)_{i\in\mathbb{Z}}$ be an $\mathbb{R}$-valued stationary mean-zero Gaussian process with covariance function $C_i = \mathbb{E}\, X_0 X_i$, $i \in \mathbb{Z}$, satisfying $\sum_{i\in\mathbb{Z}} |C_i| < \infty$. Let $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $P_n(\cdot) = \mathbb{P}(Z_n \in \cdot)$. Show that (P_n) satisfies the* LDP *on $\mathbb{R}$ with rate n and with rate function $\Lambda^*(x) = x^2/2C$, where $C = \sum_{i\in\mathbb{Z}} C_i$.*

EXERCISE V.13 *Let $(X_i)_{i\in\mathbb{N}}$ be i.i.d. $\mathbb{R}^2$-valued random variables with a marginal law that is bivariate standard normal, i.e., $\mathbb{P}(X_1 \in dx) = \frac{1}{2\pi} e^{-|x|^2/2}$, $x \in \mathbb{R}^2$. Let $\{N(t)\colon t \geq 0\}$ be a rate 1 Poisson process, independent of $(X_i)_{i\in\mathbb{N}}$. Let $Z_n = \frac{1}{n}\sum_{i=1}^{N(n)} X_i$ and $P_n(\cdot) = \mathbb{P}(Z_n \in \cdot)$. Show that (P_n) satisfies the* LDP *on $\mathbb{R}^2$ with rate n and with rate function $\Lambda^*(x) = 1 + |x|(t^* + t^{*-1})$, where $t^* = t^*(|x|)$ solves the equation $|x| = te^{t^2/2}$.*

(3) For i.i.d. sequences the steepness condition can be removed (see Dembo and Zeitouni [A2] Corollary 6.1.6). For dependent sequences, however, this is in general not the case.

(4) Theorem V.6 can be extended to topological vector spaces, in the context of the general setting described in Section III.1 and Comment (7) in Section III.2. For further details, see Dembo and Zeitouni [A2] Section 4.5. Bryc's Theorem mentioned at the end of Section III.3 can be viewed as an infinite-dimensional version of the Gärtner-Ellis Theorem. For further details, see O'Brien and Vervaat [D46].

(5) Theorem V.6(c) fails to capture situations where the rate function is not strictly convex on the relative interior of its domain (see Lemma V.8). Thus, it is not suited to handle sequences with a strong dependence, for which strict convexity may fail, at least not without additional work. Examples will be given in Chapters VII and IX.

(6) Orey [D48] and Orey and Pelikan [D49] prove the LDP for the empirical process associated with stationary and ergodic sequences satisfying certain mixing conditions. The proof uses the LDP for i.i.d. sequences and runs via a Radon-Nikodym argument of the type that was used in Sections IV.1 and IV.2 for Markov sequences. Only in rare cases can the rate function be computed explicitly. See also Deuschel and Stroock [A3] Section 5.4, Dembo and Zeitouni [A2] Section 6.4. Related results for Gibbs measures can be found in Olla [D47], Föllmer and Orey [D21], Comets [D10], and Deuschel, Stroock and Zessin [D16].

V.4 Relation to earlier results

Theorem V.6 gives us the LDP under fairly mild regularity assumptions. The rate function is identified as Λ^*, and is always convex. In special cases it is possible to evaluate Λ and Λ^* explicitly, for instance, in the i.i.d. case and the Markov case. In general, however, they are not easily accessible.

In this section we return to the examples of Chapters I, II and IV to see how these relate to the Gärtner-Ellis Theorem. In this way we come full circle.

i.i.d. case:

1. Let (X_i) be i.i.d. $\mathbb{R}$-valued random variables satisfying either (I.5) or (I.17), (I.18) in Comment (4) in Section I.3. Let $Z_n = \frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then

$$\varphi_n(nt) = \mathbb{E}\left(e^{ntZ_n}\right) = \mathbb{E}\left(e^{t\sum_{i=1}^n X_i}\right) = [\varphi(t)]^n, \qquad t \in \mathbb{R},$$

with φ the moment generating function of X_1. Hence $\Lambda(t) = \log\varphi(t)$, and so Theorem V.6 reduces to Theorem I.4. Condition (V.1) coincides with (I.17). Conditions (c)(1-2) in Theorem V.6 are automatic (recall Exercise I.21), while condition (c)(3) in Theorem V.6 coincides with (I.5) or (I.18).

2. Let (X_i) be i.i.d. Γ-valued random variables satisfying (II.1). Let $Z_n = L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. Then

$$\varphi_n(nt) = \mathbb{E}\left(e^{\langle nt, Z_n\rangle}\right) = \mathbb{E}\left(e^{\langle t, \sum_{i=1}^n \delta_{X_i}\rangle}\right) = \left[\sum_{s=1}^r \rho_s e^{t_s}\right]^n, \qquad t \in \mathbb{R}^r, r = |\Gamma|.$$

Hence $\Lambda(t) = \log(\sum_s \rho_s e^{t_s})$. Under condition (II.1), all the conditions in Theorem V.6 are met and so we get

$$\Lambda^*(\nu) = \sup_{t\in\mathbb{R}^r}\left[\sum_s \nu_s t_s - \log\left(\sum_s \rho_s e^{t_s}\right)\right], \qquad \nu \in \mathfrak{M}_1(\Gamma).$$

Since $\rho_s > 0$ for all s, $t \mapsto \log(\sum_s \rho_s e^{t_s})$ is strictly convex, and so the supremum is attained at t satisfying

$$\nu_s = \rho_s e^{t_s}\left(\sum_u \rho_u e^{t_u}\right)^{-1} \qquad \forall s.$$

Hence $t_s = \log(\frac{\nu_s}{\rho_s}) + C$, for some constant C. Substitution yields

$$\Lambda^*(\nu) = \sum_s \nu_s \log\left(\frac{\nu_s}{\rho_s}\right),$$

where C cancels because $\sum_s \nu_s = \sum_s \rho_s = 1$. This is precisely the relative entropy in Theorem II.2. Thus, Sanov's Theorem is a higher-dimensional version of Cramér's Theorem in disguise.

Markov case:

3. Let (X_i) be a stationary Γ-valued Markov chain satisfying (IV.1). Let $Z_n = L_n^2 = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i,X_{i+1})}$ with periodic boundary conditions. Then

$$\varphi_n(nt) = \mathbb{E}\left(e^{\langle nt, Z_n\rangle}\right) = \sum_{i,j=1}^r \pi_i Q_{ij}^n(t) e^{t_{ji}}, \qquad t \in \mathbb{R}^r \times \mathbb{R}^r, r = |\Gamma|,$$

where $Q(t)$ is the matrix

$$Q_{ij}(t) = P_{ij}e^{t_{ij}}, \qquad i,j \in \Gamma.$$

Perron-Frobenius theory (Seneta [C8] Chapter 1) tells us that

$$\Lambda(t) = \log \lambda(t),$$

with $\lambda(t)$ the unique largest eigenvalue of $Q(t)$. Since $t \mapsto \log \lambda(t)$ is analytic, all the conditions in Theorem V.6 are met and so we get

$$\Lambda^*(\nu) = \sup_{t \in \mathbb{R}^r \times \mathbb{R}^r} \Big[\sum_{i,j} \nu_{ij} t_{ij} - \log \lambda(t) \Big], \qquad \nu \in \widetilde{\mathfrak{M}}_1(\Gamma \times \Gamma),$$

EXERCISE V.14 *Prove that* $t \mapsto \log \lambda(t)$ *is strictly convex.*

By this strict convexity, the supremum is attained at t satisfying

$$\nu_{ij} = \frac{1}{\lambda(t)} \frac{\partial}{\partial t_{ij}} \lambda(t) \qquad \forall i, j.$$

Let $r(t)$ and $l(t)$ denote the right and left eigenvector of $Q(t)$ corresponding to $\lambda(t)$, normalized as $\sum_i l_i(t) r_i(t) = 1$, which are also analytic in t. Then we may compute

$$\begin{aligned} \tfrac{\partial}{\partial t_{ij}} \lambda(t) &= \tfrac{\partial}{\partial t_{ij}} \sum_{k,l} l_k(t) Q_{kl}(t) r_l(t) \\ &= \sum_{k,l} l_k(t) \Big[\tfrac{\partial}{\partial t_{ij}} Q_{kl}(t) \Big] r_l(t) = l_i(t) Q_{ij}(t) r_j(t). \end{aligned}$$

Here two derivatives cancel because $(\partial/\partial t_{ij}) \sum_k l_k(t) r_k(t) = (\partial/\partial t_{ij}) 1 = 0$. Hence, we find that the supremum is attained at t satisfying

$$\nu_{ij} = \frac{1}{\lambda(t)} l_i(t) Q_{ij}(t) r_j(t) \qquad \forall i, j,$$

which gives

$$t_{ij} = \log \Big(\frac{Q_{ij}(t)}{P_{ij}} \Big) = \log \Big(\frac{\lambda(t) \nu_{ij}}{l_i(t) P_{ij} r_j(t)} \Big).$$

Substitution yields

$$\begin{aligned} \Lambda^*(\nu) &= \sum_{i,j} \nu_{ij} \log \Big(\tfrac{\lambda(t) \nu_{ij}}{l_i(t) P_{ij} r_j(t)} \Big) - \log \lambda(t) \\ &= \sum_{i,j} \nu_{ij} \log \Big(\tfrac{\nu_{ij}}{l_i(t) r_i(t) P_{ij}} \tfrac{r_i(t)}{r_j(t)} \Big) = \sum_{i,j} \nu_{ij} \log \Big(\tfrac{\nu_{ij}}{\bar{\nu}_i P_{ij}} \Big), \end{aligned}$$

where $\lambda(t)$ cancels because $\sum_{i,j} \nu_{ij} = 1$, the ratio $r_i(t)/r_j(t)$ cancels because $\sum_j \nu_{ij} = \sum_j \nu_{ji}$, and we use that $\bar{\nu}_i = l_i(t) r_i(t)$. Thus, we end up with precisely the relative entropy in Theorem IV.3.

V.5 Conclusion

What the Gärtner-Ellis Theorem shows is that the rate functions computed in Chapters I, II and IV are all Legendre transforms in disguise. By now the reader will hardly be surprised. The Legendre transform of the moment generating function is what arises naturally as a lower bound in the LDP after we do an exponential Chebyshev estimate and optimize over the parameter in the exponential. The difficult half of the LDP sits in the upper bound. If that holds too, then the rate function must be the Legendre transform.

We have come to the end of the theoretical Part A. It is time to move to the applications.

Part B

APPLICATIONS

CHAPTER VI

STATISTICAL HYPOTHESIS TESTING

In this chapter we give an application of large deviation theory in statistics that goes back to Chernoff [D9]. Namely, we show how Cramér's Theorem can be used in statistical hypothesis testing to assess test optimality.

VI.1 The statistical problem

The following is a standard problem in statistics:

Statistical problem: Let $X_1, \ldots, X_n$ be i.i.d. $\mathbb{R}$–valued random variables with an unknown marginal law μ. Suppose we know that either $\mu = \mu_0$ or $\mu = \mu_1$, where μ_0 and μ_1 are given, but we do not know which. What is the best statistical test to decide, based on the observation of the sample $X_1, \ldots, X_n$, which of the two laws occurs, and how good is this test for large n?

There are two hypotheses:

$$\text{hypothesis } H_0 : \qquad \mu = \mu_0,$$
$$\text{hypothesis } H_1 : \qquad \mu = \mu_1.$$

DEFINITION VI.1 *A decision test is a measurable function $T_n \colon \mathbb{R}^n \to \{0, 1\}$ such that*

$$\begin{aligned} H_0 \textit{ is accepted when } T_n(X_1, \ldots, X_n) &= 0, \\ H_1 \textit{ is accepted when } T_n(X_1, \ldots, X_n) &= 1. \end{aligned}$$

The *performance* of a decision test is determined by the error probabilities

$$\begin{aligned} \alpha_n &= \mathbb{P}(T_n \text{ rejects } H_0 | \mu = \mu_0) &= \mathbb{P}(T_n = 1 | \mu = \mu_0), \\ \beta_n &= \mathbb{P}(T_n \text{ rejects } H_1 | \mu = \mu_1) &= \mathbb{P}(T_n = 0 | \mu = \mu_1), \end{aligned}$$

where $\mathbb{P}$ is the joint law of $X_1, \ldots, X_n$ and μ. The latter is assumed to be distributed according to an *a priori* law on $\{\mu_0, \mu_1\}$. The *Bayesian error probability* is

$$\Delta_n = \alpha_n \mathbb{P}(\mu = \mu_0) + \beta_n \mathbb{P}(\mu = \mu_1).$$

A good decision test will have both α_n and β_n small. A sensible criterion for optimality is to seek a decision test that minimizes β_n subject to a pre-set constraint on α_n, or vice versa. The optimal decision test was identified by Neyman and Pearson (see Definition VI.2 below), and is based on the so-called log-likelihood ratios.

Henceforth we will assume that $\mu_0 \neq \mu_1$ but that they are equivalent (i.e., mutually absolutely continuous w.r.t. each other). Define the likelihood ratios

$$L_{10} = \frac{d\mu_1}{d\mu_0} \qquad \text{and} \qquad L_{01} = \frac{d\mu_0}{d\mu_1}$$

and define new random variables

$$Y_i = \log L_{10}(X_i) = -\log L_{01}(X_i), \qquad i = 1, \ldots, n.$$

DEFINITION VI.2 *A Neyman-Pearson test (NP-test) is a decision test of the form*

$$T_n(X_1, \ldots, X_n) = 1_{\{\frac{1}{n}(Y_1+\cdots+Y_n)>\gamma_n\}}$$

for some $\gamma_n \in \mathbb{R}$.

Definition VI.2 shows that NP-tests are naturally linked with large deviations for empirical averages of i.i.d. random variables, i.e., with Cramér's theorem.

Below we will need the following quantities:

$$\begin{aligned} \gamma_0 &= \mathbb{E}(Y_1 \,|\, \mu = \mu_0) = -H(\mu_0 \,|\, \mu_1), \\ \gamma_1 &= \mathbb{E}(Y_1 \,|\, \mu = \mu_1) = H(\mu_1 \,|\, \mu_0), \end{aligned}$$

where $H(\mu_1 \,|\, \mu_0) = \int d\mu_1 \log \frac{d\mu_1}{d\mu_0}$ denotes the relative entropy of μ_0 with respect to μ_1. Note that $-\infty \le \gamma_0 < 0 < \gamma_1 \le \infty$.

VI.2 Large deviation estimates on test optimality

It is a classical result in statistics that NP-tests are optimal in the following sense: there are no tests with the same value of α_n and a smaller value of β_n, and vice versa. In other words, if we consider the class of all decision tests with a fixed value of α_n, then in this class the NP-test has the smallest value of β_n. The γ_n in this NP-test of course depends on the choice of α_n.

EXERCISE VI.3 *Look up the proof of NP-optimality in the literature.*

The following theorem shows that α_n and β_n decay exponentially fast with n when $\gamma_n \equiv \gamma$.

THEOREM VI.4 *Let $\gamma \in (\gamma_0, \gamma_1)$. Then, for the NP-test with $\gamma_n \equiv \gamma$,*

$$\begin{aligned} \lim_{n\to\infty} \tfrac{1}{n} \log \alpha_n &= -I_0(\gamma) < 0, \\ \lim_{n\to\infty} \tfrac{1}{n} \log \beta_n &= -[I_0(\gamma) - \gamma] < 0, \end{aligned}$$

where I_0 is the Legendre transform of the cumulant generating function of Y_1 given $\mu = \mu_0$, i.e.,

$$I_0(z) = \sup_{t\in\mathbb{R}} \big[zt - \log \varphi_0(t)\big],$$

with

$$\varphi_0(t) = \mathbb{E}\big(e^{tY_1} \,|\, \mu = \mu_0\big).$$

PROOF. The proof is an application of Cramér's Theorem. First note that

$$\varphi_0(t) = \int d\mu_0 \; e^{t \log \frac{d\mu_1}{d\mu_0}}.$$

Hence, for all $0 \le t \le 1$,

$$\varphi_0(t) \le \int_{\{\frac{d\mu_1}{d\mu_0} < 1\}} d\mu_0 + \int_{\{\frac{d\mu_1}{d\mu_0} \ge 1\}} d\mu_1 \le 2,$$

which means that φ_0 exists in a right-neighborhood of the origin. Now, by Definitions VI.1 and VI.2, we have

$$\alpha_n = \mathbb{P}\Big(\frac{1}{n}(Y_1 + \cdots + Y_n) > \gamma \;\Big|\; \mu = \mu_0\Big),$$

so the first claim in Theorem VI.4 follows from the (one-sided version of the) LDP in Theorem I.4 (recall Comment (5) in Section I.4). Note that $\gamma > \gamma_0 = \mathbb{E}(Y_1|\mu = \mu_0)$, so $I_0(\gamma) > 0$ (see Lemma I.14(v)).

Similarly, we have

$$\beta_n = \mathbb{P}\Big(\frac{1}{n}(Y_1 + \cdots + Y_n) \leq \gamma \;\Big|\; \mu = \mu_1\Big)$$

and hence the same LDP applies, now with the moment generating function

$$\varphi_1(t) = \mathbb{E}(e^{tY_1}|\mu = \mu_1),$$

which exists in a left-neighborhood of the origin. However, an easy computation shows that $\varphi_1(t) = \varphi_0(t+1)$, so that $I_1(z) = I_0(z) - z$ for the rate function I_1 associated with φ_1, proving the second claim in Theorem VI.4. Note that $\gamma < \gamma_1 = \mathbb{E}(Y_1|\mu = \mu_1)$, so $I_1(\gamma) > 0$. □

Qualitatively $\log\varphi_0$ looks like (compare with Fig. 3 in Section I.4):

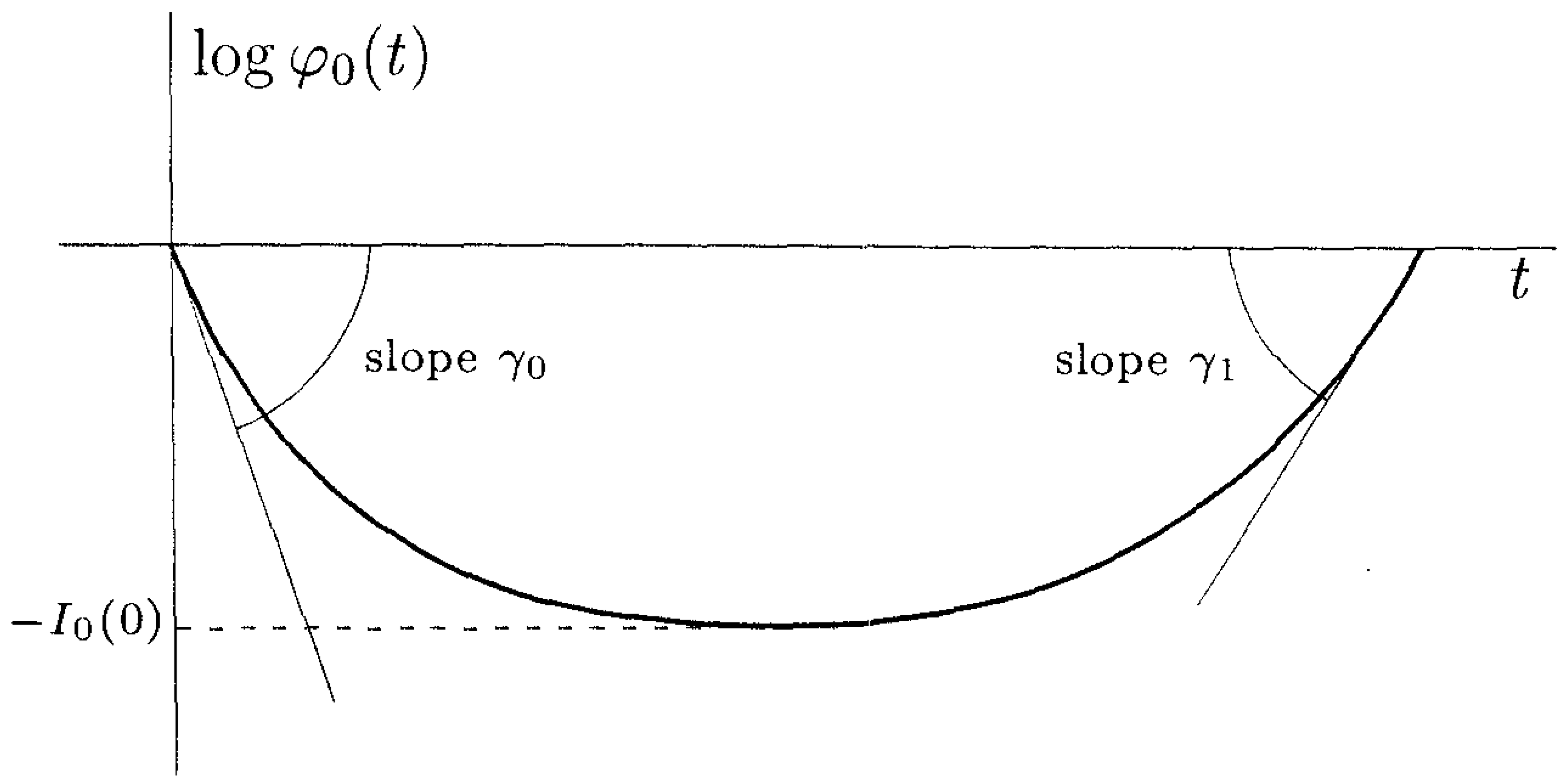

FIG. 10. The graph of $t \mapsto \log\varphi_0(t)$ on $[0,1]$

As a corollary of Theorem VI.4 we obtain:

THEOREM VI.5 *Suppose that* $0 < \mathbb{P}(\mu = \mu_0) < 1$. *Then*

$$\liminf_{n\to\infty} \frac{1}{n} \log\Big(\inf_{T_n} \Delta_n(T_n)\Big) = -I_0(0),$$

where the infimum runs over all decision tests.

PROOF. It suffices to consider NP-tests (as these are optimal). Let α_n^* and β_n^* be the error probabilities for the NP-test with $\gamma_n \equiv \gamma = 0$. Then, for any other NP-test, either $\alpha_n \geq \alpha_n^*$ (when $\gamma_n \leq 0$) or $\beta_n \geq \beta_n^*$ (when $\gamma_n \geq 0$). Hence, by the definition of Δ_n,

$$\begin{aligned}\inf_{T_n} \Delta_n(T_n) &\geq \inf_{T_n} \min\{\alpha_n \mathbb{P}(\mu = \mu_0), \beta_n \mathbb{P}(\mu = \mu_1)\} \\ &\geq \min\{\alpha_n^* \mathbb{P}(\mu = \mu_0), \beta_n^* \mathbb{P}(\mu = \mu_1)\}.\end{aligned}$$

Since $0 < \mathbb{P}(\mu = \mu_0) = 1 - \mathbb{P}(\mu = \mu_1) < 1$, it follows from (I.2) that

$$\liminf_{n\to\infty} \frac{1}{n} \log\Big(\inf_{T_n} \Delta_n(T_n)\Big) \geq \min\Big\{\lim_{n\to\infty} \frac{1}{n} \log \alpha_n^*, \lim_{n\to\infty} \frac{1}{n} \log \beta_n^*\Big\} = -I_0(0).$$

Moreover, equality is obtained for the NP-test with $\gamma_n \equiv \gamma = 0$. □

Theorem VI.5 shows that the Bayesian error probability is smallest, in the limit as $n \to \infty$, for the *zero threshold* NP-test. This test has an exponential rate $I_0(0) = -\log \inf_{t\in\mathbb{R}} \varphi_0(t) = I_1(0) = -\log \inf_{t\in\mathbb{R}} \varphi_1(t)$ that is best possible. This rate is called the Chernoff information of the laws μ_0 and μ_1.

EXERCISE VI.6 *Compute $I_0(0)$ when μ_0, μ_1 are Exponential with parameters $\delta_0, \delta_1 > 0$.*

There are many variations on the statistical hypothesis testing problem described above. For instance, we may be given μ_0 but not μ_1, and only know that μ_1 is drawn from a given class. Or the sample $X_1, \ldots, X_n$ may be drawn from a Markov chain with transition kernel P_0 or P_1. For more details, we refer the reader to the statistics literature. See also Dembo and Zeitouni [A2] Sections 3.5 and 7.1.

CHAPTER VII

RANDOM WALK IN RANDOM ENVIRONMENT

In this chapter we give an application of large deviation theory in probability. Namely, we consider a one-dimensional random walk with *random local drift* and derive the LDP for the speed of the random walk conditioned on the drift field. As we will see, a rather rich behavior for the rate function shows up, depending on the choice of the probability law governing the drift field. The results described below were first derived in Greven and den Hollander [D32]. For the proof, however, we will follow a different and simpler line of argument, due to Comets, Gantert and Zeitouni [D11]. The LDP will be derived with the help of the Gärtner-Ellis Theorem, and we will arrive at an interesting characterization of the rate function in terms of random continued fractions.

VII.1 Random drifts

Let

$$\omega = (\omega_x)_{x\in\mathbb{Z}}$$

be an i.i.d. collection of $(0,1)$-valued random variables with common distribution α. For fixed ω, let

$$X = (X_n)_{n\in\mathbb{N}_0}$$

be the Markov chain on $\mathbb{Z}$, starting at $X_0 = 0$, with transition probabilities

$$\mathbb{P}_\omega(X_{n+1} = y | X_n = x) = \begin{cases} \omega_x & \text{if } y = x+1, \\ 1-\omega_x & \text{if } y = x-1, \\ 0 & \text{otherwise.} \end{cases}$$

Here the symbol $\mathbb{P}_\omega$ denotes the probability law on path space given ω. The process (X, ω) is an example of a *random walk in random environment*, and X has probability law $\mathbb{P} = \int \alpha^{\mathbb{Z}}(d\omega)\mathbb{P}_\omega$.

Henceforth we abbreviate

$$\rho_x = \frac{1-\omega_x}{\omega_x}, \qquad \langle f \rangle = \int f(\omega)\, \alpha^{\mathbb{Z}}(d\omega),$$

and we write $\rho = \rho_0$. Furthermore, we assume that α is non-degenerate and that $\text{supp}(\alpha)$ is bounded away from 0 and 1. When we write "ω-a.s." we mean "for all ω in a set of measure 1 under $\alpha^{\mathbb{Z}}$".

The above model has been studied extensively in the literature. We cite two results that are relevant for our discussion.

Solomon [D54] proved that X is ω-a.s.

$$\begin{array}{ll} \text{recurrent} & \text{if } \langle \log \rho \rangle = 0, \\ \text{transient to the left} & \text{if } \langle \log \rho \rangle > 0, \\ \text{transient to the right} & \text{if } \langle \log \rho \rangle < 0. \end{array}$$

Moreover,

$$\lim_{n\to\infty} \frac{1}{n} X_n = v_\alpha \quad \text{exists and is constant } \mathbb{P}\text{-a.s.},$$

with

$$v_\alpha = \begin{cases} \frac{1-\langle \rho \rangle}{1+\langle \rho \rangle} & \text{if } \langle \rho \rangle < 1, \\ -\frac{1-\langle \rho^{-1} \rangle}{1+\langle \rho^{-1} \rangle} & \text{if } \langle \rho^{-1} \rangle < 1, \\ 0 & \text{if } \langle \rho \rangle^{-1} \leq 1 \leq \langle \rho^{-1} \rangle. \end{cases}$$

This result identifies the speed of the random walk depending on the choice of α. The speed is of course zero in the recurrent case. But, interestingly, in the transient case it can be either zero or non-zero. Note that, by Jensen's inequality,

$$\log \langle \rho^{-1} \rangle^{-1} < \langle \log \rho \rangle < \log \langle \rho \rangle.$$

Hence recurrence, transience and sign of the speed all depend on where the value 0 is located with respect to the three values in the last display. The strict inequalities hold because α is non-degenerate. Therefore the zero vs. non-zero speed dichotomy in the transient case occurs *only* in the random environment situation.

Sinai [D53] proved that in the recurrent case

$$\frac{\sigma^2}{(\log n)^2} X_n \underset{n\to\infty}{\longrightarrow} Z \qquad \text{in law w.r.t. } \mathbb{P},$$

with $\sigma^2 = \langle (\log \rho)^2 \rangle \in (0, \infty)$ and with Z some random variable that is independent of α. This result says that the random walk is extremely subdiffusive: the random local drifts slow it down to a logarithmic scale. The law of Z has been identified by Kesten [D41]:

$$\mathbb{P}(Z \in dx) = z(x)dx \quad \text{with} \quad z(x) = \frac{2}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\Big[-\frac{\pi^2}{8}(2k+1)^2 |x| \Big].$$

(Incidentally, $2z(t)$, $t > 0$, is the probability that standard Brownian motion starting at 0 does not exit the interval $[-1, 1]$ prior to time t.)

Several refinements of the above results have been derived over the years. We refer the reader to the literature.

VII.2 The LDP for the speed

Our goal in the rest of this chapter will be to derive the LDP for the speed of the random walk *conditioned* on ω. Throughout what follows we will assume that both positive and negative drifts occur in the random environment:

$$\text{supp}(\alpha) \cap (0, \tfrac{1}{2}] \neq \emptyset, \quad \text{supp}(\alpha) \cap [\tfrac{1}{2}, 1) \neq \emptyset. \tag{VII.1}$$

In other words, the random walk is thoroughly confused. In Section VII.8 we will see what happens when (VII.1) fails. We will also assume that

$$\langle \log \rho \rangle \leq 0, \tag{VII.2}$$

i.e., the random walk is either recurrent or is transient to the right. The opposite case can be obtained via mirror reflection (i.e., $\rho_x \to \rho_{-x}^{-1}$, $X_n \to -X_n$). Our main results read:

THEOREM VII.3 *Let* $P_n^\omega(\cdot) = \mathbb{P}_\omega(\frac{1}{n}X_n \in \cdot)$. *Then,* ω-*a.s.,* (P_n^ω) *satisfies the* LDP *on* $\mathbb{R}$ *with rate* n *and with a deterministic (!) rate function* I *that can be computed in terms of a variational problem.*

THEOREM VII.4 *It follows from the solution of the variational problem that (see Fig. 11):*
(a) I *is continuous and convex on* $[-1,1]$ *and infinite elsewhere.*
(b) $I(-\theta) = I(\theta) - \theta\langle\log\rho\rangle$ *for* $\theta \in (0,1]$.
(c) I *is zero on* $[0, v_\alpha]$ *and strictly positive on* $(v_\alpha, 1]$.
(d) I *is strictly convex and analytic on* $(v_\alpha, 1)$.

Here are qualitative pictures of $\theta \mapsto I(\theta)$ on $[-1,1]$ in the three respective cases:

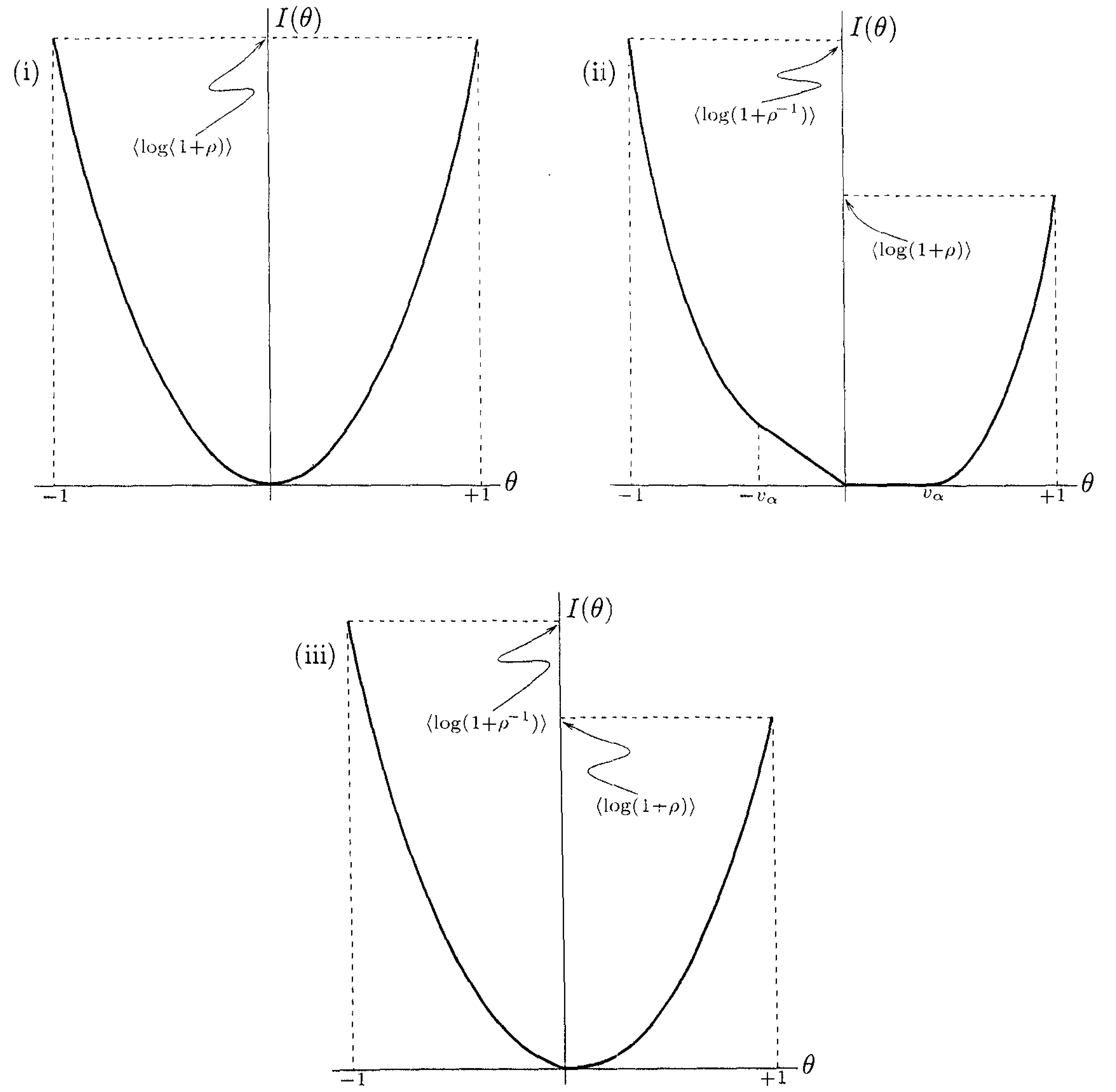

FIG. 11. (i) recurrent; (ii) transient: positive speed; (iii) transient: zero speed

Thus, we see that both in the recurrent case and in the transient case with zero speed the rate function has a unique zero at $\theta = 0$ and is strictly convex everywhere. On the other hand, in the transient case with positive speed the rate function has two linear pieces: one horizontal piece for $\theta \in [0, v_\alpha]$ and one tilted piece for $\theta \in [-v_\alpha, 0]$ (see Fig. 11).

The flat piece means that speeds *smaller* than the typical speed v_α are not exponentially costly. This may be explained as follows. Because of our assumption (VII.1), the random environment contains arbitrarily long stretches of sites where the local drifts are either neutral or point to the center of the stretch. Such stretches tend to "trap" the walker. The longer the stretch, the easier it is for the walker to "loose time inside". Between 0 and θn the longest stretches have a length of order $\log n$, and for the walker to loose a time of order n in those stretches has a cost that is subexponential in n.

The tilted piece has a similar explanation. In fact, Theorem VII.4(b) shows that there is a symmetry relation between forward and backward speeds. In the recurrent case the rate function is symmetric, in the transient case it is not. Note that in the transient case the slope of the rate function is discontinuous at $\theta = 0$.

EXERCISE VII.5 *Show that* $I(1) = \langle \log(1+\rho) \rangle$ *and* $I(-1) = \langle \log(1+\rho^{-1}) \rangle$.

The proofs of Theorems VII.3 and VII.4 will be given in four steps, organized as Sections VII.3–VII.6 below.

VII.3 The LDP for the hitting times

The idea behind the proof of the LDP in Theorem VII.3 is that we first look at how fast the random walk manages to reach points to the right of the origin. Let

$$T_k = \inf\{n \geq 0 \colon X_n = k\}, \qquad k \in \mathbb{N}_0,$$

be the hitting times to the right of the origin. Put

$$\tau_k = T_k - T_{k-1}, \qquad k \in \mathbb{N},$$

and define

$$\varphi(r, \omega) = \mathbb{E}_\omega(e^{r\tau_1}), \qquad r \in \mathbb{R}.$$

LEMMA VII.6 *Let* $Q_k^\omega(\cdot) = \mathbb{P}_\omega(\frac{1}{k}T_k \in \cdot)$. *Then,* ω*-a.s.,* (Q_k^ω) *satisfies the weak* LDP *on* $\mathbb{R}$ *with rate* k *and with weak rate function*

$$J(u) = \sup_{r \in \mathbb{R}} [ur - \log \lambda(r)], \qquad u \in \mathbb{R}, \tag{VII.7}$$

where

$$\log \lambda(r) = \langle \log \varphi(r, \cdot) \rangle = \int \alpha^{\mathbb{Z}}(d\omega) \log \varphi(r, \omega), \qquad r \in \mathbb{R}. \tag{VII.8}$$

PROOF. The key observation is that, for fixed ω, $(\tau_k)_{k \in \mathbb{N}}$ is a sequence of *independent* random variables. However, they are not identically distributed: the law of τ_k depends on all the ω_x with $x < k$. But, because $\alpha^{\mathbb{Z}}$ is ergodic under translations, it is possible to handle the inhomogeneity as follows.

Let $\Lambda_k^\omega(r) = \log \mathbb{E}_\omega(e^{rT_k})$. Then

$$\Lambda_k^\omega(r) = \log \mathbb{E}_\omega(e^{r(\tau_1 + \cdots + \tau_k)}) = \log \prod_{l=1}^{k} \mathbb{E}_\omega(e^{r\tau_l}) = \log \prod_{l=1}^{k} \varphi(r, \sigma^{l-1}\omega),$$

where σ is the left-shift acting on ω (i.e., $(\sigma\omega)_x = \omega_{x+1}$). It follows from Birkhoff's Ergodic Theorem that

$$\lim_{k\to\infty} \frac{1}{k}\Lambda_k^\omega(r) = \log\lambda(r) \qquad \omega - \text{a.s.}$$

In Section VII.6 we will see that $r \mapsto \log\lambda(r)$ is lower semi-continuous on $\mathbb{R}$ and differentiable on the interior of its domain $[-\infty, 0]$ (see Fig. 12 below). Hence, the claim follows from (the one-sided version of) Theorem V.6 (the Gärtner-Ellis Theorem): J is the Legendre transform of $\log\lambda$. Note here that Theorem V.6 is applied for fixed ω, but that *the rate function is ω-a.s. constant.* In Section VII.6 we will also see that the level sets of J are closed but not compact (see Fig. 13 below), which is why we only get a weak LDP. □

There is an analogous weak LDP for the hitting times to the left of the origin. This follows from the same type of argument, leading to the weak rate function

$$\widetilde{J}(u) = \sup_{r\in\mathbb{R}}[ur - \log\widetilde{\lambda}(r)], \qquad u \in \mathbb{R},$$

where

$$\log\widetilde{\lambda}(r) = \langle \log\widetilde{\varphi}(r,\cdot)\rangle = \int \alpha^{\mathbb{Z}}(d\omega)\log\widetilde{\varphi}(r,\omega), \qquad r \in \mathbb{R},$$

with

$$\widetilde{\varphi}(r,\omega) = \mathbb{E}_\omega(e^{r\mathcal{T}_{-1}} 1_{\{\mathcal{T}_{-1}<\infty\}}).$$

(The extra indicator is needed when the random walk is transient; recall (VII.2).) The following symmetry relation will be needed later on.

LEMMA VII.9 $\log\widetilde{\lambda}(r) = \log\lambda(r) + \langle\log\rho\rangle$ *for all* $r < 0$.

The proof is deferred to the end of Section VII.5.

VII.4 From hitting times to speed

Since

$$X_{T_k} = k, \qquad k \in \mathbb{N}_0,$$

it is possible to invert the weak LDP in Lemma VII.6 to obtain the LDP in Theorems VII.3 and VII.4. Roughly, the idea is that

$$\{X_n \approx \lceil\theta n\rceil\} \approx \{T_{\lceil\theta n\rceil} \approx n\} \qquad \Longrightarrow \qquad e^{-nI(\theta)} \approx e^{-\theta n J(\frac{1}{\theta})},$$

which leads to the relation $I(\theta) = \theta J(\frac{1}{\theta})$.

LEMMA VII.10 *Theorem* VII.3 *holds with the rate function* I *identified as:*
(i) $I(\theta) = \theta\sup_{r\in\mathbb{R}}[\frac{r}{\theta} - \log\lambda(r)]$ *for* $\theta \in (0,1]$.
(ii) $I(-\theta) = I(\theta) - \theta\langle\log\rho\rangle$ *for* $\theta \in (0,1]$.
(iii) $I(0) = 0$ *and* $I(\theta) = \infty$ *for* $\theta \notin [-1,1]$.

PROOF. (i) Fix $\theta \in (0,1)$. Clearly,

$$\mathbb{P}_\omega(X_n \geq \lfloor\theta n\rfloor) \leq \mathbb{P}_\omega(T_{\lfloor\theta n\rfloor} \leq n).$$

Hence, Lemma VII.6 gives us the upper bound

$$\limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}_\omega(X_n \geq \lfloor\theta n\rfloor) \leq -\theta \inf_{u\leq\frac{1}{\theta}} J(u) = -\theta J\Big(\frac{1}{\theta}\Big),$$

where we use that J is continuous on $(1,\infty)$ and non-increasing on $\mathbb{R}$ (see Fig. 13 below). Similarly, for any $\epsilon > 0$,

$$\mathbb{P}_\omega(X_n \geq \lfloor \theta n \rfloor) \geq \mathbb{P}_\omega\big(n \leq T_{\lfloor \theta n \rfloor + \lfloor \epsilon n \rfloor} \leq n + \lfloor \epsilon n \rfloor\big).$$

Hence, Lemma VII.6 also gives us the lower bound

$$\begin{aligned}&\liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}_\omega(X_n \geq \lfloor \theta n \rfloor)\\ &\geq -(\theta+\epsilon) \inf_{\frac{1}{\theta+\epsilon} \leq u \leq \frac{1+\epsilon}{\theta+\epsilon}} J(u) = -(\theta+\epsilon) J(\tfrac{1+\epsilon}{\theta+\epsilon}) \underset{\epsilon \downarrow 0}{\longrightarrow} -\theta J(\tfrac{1}{\theta}).\end{aligned}$$

Thus, we have proved the claim for large deviations in the upward direction. An analogous argument works for large deviations in the downward direction. Hence, we have the LDP for semi-intervals (the case $\theta = 1$ can be incorporated by hand because of Exercise VII.5). The LDP for arbitrary open and closed sets now follows easily because I is convex (see Fig. 11).
(ii) It follows from Lemma VII.9 that $\widetilde{J}(u) = J(u) - \langle \log \rho \rangle$, $u \in \mathbb{R}$. By part (i) of the lemma, we have

$$I(-\theta) = \theta \widetilde{J}\Big(\frac{1}{\theta}\Big), \qquad \theta \in (0,1],$$

from which the claim follows.
(iii) Since $\mathrm{supp}(\alpha)$ is bounded away from 0 and 1, it can be proved, via a probabilistic argument applied directly to the random walk, that $I(0) = \lim_{\theta \to 0} I(\theta)$.

EXERCISE VII.11 *Give the proof of the last relation.*

The claim that I is infinite outside $[-1,1]$ is obvious. □

VII.5 Random continued fractions

It turns out that the cumulant generating function $\log \varphi(r,\omega)$ has a beautiful representation in terms of a continued fraction. Even though this representation may seem a bit prohibitive, it is actually quite helpful for numerical calculations.

LEMMA VII.12 (i) *For any $r \leq 0$,*

$$\varphi(r,\omega) = \cfrac{1}{e^{-r}(1+\rho_0) - \cfrac{\rho_0}{e^{-r}(1+\rho_{-1}) - \cfrac{\rho_{-1}}{\cdots}}},$$

with this continued fraction converging ω-a.s.
(ii) *For any $r > 0$, $\varphi(r,\omega) = \infty$ ω-a.s.*

PROOF. (i) It is obvious from the definition of $\varphi(r,\omega)$ that $0 < \varphi(r,\omega) \leq 1$ for all $r \leq 0$ and all ω. We decompose according to the first step of the random walk:

$$\tau_1 = 1_{\{X_1=1\}} + 1_{\{X_1=-1\}}(1 + \tau_1^* + \tau_1^{**}). \qquad \text{(VII.13)}$$

Here

- τ_1^* is the first hitting time of 0 for the path starting at -1, which has probability law $\mathbb{P}_{\sigma^{-1}\omega}(\tau_1 \in \cdot\,)$.
- τ_1^{**} is the first hitting time of 1 for the path starting at 0, which has probability law $\mathbb{P}_\omega(\tau_1 \in \cdot\,)$ and is independent of τ_1^* given that $\tau_1^* < \infty$.

Therefore we compute

$$\varphi(r,\omega) = \mathbb{E}_\omega(e^{r\tau_1}) = \omega_0 e^r + (1-\omega_0)e^r\varphi(r,\sigma^{-1}\omega)\varphi(r,\omega) \qquad \text{(VII.14)}$$

(σ^{-1} is the right-shift). Since $\rho_0 = (1-\omega_0)/\omega_0$, this relation can be rewritten as the recursion

$$\varphi(r,\omega) = \frac{1}{e^{-r}(1+\rho_0) - \rho_0\varphi(r,\sigma^{-1}\omega)}, \qquad \text{(VII.15)}$$

which can be iterated to produce the continued fraction claimed above (convergence follows by standard estimates).

(ii) Fix $r > 0$. Let $A = \{\omega\colon \varphi(r,\omega) = \infty\}$. We see from (VII.14) that $\omega \in A$ when $\sigma^{-1}\omega \in A$. Since $\alpha^{\mathbb{Z}}$ is ergodic under σ, it follows that $\alpha^{\mathbb{Z}}(A) = 0$ or 1. We will show that $\alpha^{\mathbb{Z}}(A) > 0$ and thereby complete the proof.

For $N \in \mathbb{N}$, let $B_N = \{\omega\colon \omega_x \leq \frac{1}{2}$ for $x = 0,-1,\ldots,-N\}$. Then $\alpha^{\mathbb{Z}}(B_N) > 0$ for all $N \in \mathbb{N}$ by (VII.1). Next, an easy comparison argument gives that

$$\varphi(r,\omega) = \mathbb{E}_\omega\left(e^{r\tau_1}\right) \geq \mathbb{E}_{\omega(N)}\left(e^{r\tau_1}\right) \qquad \forall \omega \in B_N,$$

where $\omega(N)$ is the environment given by $\omega(N)_x = \frac{1}{2}$ for $x = 0,-1,\ldots,-N$ and $\omega(N)_x = \omega_x$ for all other x. Because SRW on $\mathbb{Z}$ is null-recurrent, the RHS tends to infinity as $N \to \infty$, uniformly in $\omega \in B_N$. Hence, for any $K > 0$, there exists an $N_0 = N_0(r,K)$ such that $\varphi(r,\omega) \geq K$ for all $\omega \in B_{N_0}$. Thus, $\varphi(\sigma^{-1}r,\omega) \geq K$ for all $\omega \in B_{N_0+1}$, and it now follows from (VII.14) that

$$\varphi(r,\omega) \geq e^r\left[\omega_0 + (1-\omega_0)K\varphi(r,\omega)\right] \geq \frac{1}{2}e^r K\varphi(r,\omega) \qquad \forall \omega \in B_{N_0+1}.$$

However, this is a contradiction when $\varphi(r,\omega) < \infty$ and $K \geq 2e^{-r}$. For such K we therefore have $B_{N_0+1} \subset A$, and so $\alpha^{\mathbb{Z}}(A) > 0$. □

We see from Lemma VII.12 how $\varphi(r,\omega)$ depends on $(\omega_{-x})_{x\in\mathbb{N}_0}$. For $r \leq 0$, the dependence on ω_x decays as $x \to -\infty$ because the continued fraction converges. For $r < 0$ this decay is actually exponentially fast in x (see Greven and den Hollander [D32]).

VII.6 Analysis of the rate functions

In this section we list some important properties of $\log\lambda(r)$ and show how these can be used to prove Theorem VII.4.

LEMMA VII.16 *(See Fig. 12.)*
(i) $r \mapsto \log\lambda(r)$ *is continuous, strictly increasing and strictly convex on* $(-\infty,0]$.
(ii) $r \mapsto \log\lambda(r)$ *is analytic on* $(-\infty,0)$.
(iii) $\log\lambda(0) = 0$ *and* $\lim_{r\to-\infty}\log\lambda(r) = -\infty$.
(iv) $\lim_{r\uparrow 0}\frac{d}{dr}\log\lambda(r) = \frac{1}{v_\alpha}$ *and* $\lim_{r\to-\infty}\frac{d}{dr}\log\lambda(r) = 1$.

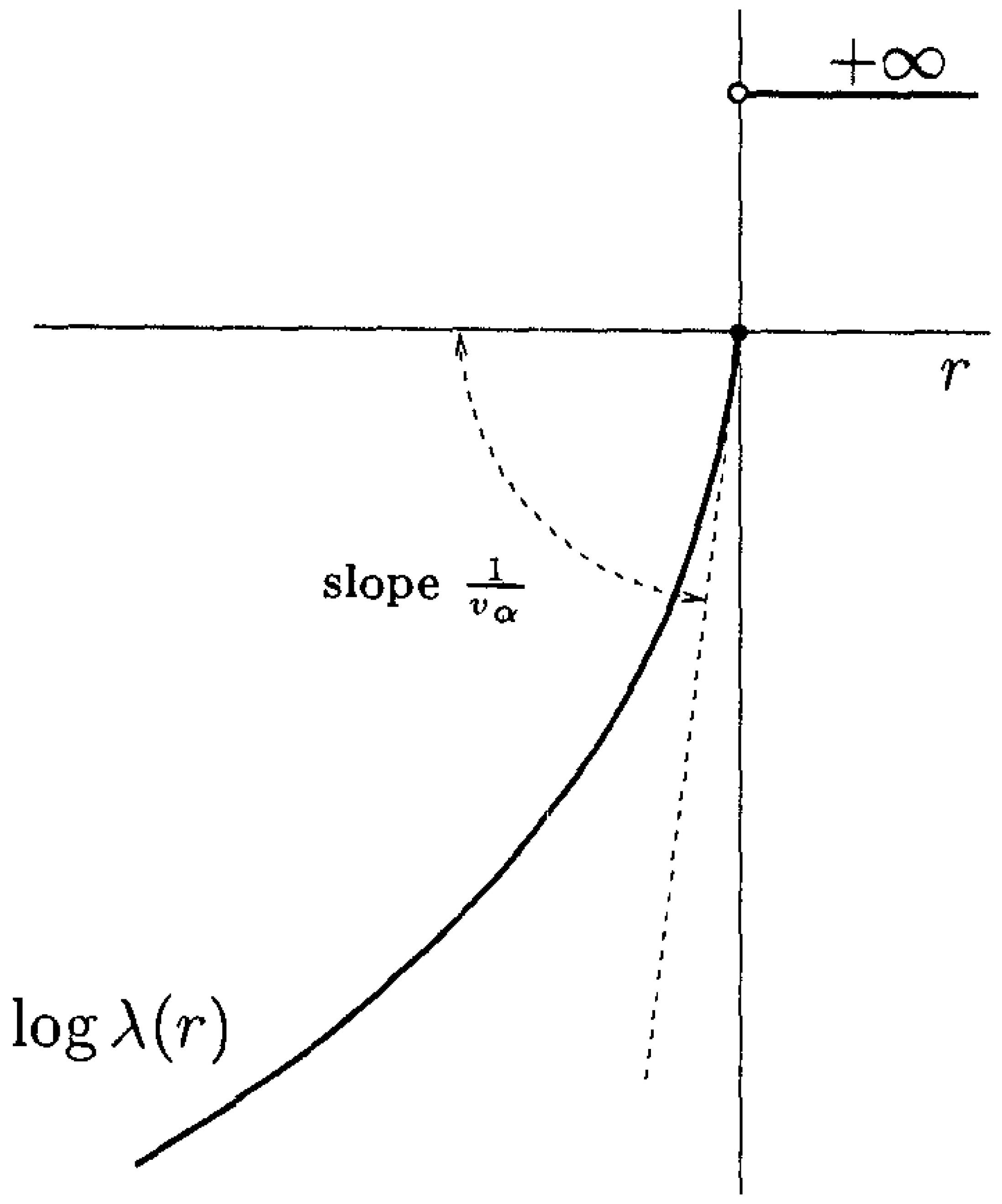

FIG. 12. The function $r \mapsto \log\lambda(r)$

PROOF. (i–iii) Immediate from the definitions. These properties are obvious for $r \mapsto \log\varphi(r,\omega)$ and are preserved under taking the average over ω. In particular, the analyticity is preserved because $\omega \mapsto \log\varphi(r,\omega)$ is uniformly integrable for $r < 0$.
(iv) Note that

$$\frac{d}{dr}\log\lambda(r) = \left\langle \frac{\partial}{\partial r}\log\varphi(r,\cdot)\right\rangle,$$

with

$$\frac{\partial}{\partial r}\log\varphi(r,\omega) = \frac{\mathbb{E}_\omega(\tau_1 e^{r\tau_1})}{\mathbb{E}_\omega(e^{r\tau_1})}.$$

Since $\tau_1 \geq 1$, the claim for $r \to -\infty$ follows immediately. Letting $r \uparrow 0$, we get

$$\lim_{r\uparrow 0}\frac{d}{dr}\log\lambda(r) = \langle \mathbb{E}_\omega(\tau_1)\rangle.$$

Now return to (VII.13). Averaging over the random walk, we have

$$\mathbb{E}_\omega(\tau_1) = \omega_0 + (1-\omega_0)[1 + \mathbb{E}_{\sigma^{-1}\omega}(\tau_1) + \mathbb{E}_\omega(\tau_1)]$$

or

$$\mathbb{E}_\omega(\tau_1) = (1+\rho_0) + \rho_0\mathbb{E}_{\sigma^{-1}\omega}(\tau_1).$$

Averaging over ω, we get

$$\langle\mathbb{E}_\omega(\tau_1)\rangle = 1 + \langle\rho\rangle + \langle\rho\rangle\ \langle\mathbb{E}_\omega(\tau_1)\rangle,$$

from which it follows that $\langle\mathbb{E}_\omega(\tau_1)\rangle = \frac{1+\langle\rho\rangle}{1-\langle\rho\rangle} = \frac{1}{v_\alpha}$. □

Using Lemma VII.16, we arrive at the following picture for the weak rate function J in Lemma VII.6:

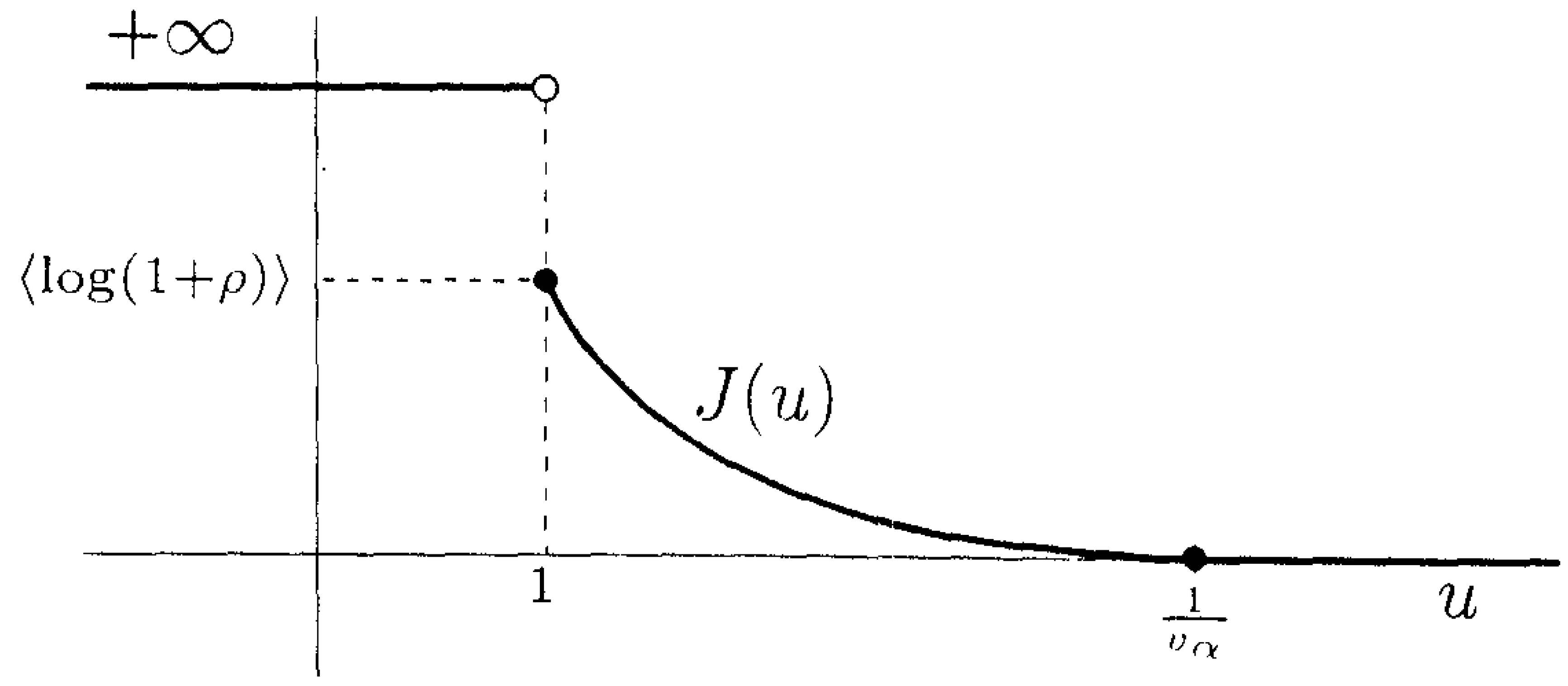

FIG. 13. The weak rate function for the hitting times

EXERCISE VII.17 *Show that* $J(1) = \langle\log(1+\rho)\rangle$. *Compare with Exercise* VII.5.

With the help of the relation $I(\theta) = \theta J(\frac{1}{\theta})$ for $\theta \in (0, 1]$ established in Lemma VII.10 we now also get the three pictures in Fig. 11, including all the properties stated in Theorem VII.4. We see from Theorem VII.10(i) that, for $\theta \in (v_\alpha, 1)$,

$$\begin{aligned} I(\theta) &= r(\theta) - \theta\log\lambda\big(r(\theta)\big), \\ I'(\theta) &= -\log\lambda\big(r(\theta)\big), \\ I''(\theta) &= -\tfrac{1}{\theta}r'(\theta), \end{aligned} \tag{VII.18}$$

with $r(\theta)$ the unique solution of the equation

$$\frac{1}{\theta} = \frac{d}{dr}\log\lambda(r). \tag{VII.19}$$

These relations show how to read off Fig. 11 from Fig. 12.

We conclude this section with a proof of Lemma VII.9.

PROOF. The same type of splitting as in (VII.13) gives the mirror version of (VII.14):

$$\widetilde{\varphi}(r,\omega) = \mathbb{E}_\omega(e^{r\tau_{-1}}1_{\{\tau_{-1}<\infty\}}) = (1-\omega_0)e^r + \omega_0 e^r\widetilde{\varphi}(r,\sigma\omega)\widetilde{\varphi}(r,\omega). \tag{VII.20}$$

Combining (VII.14) and (VII.20), we get after an easy computation that

$$\frac{\rho_0\varphi(r,\omega)}{\widetilde{\varphi}(r,\omega)} = \frac{1-\varphi(r,\omega)\widetilde{\varphi}(r,\sigma\omega)}{1-\varphi(r,\sigma^{-1}\omega)\widetilde{\varphi}(r,\omega)}.$$

Taking logarithms and averaging over ω on both sides, we see that the RHS vanishes and so we get the symmetry relation in Lemma VII.9. Integrability is in order because $r < 0$. □

VII.7 Comparison with homogeneous random walk

The following inequality has an interesting consequence.

LEMMA VII.21 *Let* $\eta = \exp\langle\log\rho\rangle$. *Then, for all* $r < 0$,

$$\frac{d}{dr}\log\lambda(r) > \frac{1+\eta\lambda^2(r)}{1-\eta\lambda^2(r)}.$$

PROOF. By differentiating (VII.15) with respect to r, we get

$$\frac{\partial}{\partial r}\varphi(r,\omega)=\varphi^2(r,\omega)\Big[e^{-r}(1+\rho_0)+\rho_0\frac{\partial}{\partial r}\varphi(r,\sigma^{-1}\omega)\Big].$$

Using (VII.15) once more to write $e^{-r}(1+\rho_0)=\varphi^{-1}(r,\omega)+\rho_0\varphi(r,\sigma^{-1}\omega)$, we obtain the recursion relation

$$\frac{\partial}{\partial r}\log\varphi(r,\omega)=1+\rho_0\varphi(r,\omega)\varphi(r,\sigma^{-1}\omega)\Big[1+\frac{\partial}{\partial r}\log\varphi(r,\sigma^{-1}\omega)\Big].$$

Iteration yields

$$\frac{\partial}{\partial r}\log\varphi(r,\omega)=1+2\sum_{x=-\infty}^{0}\prod_{y=x}^{0}\rho_y\varphi(r,\sigma^y\omega)\varphi(r,\sigma^{y-1}\omega).$$

Taking the expectation over ω, we get

$$\begin{aligned}\tfrac{d}{dr}\log\lambda(r) &= 1+2\sum_{x=-\infty}^{0}\Big\langle\prod_{y=x}^{0}\rho_y\varphi(r,\sigma^y\omega)\varphi(r,\sigma^{y-1}\omega)\Big\rangle\\ &> 1+2\sum_{x=-\infty}^{0}\exp\Big\{\sum_{y=x}^{0}\Big\langle\log[\rho_y\varphi(r,\sigma^y\omega)\varphi(r,\sigma^{y-1}\omega)]\Big\rangle\Big\}\\ &= 1+2\sum_{x=-\infty}^{0}\exp\Big\{(|x|+1)[\langle\log\rho\rangle+2\log\lambda(r)]\Big\}\\ &= 1+2\sum_{x=-\infty}^{0}[\eta\lambda^2(r)]^{|x|+1}\\ &= \tfrac{1+\eta\lambda^2(r)}{1-\eta\lambda^2(r)},\end{aligned}$$

where the strict inequality holds because α is non-degenerate. □

Lemma VII.21 implies the following bound on I. For $\eta\in(0,1]$, let $\lambda_\eta(r)$, $r_\eta(\theta)$ and $I_\eta(\theta)$ denote the analogous quantities for the *homogeneous* random walk with $\rho_x\equiv\eta$, i.e., the random walk with drift $\frac{1-\eta}{1+\eta}$.

LEMMA VII.22 *Let $\eta=\exp\langle\log\rho\rangle$. Then, for all $\theta\in(v_\alpha,1)$,*

$$\begin{aligned}\lambda(r(\theta)) &< \lambda_\eta(r_\eta(\theta)),\\ I'(\theta) &> I'_\eta(\theta).\end{aligned}$$

PROOF. Use Lemma VII.21 twice in combination with (VII.19), to write

$$\frac{1+\eta\lambda_\eta^2(r_\eta(\theta))}{1-\eta\lambda_\eta^2(r_\eta(\theta))}=[\log\lambda_\eta]'(r_\eta(\theta))=\frac{1}{\theta}=[\log\lambda]'(r(\theta))>\frac{1+\eta\lambda^2(r(\theta))}{1-\eta\lambda^2(r(\theta))}.$$

This proves the first claim. The second claim now follows from (VII.18). □

The second inequality in Lemma VII.22 shows that large deviations *above* the typical speed v_α are more expensive in the random medium than in the appropriate averaged medium. For instance, in the recurrent case we have $\eta=1$ and $v_\alpha=0$, so $I'(\theta)>I'_1(\theta)$ for all $\theta\in(0,1)$. Since $I(0)=I_1(0)=0$, it follows that $I(\theta)>I_1(\theta)$ for all $\theta\in(0,1)$. The latter is the rate function for SRW given by $I_1(\theta)=\frac{1}{2}(1+\theta)\log(1+\theta)+\frac{1}{2}(1-\theta)\log(1-\theta)$ (recall Theorem I.3).

The situation described in the preceeding paragraph can be understood in the light of the remark made below Fig. 11. Since the random walk gets "trapped" in slow stretches, it has a harder time to move at a speed above the typical speed $v_\alpha = 0$ than SRW does. A similar comparison holds when $v_\alpha > 0$. The reverse situation occurs for large deviations below the typical speed v_α, as can be seen from Fig. 11(ii).

VII.8 Concluding remarks

(1) The continued fraction in Lemma VII.12 is easy to compute numerically and allows for a good numerical computation of the rate function I for any reasonable choice of α.

(2) It is shown in Greven and den Hollander [D32] that if (VII.1) fails, then the rate function changes dramatically (see Fig. 14): there is a unique zero at $\theta = v_\alpha$, the two linear pieces shrink to $[-\bar{v}_\alpha, 0]$ respectively $[0, \bar{v}_\alpha]$ for some $0 < \bar{v}_\alpha < v_\alpha$, and both linear pieces become tilted. So, on $[-1, 1]$ the rate function I now looks like:

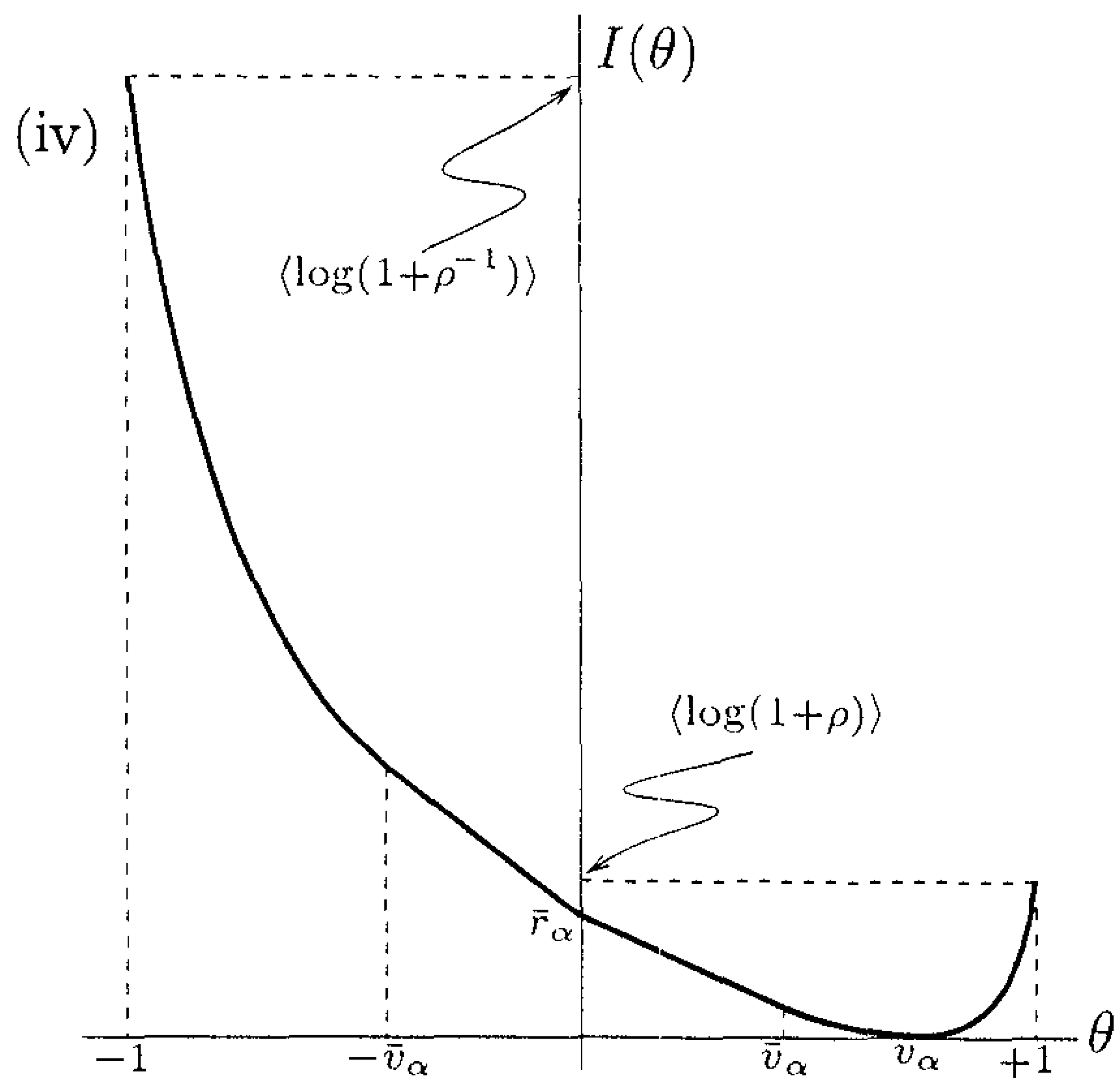

FIG. 14. (iv) transient: uniformly positive drifts (see Fig. 11)

This case corresponds to the situation where the local drifts are uniformly positive, so that the walker sees no "traps" (recall the remark made at the end of Section VII.2). What happens is that $\varphi(r, \omega)$ converges ω-a.s. for all $r \leq \bar{r}_\alpha$, where $\bar{r}_\alpha$ is given by (compare with Lemma VII.12)

$$e^{\bar{r}_\alpha} = \sqrt{\frac{1}{4m(1-m)}} > 1 \quad \text{with } m = \text{essinf}[\text{supp}(\alpha)] > \tfrac{1}{2},$$

while $\bar{v}_\alpha$ is identified as (compare with Lemma VII.16(d))

$$\lim_{r \uparrow \bar{r}_\alpha} \frac{d}{dr} \log \lambda(r) = \frac{1}{\bar{v}_\alpha}.$$

We refer to Comets, Gantert and Zeitouni [D11] for a derivation of these results along the lines of the preceeding sections. The value of $\bar{v}_\alpha$ is not known explicitly. Nor is it known how to obtain a comparison with a homogeneous random walk, like in Lemmas VII.21 and VII.22.

(3) The subdiffusive behavior in the recurrent case (recall Section VII.1) suggests that $\lim_{\theta\downarrow 0} I''(\theta) = \infty$, in the spirit of Lemma I.14(vi). This has, however, not been verified analytically.

EXERCISE VII.23 *Show that the latter property is equivalent to*

$$\lim_{r\uparrow 0} \frac{[\log\lambda]''(r)}{\{[\log\lambda]'(r)\}^3} = 0.$$

A proof of the last display would require sharp control over the tail of the distribution of τ_1 (see Section VII.4), which is not available.

(4) Finer large deviation estimates are needed to understand the behavior in the flat piece of I. Depending on the choice of α, there is a remarkably rich behavior here as well, for which we refer the reader to Gantert and Zeitouni [D22], Pisztora and Povel [D50].

(5) Much of the argument in Sections VII.3–VII.7 carries over to random environments ω that are stationary under translations and satisfy a suitable mixing condition. See Comets, Gantert and Zeitouni [D11].

(6) Theorems VII.3 and VII.4 give us the LDP under the law $\mathbb{P}_\omega$ for fixed ω. This is called the *quenched* LDP. There is also an LDP under the law $\mathbb{P} = \int \alpha^{\mathbb{Z}}(d\omega)\mathbb{P}_\omega$, called the *annealed* LDP. The latter has a different rate function, which lies below the I-function in Theorem VII.3.

EXERCISE VII.24 *Prove the last statement.*

For more details, see Dembo, Peres and Zeitouni [D15], as well as the overview in Gantert and Zeitouni [D23].

(7) Zerner [D62] has extended Theorem VII.3 to random walk in random environment on $\mathbb{Z}^d$ with $d \geq 2$. The main assumptions are: (i) nearest-neighbor steps, (ii) the convex hull of the support of the law of a single jump contains 0 (which is the analogue of (VII.1)). The rate function is linked to certain "Lyapunov exponents" associated with the hitting times of far away points, in the same spirit as Lemmas VII.6 and VII.10. These Lyapunov exponents are not known explicitly, which is why unfortunately little is known about the shape of the rate function other than that it is convex with $I(0) = 0$. It is believed that $\{\theta \in \mathbb{R}^d \colon I(\theta) = 0\}$ has an empty interior, but this remains open.

(8) Sznitman and Zerner [D59] give a sufficient criterion for transience with nonzero speed when $d \geq 2$. It is not known whether this criterion is optimal or not. The precise criterion separating recurrence from transience is open. Thus, much remains to be done when $d \geq 2$.

CHAPTER VIII

HEAT CONDUCTION WITH RANDOM SOURCES AND SINKS

In this chapter we give an application of large deviation theory in analysis. Namely, we show how to apply the LDP for the occupation time measure of a continuous-time symmetric Markov chain, as formulated in Theorem IV.14(ii), to study a partial differential equation describing *heat conduction in a random medium*. It turns out that the solution of this equation has an interesting spatial structure, admitting a description in terms of a certain non-linear "ground state problem". The results described below are taken from Gärtner and Molchanov [D29], [D30] and Gärtner and den Hollander [D26].

VIII.1 The parabolic Anderson model

Let $u\colon \mathbb{Z}^d \times [0,\infty) \to \mathbb{R}$ be given by the Cauchy problem

$$\begin{array}{rcll} \frac{\partial}{\partial t}u(x,t) & = & \Delta u(x,t) + \xi(x)u(x,t), & x \in \mathbb{Z}^d, t \geq 0, \\ u(x,0) & \equiv & 1, & x \in \mathbb{Z}^d. \end{array} \tag{VIII.1}$$

Here, Δ is the discrete Laplacian

$$\Delta u(x,t) = \sum_{y \in \mathbb{Z}^d\colon\, y \sim x} [u(y,t) - u(x,t)],$$

with $y \sim x$ meaning that y is a nearest-neighbor site of x, and

$$\xi = \{\xi(x)\colon x \in \mathbb{Z}^d\}$$

is an i.i.d. $\mathbb{R}$-valued random field playing the role of a *random medium*. Equation (VIII.1), which is referred to in the literature as the parabolic Anderson model, comes up in various areas of chemistry and biology, e.g. catalytic reactions and population dynamics. A vector version of (VIII.1) is used in physics to describe magneto-hydrodynamics. For more background the reader is referred to the monograph by Carmona and Molchanov [B3].

Equation (VIII.1) is a discrete heat equation. Its RHS has two competing parts:

- Δ tends to make u spatially flat,
- ξ tends to make u spatially irregular.

The sites x with $\xi(x) > 0$ can be thought of as heat sources, those with $\xi(x) < 0$ as heat sinks. Our goal will be to study the behavior of the random field $\{u(x,t)\colon x \in \mathbb{Z}^d\}$ for large t, in particular, the *moments* and the *correlations* when we take the average w.r.t. ξ. As will be explained in Section VIII.5, this asymptotics will give us some insight into the size and shape of the dominant peaks in the u-field.

We begin by writing down a formal solution of (VIII.1) using the Feynman-Kac formula, which will serve as the starting point for our analysis. Let

$$Z = \{Z(t) \colon t \geq 0\}$$

denote continuous-time simple random walk (SRW) on $\mathbb{Z}^d$ jumping at rate $2d$, i.e., the Markov process with generator Δ. Write $\mathbb{P}_x, \mathbb{E}_x$ to denote probability and expectation on path space given $Z(0) = x$. Let $\langle\cdot\rangle$ denote expectation w.r.t. ξ. Throughout the sequel we will make the following assumption:

$$H(t) = \log\langle e^{t\xi(0)}\rangle < \infty \qquad \forall t \geq 0. \tag{VIII.2}$$

LEMMA VIII.3 *Assume* (VIII.2). *Then* (VIII.1) *has a unique nonnegative solution ξ-a.s., which admits the Feynman-Kac representation*

$$u(x,t) = \mathbb{E}_x\Big(\exp\Big[\int_0^t \xi(Z(s))ds\Big]\Big). \tag{VIII.4}$$

For all $t \geq 0$, the random field $\{u(x,t) \colon x \in \mathbb{Z}^d\}$ is stationary, ergodic and mixing under translations, and all its moments and correlations are finite.

PROOF. It is easy to check, at least informally, that (VIII.4) is a solution of (VIII.1). Namely, fix $t \geq 0$ and pick a small $h > 0$. Using the Markov property at time h, we may write

$$u(x,t+h) = \mathbb{E}_x\Big(e^{\int_0^h \xi(Z(s))ds} u(Z(h),t)\Big).$$

From this we draw the following identity:

$$\begin{aligned}
\tfrac{1}{h}[u(x,t+h) - u(x,t)] \;&=\; \mathbb{E}_x\Big(\tfrac{1}{h}\big[u(Z(h),t) - u(x,t)\big]\Big) \\
&\quad + \mathbb{E}_x\Big(\tfrac{1}{h}\Big\{e^{\int_0^h \xi(Z(s))ds} - 1\Big\}\Big)\; u(x,t) \\
&\quad + \mathbb{E}_x\Big(\tfrac{1}{h}\Big\{e^{\int_0^h \xi(Z(s))ds} - 1\Big\}\big[u(Z(h),t) - u(x,t)\big]\Big).
\end{aligned}$$

Let $h \downarrow 0$. Then the first term tends to $\Delta u(x,t)$, the second term to $\xi(x)u(x,t)$, and the third term to zero. so we indeed get (VIII.1) for the right time-derivative. A similar argument works for the left time-derivative.

The fact that (VIII.4) is the unique solution of (VIII.1) is amply guaranteed by (VIII.2), for which we refer to Gärtner and Molchanov [D29]. The fact that $u(\cdot,t)$ is stationary is obvious, because $\mu(\cdot,0) \equiv 1$, $\xi(\cdot)$ is stationary and SRW has i.i.d. increments. The fact that $u(\cdot,t)$ is ergodic follows from the observation that $u(\cdot,t)$ is a measurable function of $\xi(\cdot)$, which itself is ergodic. The fact that $u(\cdot,t)$ is mixing (which means that $u(x,t)$ and $u(y,t)$ are asymptotically independent as $|x-y| \to \infty$) holds because SRW can only travel a finite distance in the finite time t. The fact that all moments and correlations are finite is immediate from (VIII.2) and (VIII.4). □

Lemma VIII.3 brings us once again in the realm of large deviations. Apparently, to understand the behavior of the u-field all we need to do is to investigate the large deviation properties of SRW, which is a Markov chain on $\mathbb{Z}^d$. However, the functional appearing in the exponent is far from simple as it depends on the ξ-field, which is random.

Of course, we expect the u-field to develop peaks in the vicinity of where the ξ-field is large and we expect these peaks to spread out and grow with time, creating an irregular landscape of low and high peaks. It is intuitively clear that the asymptotic behavior of the peaks for $t \to \infty$ is determined by the right tail of the distribution of $\xi(0)$. In order to describe how the peaks behave, we make a certain *scaling assumption* on the right tail. This assumption is formulated in terms of the cumulant generating function H and reads

$$\lim_{t\to\infty} \frac{1}{t}[H(ct) - cH(t)] = \rho c \log c \qquad \forall c \in (0,1), \tag{VIII.5}$$

where $\rho \in (0,\infty)$ is a parameter. The basic example satisfying (VIII.5) is when the law of $\xi(0)$ is the double exponential:

$$\mathbb{P}(\xi(0) > s) = \exp[-e^{s/\rho}], \qquad s \in \mathbb{R}.$$

EXERCISE VIII.6 *Show that for this example* $H(t) = \log\Gamma(\rho t + 1)$ *with* Γ *the Gamma function, and check that it satisfies* (VIII.5).

Thus, (VIII.5) says that the right tail of the distribution of $\xi(0)$ behaves like the double exponential. Since the double exponential drops to zero rapidly beyond the value $s = \rho$, we may think of the parameter ρ in (VIII.5) as measuring the degree of disorder in the ξ-field.

The fact that the double exponential distribution is unbounded from above is important: (VIII.4) shows that all cases where ξ is bounded from above can be reduced, via a trivial exponential factor, to the case where ξ is negative, for which there is no growth at all.

EXERCISE VIII.7 *Show that if* $\xi(0)$ *is bounded from above, then the limit in* (VIII.5) *is zero* ($\rho = 0$).

Although (VIII.5) represents a rather limited class of examples, we will see that it is in some sense a *critical class* for a nice description of the dominant peaks in the u-field. In Section VIII.7 we will comment on what happens under different types of scaling.

VIII.2 Growth rate of the moments

As a first step towards understanding the evolution of the u-field, we monitor the growth of the integer moments of $u(0,t)$. The main result is the following.

THEOREM VIII.8 *Assume* (VIII.2) *and* (VIII.5). *Then, for any* $p \in \mathbb{N}$,

$$\langle u^p(0,t)\rangle = \exp[H(pt) - 2dpt\,\chi(\rho) + o(t)],$$

where

$$\chi(\rho) = \frac{1}{2d} \inf_{\nu\in\mathfrak{M}_1(\mathbb{Z}^d)} \{I(\nu) + \rho J(\nu)\}, \tag{VIII.9}$$

with

$$\begin{aligned} I(\nu) &= \tfrac{1}{2} \sum_{x,y\in\mathbb{Z}^d:\ x\sim y} \left(\sqrt{\nu(x)} - \sqrt{\nu(y)}\right)^2, \\ J(\nu) &= -\sum_{x\in\mathbb{Z}^d} \nu(x)\log\nu(x). \end{aligned} \tag{VIII.10}$$

Theorem VIII.8 identifies the growth rate of the p-th moment up to exponential order, thus giving us information on the *mean height* of the peaks in the u-field. The leading term already gives us something interesting. Namely, assumption (VIII.5) implies that

$$H(t) = \rho t \log(\rho t) - \rho t + o(t).$$

Therefore it follows that for every $p > q \geq 1$:

$$\frac{\langle u^p(0,t)\rangle^{1/p}}{\langle u^q(0,t)\rangle^{1/q}} = \exp\left[\rho t \log\left(\frac{p}{q}\right) + o(t)\right].$$

This formula says that the second moment grows faster than the square of the first moment by an exponential factor, and similarly for the higher moments. This phenomenon, which is referred to in the literature as *intermittency*, is to be interpreted as follows. The dominant peaks in the u-field are localized on *random islands*, which are far apart and occupy a fraction of the lattice that vanishes as $t \to \infty$. Nevertheless, on these islands the peaks are so high that they determine the growth of the moments. Each higher moment is determined by a smaller fraction of the peaks.

Theorem VIII.8 is proved in Section VIII.5.

VIII.3 Analysis of the variational problem

We next analyze the variational problem in Theorem VIII.8. It turns out that (VIII.9) factorizes into the d-fold product of the one-dimensional version of (VIII.9) and that, consequently, $\chi(\rho)$ does not depend on d. Moreover, the one-dimensional variational problem has a minimizer that is linked to the following non-linear problem.

THEOREM VIII.11 *Consider the non-linear difference equation*

$$\Delta v + 2\rho v \log v = 0, \qquad v\colon \mathbb{Z} \to (0,\infty). \tag{VIII.12}$$

This equation has a ground state v_ρ, i.e., a solution with minimal ℓ^2-norm, and

$$\chi(\rho) = \rho \log \|v_\rho\|_2.$$

Theorem VIII.11 is proved in Section VIII.6. We will see that, modulo some transformation, (VIII.12) is the Euler-Lagrange equation associated with the one-dimensional version of (VIII.9).

In Gärtner and den Hollander [D26] it is shown that for ρ sufficiently large (namely, $\rho \geq 15.7$) the ground state of (VIII.12) is unique. Numerical calculation suggests that the ground state is actually unique for all ρ, but this remains open. In addition, it is shown that (see Fig. 15)

$$\chi(\rho) \sim \frac{1}{4}\rho \log\left(\frac{1}{\rho}\right) \quad \text{as } \rho \downarrow 0, \qquad \chi(\rho) \to 1 \quad \text{as } \rho \to \infty.$$

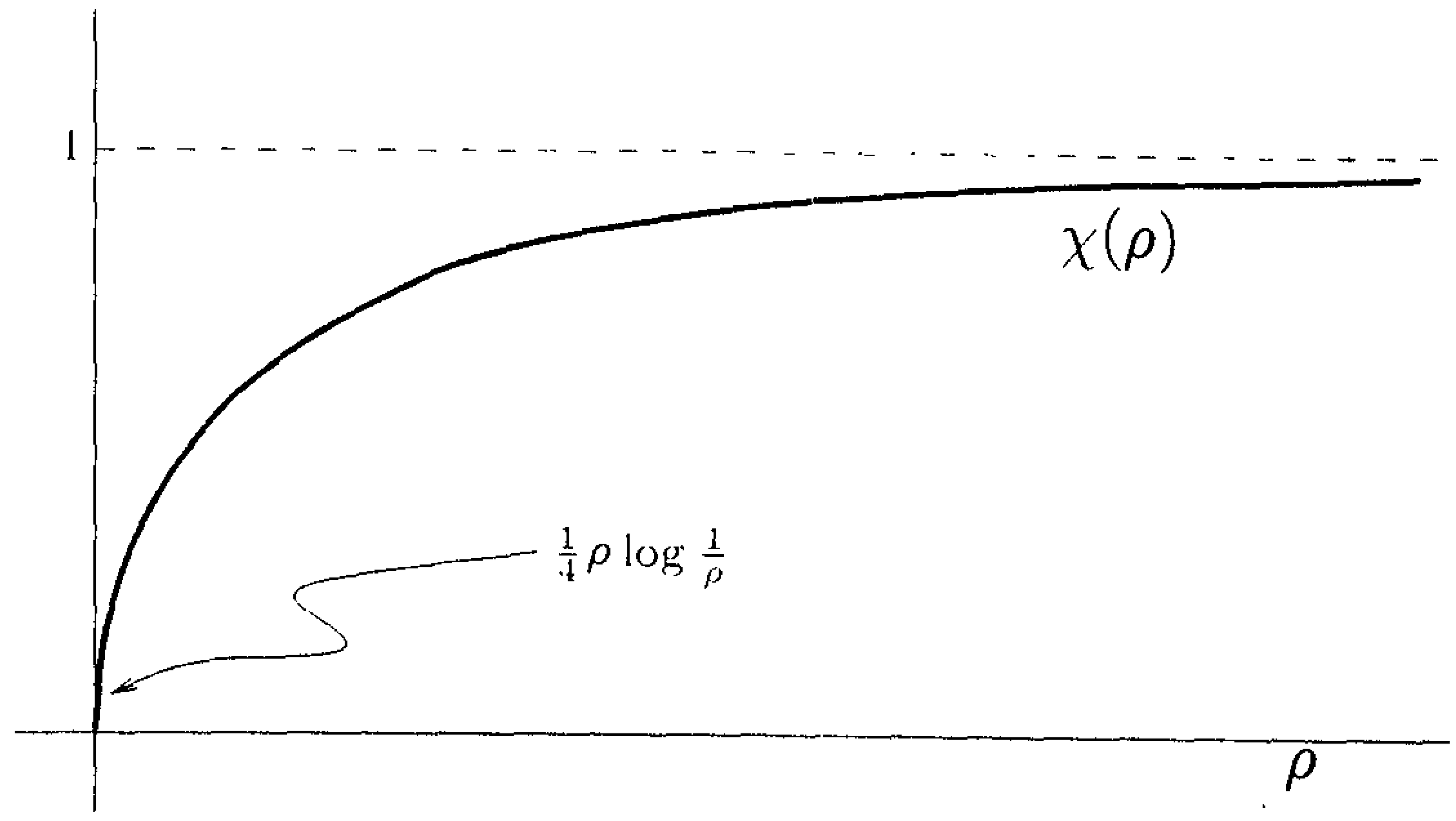

FIG. 15. The χ-function

VIII.4 Correlation structure

To understand the *correlation structure* of the u-field, one needs to extend Theorem VIII.8 to mixed moments

$$\langle u(x_1,t) \times \cdots \times u(x_p,t) \rangle, \qquad x_1,\ldots,x_p \in \mathbb{Z}^d,$$

and push the asymptotics for $t \to \infty$ all the way up to and including order 1. Indeed, this accuracy is needed to be able to compute the limit of the ratios

$$\frac{\langle u(x_1,t) \times \cdots \times u(x_p,t) \rangle}{\langle u^p(0,t) \rangle}.$$

Such an extension is derived in Gärtner and den Hollander [D26] under the stronger scaling assumption

$$\lim_{t\to\infty} tH''(t) = \rho. \tag{VIII.13}$$

EXERCISE VIII.14 *Prove that* (VIII.13) *implies* (VIII.5).

For $p = 2$ the result, under the assumption that the ground state in Theorem VIII.11 is unique, reads

$$\lim_{t\to\infty} \frac{\langle u(x,t)u(y,t) \rangle}{\langle u^2(0,t) \rangle} = c_\rho(x,y) \qquad \forall x,y \in \mathbb{Z}^d, \tag{VIII.15}$$

with

$$c_\rho(x,y) = \frac{1}{\|w_\rho\|_2^2} \sum_{x\in\mathbb{Z}^d} w_\rho(x+z)w_\rho(y+z) \qquad \text{and} \qquad w_\rho = v_\rho^{\otimes d}. \tag{VIII.16}$$

A similar behavior shows up for all $p > 2$, i.e., all the p-th order correlation coefficients converge to a limit in terms of p appropriate shifts of w_ρ.

The last two displays should be interpreted as saying that the high peaks in the u-field have a shape that is a multiple of w_ρ (see Fig. 16). The idea is that, since the dominant peaks are located on sparse islands, the main contribution to

$c_\rho(x,y)$ comes from those realizations where x and y both lie in the same island. The ergodic theorem tells us that

$$\langle u(x,t)u(y,t)\rangle = \lim_{N\to\infty} \frac{1}{|T_N|} \sum_{z\in T_N} u(z+x,t)u(z+y,t),$$

with $T_N \in [-N,N)^d \cap \mathbb{Z}^d$. For the ratio in the LHS of (VIII.15) it is irrelevant what the absolute heights of the peaks are, only their relative heights count, which are given by $w_\rho(z+x)$ and $w_\rho(z+y)$, respectively. A similar picture applies to the higher-order correlations. Therefore the above result for the correlation coefficients suggests that *the asymptotic shape of the peaks is* w_ρ. Nevertheless, a rigorous proof of this interpretation is not known, since it is not clear how to justify an interchange of the limits $N \to \infty$ and $t \to \infty$.

A qualitative picture of v_ρ is:

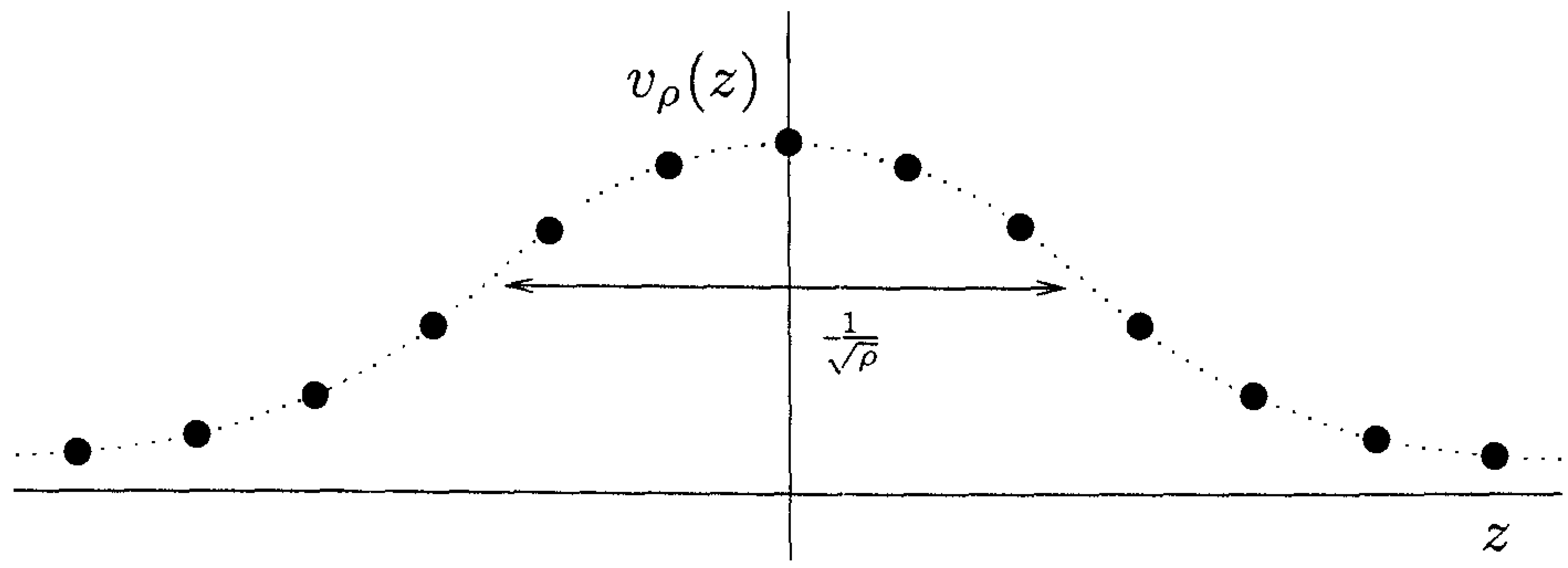

FIG. 16. The ground state v_ρ centered at the origin

The d-fold product of this picture is the shape of the high peaks in $\mathbb{Z}^d$.

In Gärtner and den Hollander [D26] it is further shown that, still under (VIII.13),

$$\lim_{\rho\downarrow 0} c_\rho\Big(\lfloor x/\sqrt{\rho}\rfloor, \lfloor y/\sqrt{\rho}\rfloor\Big) = e^{-\frac{1}{4}|x-y|^2}, \qquad \lim_{\rho\to\infty} c_\rho(x,y) = \delta_{x,y}. \qquad \text{(VIII.17)}$$

This result says that the islands have a typical size of order $1/\sqrt{\rho}$ for small ρ, blow up for $\rho \downarrow 0$ and contract to single points for $\rho \to \infty$. Thus, the double exponential is actually the *critical class* for having a non-trivial limiting correlation structure as $t \to \infty$.

The proof of (VIII.15), (VIII.16) and (VIII.17) is beyond our scope. It boils down to a refined large deviation analysis of the "transformed random walk" whose generator G_ρ is given by

$$(G_\rho f)(x) = \sum_{y\in\mathbb{Z}^d:\, y\sim x} \frac{w_\rho(y)}{w_\rho(x)}[f(y)-f(x)].$$

In conclusion, we see that the variational problem (VIII.9) in Theorem VIII.8 and the associated difference equation (VIII.12) in Theorem VIII.11 are vital for an understanding of how the u-field evolves.

VIII.5 Derivation of the growth rate

In this section we give the proof of Theorem VIII.8 for the case $p = 1$. At the end we will indicate how to generalize to $p \geq 2$. Let

$$\ell_t(z) = \int_0^t 1_{\{Z(s)=z\}} ds, \qquad z \in \mathbb{Z}^d, t \geq 0,$$

denote the local times of our SRW. Then, from the Feynman-Kac formula in (VIII.4), we have

$$u(0,t) = \mathbb{E}_0\Big(\exp\Big[\sum_{z\in\mathbb{Z}^d} \xi(z)\ell_t(z)\Big]\Big).$$

Take the expectation over ξ, use Fubini and the i.i.d. property of ξ, and recall (VIII.2), to get

$$\langle u(0,t)\rangle = \mathbb{E}_0\Big(\exp\Big[\sum_{z\in\mathbb{Z}^d} H(\ell_t(z))\Big]\Big).$$

Since $\sum_{z\in\mathbb{Z}^d} \ell_t(z) = t$, the exponent in the last display may be rewritten as

$$\sum_{z\in\mathbb{Z}^d} H(\ell_t(z)) = H(t) + t\sum_{z\in\mathbb{Z}^d} \frac{1}{t}\Big[H(L_t(z)t) - L_t(z)H(t)\Big],$$

where

$$L_t = \frac{1}{t}\ell_t$$

is the occupation time measure. Since we have assumed (VIII.5), it thus seems *plausible* that

$$\sum_{z\in\mathbb{Z}^d} H(\ell_t(z)) = H(t) + t\sum_{z\in\mathbb{Z}^d}\big[\rho L_t(z)\log L_t(z) + o(1)\big] = H(t) - t\rho J(L_t) + o(t),$$

where we insert the definition of the J-functional in Theorem VIII.8. This leads us to the expression

$$\langle u(0,t)\rangle = e^{H(t)+o(t)} \int_{\mathfrak{M}_1(\mathbb{Z}^d)} e^{-t\rho J(\nu)}\mathbb{P}_0(L_t \in d\nu).$$

Next, since (L_t) satisfies the weak LDP on $\mathfrak{M}_1(\mathbb{Z}^d)$ with rate t and with weak rate function given by the I-functional in Theorem VIII.8, it is *plausible* via Theorem III.2 (Varadhan's Lemma) that

$$\begin{aligned}\langle u(0,t)\rangle &= \exp\Big[H(t) - t\inf_{\nu\in\mathfrak{M}_1(\mathbb{Z}^d)}\{I(\nu) + \rho J(\nu)\} + o(t)\Big] \\ &= \exp[H(t) - 2dt\,\chi(\rho) + o(t)]\end{aligned}$$

(recall (VIII.9)), which is the claim in Theorem VIII.8 for $p = 1$. Note here that the I-functional equals

$$I(\nu) = \langle\sqrt{\nu}, (-\Delta)\sqrt{\nu}\rangle,$$

which comes from Theorem IV.14(b). Only the weak LDP holds: I does not have compact level sets because $\mathbb{Z}^d$ is infinite.

The above two plausible steps need justification (see also Remark III.24.3). For this we do a standard compactification argument, namely, we restrict the problem to a large box $T_N = (-N, N]^d \cap \mathbb{Z}^d$ and let $N \to \infty$ afterwards.

Upper bound:

Suppose we wrap the SRW around T_N, i.e., it moves on T_N with periodic boundary conditions. Let $\{\ell_t^N(z) \colon z \in T_N\}$ denote the local times of the wrapped SRW. We can relate these to the local times of the unwrapped SRW as follows:

$$\ell_t^N(z) = \sum_{y \in (2N\mathbb{Z})^d} \ell_t(z+y), \qquad z \in T_N.$$

Since $H(0) = 0$ and since H is convex, we therefore have

$$\sum_{z \in \mathbb{Z}^d} H(\ell_t(z)) \leq \sum_{z \in T_N} H(\ell_t^N(z)).$$

Hence, for every N, we have the upper bound

$$\langle u(0,t) \rangle \leq \mathbb{E}_0\Big(\exp\Big[\sum_{z \in T_N} H(\ell_t^N(z))\Big]\Big).$$

But the RHS is a problem on the finite torus T_N, so the above two plausible steps become rigorous. In particular, we can now apply the LDP in Theorem IV.14(b) in combination with Theorem III.2, to obtain

$$\langle u(0,t) \rangle \leq \exp[H(t) - 2dt\,\chi^N(\rho) + o(t)],$$

with

$$\chi^N(\rho) = \frac{1}{2d} \inf_{\nu \in \mathfrak{M}_1(T_N)} \{I^N(\nu) + \rho J^N(\nu)\},$$

where I^N, J^N are the I, J-functionals in (VIII.10) restricted to T_N with periodic boundary conditions. The leading term is already what we want. All that remains to do is to show that

$$\chi^N(\rho) \to \chi(\rho) \qquad \text{as } N \to \infty.$$

This can be done via a standard argument, for which we refer the reader to Gärtner and Molchanov [D30].

Lower bound:

Suppose we kill the SRW when it hits ∂T_N. This amounts to inserting under the expectation the indicator of the event $\{\ell_t(z) = 0\ \forall z \notin T_N \setminus \partial T_N\}$. In terms of the wrapped SRW the latter translates into the event $\{\ell_t^N(z) = 0\ \forall z \in \partial T_N\}$ and so, for every N, we get the lower bound

$$\langle u(0,t) \rangle \geq \mathbb{E}_0\Big(\exp\Big[\sum_{z \in T_N} H(\ell_t^N(z))\Big] 1_{\{\ell_t^N(z)=0\ \forall z \in \partial T_N\}}\Big).$$

With the help of Theorem IV.14(b) and Theorem III.2 this in turn leads to

$$\langle u(0,t) \rangle \geq \exp[H(t) - 2dt\,\widehat{\chi}^N(\rho) + o(t)],$$

with

$$\widehat{\chi}^N(\rho) = \frac{1}{2d} \inf_{\nu \in \mathfrak{M}_1(T_N) \colon \nu(z)=0\ \forall z \in \partial T_N} \{I^N(\nu) + \rho J^N(\nu)\}.$$

Now it suffices to show that

$$\widehat{\chi}^N(\rho) \to \chi(\rho) \qquad \text{as } N \to \infty,$$

for which we again refer to Gärtner and Molchanov [D30].

EXERCISE VIII.18 *Check that $N \mapsto \widehat{\chi}^N(\rho)$ is non-increasing.*

The above argument completes the proof for the case $p = 1$. The generalization to $p \geq 2$ goes as follows. Let $\{Z^i(s)\colon s \geq 0\}$, $i = 1, \dots, p$, be p independent copies of our SRW. Let ℓ_t^i denote the local time of the i-th SRW. Then the Feynman-Kac formula in (VIII.4) gives

$$\begin{aligned}\langle u^p(0,t)\rangle &= \Big\langle \mathbb{E}_0^{\otimes p}\Big(\prod_{i=1}^p \exp\Big[\int_0^t \xi(Z^i(s))ds\Big]\Big)\Big\rangle \\ &= \Big\langle \mathbb{E}_0^{\otimes p}\Big(\exp\Big[\sum_{z\in\mathbb{Z}^d} \xi(z)\sum_{i=1}^p \ell_t^i(z)\Big]\Big)\Big\rangle \\ &= \mathbb{E}_0^{\otimes p}\Big(\exp\Big[\sum_{z\in\mathbb{Z}^d} H\Big(\sum_{i=1}^p \ell_t^i(z)\Big)\Big]\Big).\end{aligned}$$

Going through the same argument, and using that $\sum_{z\in\mathbb{Z}^d}\sum_{i=1}^p \ell_t^i(z) = pt$, we find

$$\langle u^p(0,t)\rangle = \exp\Big[H(pt) - 2dpt\,\chi_p(\rho) + o(t)\Big],$$

with

$$\chi_p(\rho) = \frac{1}{2d}\inf_{\nu^1,\dots,\nu^p\in\mathfrak{M}_1(\mathbb{Z}^d)}\Big\{\frac{1}{p}\sum_{i=1}^p I(\nu^i) + \rho J\Big(\frac{1}{p}\sum_{i=1}^p \nu^i\Big)\Big\}.$$

Here, $(\nu^1,\dots,\nu^p) \mapsto \frac{1}{p}\sum_{i=1}^p I(\nu^i)$ is the weak rate function in the LDP for the p-tuple $(L_t^1,\dots,L_t^p)$. However, because $\nu \mapsto J(\nu)$ is strictly concave, the infimum reduces to the diagonal $\nu^1 = \cdots = \nu^p$. Hence $\chi_p(\rho) = \chi(\rho)$, which completes the proof of Theorem VIII.8.

VIII.6 Transformation of the variational problem

In this section we give the proof of Theorem VIII.11. We begin by showing why the variational problem in (VIII.9) factorizes. Abbreviate $F_d = I_d + \rho J_d$, where we add the index d to exhibit the dimension dependence. We will show that

$$\inf_{\nu\in\mathfrak{M}_1(\mathbb{Z}^d)} F_d(\nu) = d\inf_{\nu\in\mathfrak{M}_1(\mathbb{Z})} F_1(\nu).$$

The proof is by induction on d.

The claim is obviously true for $d = 1$. Suppose it is true for all dimensions $\leq d$. Pick any $\nu \in \mathfrak{M}_1(\mathbb{Z}^{d+1})$, and let ν_d, ν_1 denote its projections onto the coordinates numbered $1,\dots,d$ respectively $d+1$. Define the conditional probabilities

$$\mu_x(z) = \frac{\nu(x,z)}{\nu_d(x)}, \qquad \mu_z(x) = \frac{\nu(x,z)}{\nu_1(z)}, \qquad x\in\mathbb{Z}^d, z\in\mathbb{Z}.$$

Then an easy computation gives

$$\begin{aligned}I_{d+1}(\nu) &= \sum_x \nu_d(x)I_1(\mu_x) + \sum_z \nu_1(z)I_d(\mu_z), \\ J_{d+1}(\nu) &= \sum_x \nu_d(x)J_1(\mu_x) + \sum_z \nu_1(z)J_d(\mu_z) + K(\nu),\end{aligned}$$

with

$$K(\nu) = \sum_x \nu_d(x)\Big[\sum_z \mu_x(z)\log\mu_x(z)\Big] - \sum_z \nu_1(z)\log\nu_1(z).$$

Because $\mu \mapsto \mu \log \mu$ is strictly convex and $\sum_x \nu_d(x)\mu_x(z) = \nu_1(z)$ for all z, it follows from Jensen's inequality that $K(\nu) \geq 0$ with equality if and only if μ_x does not depend on x, i.e., ν is of the form

$$\nu = \nu_d \times \nu_1.$$

Thus, we have

$$F_{d+1}(\nu) \geq \sum_x \nu_d(x) F_1(\mu_x) + \sum_z \nu_1(z) F_d(\mu_z) \qquad \forall \nu \in \mathfrak{M}_1(\mathbb{Z}^{d+1}),$$

and hence

$$\inf_{\nu \in \mathfrak{M}_1(\mathbb{Z}^{d+1})} F_{d+1}(\nu) \geq \inf_{\eta \in \mathfrak{M}_1(\mathbb{Z})} F_1(\eta) + \inf_{\zeta \in \mathfrak{M}_1(\mathbb{Z}^d)} F_d(\zeta).$$

However, equality holds because

$$F_{d+1}(\zeta \times \eta) = F_1(\eta) + F_d(\zeta) \qquad \forall \eta \in \mathfrak{M}_1(\mathbb{Z})\ \forall \zeta \in \mathfrak{M}_1(\mathbb{Z}^d),$$

so we have completed the induction step. So the problem indeed reduces to $d = 1$.

We refer the reader to Gärtner and den Hollander [D26] for the proof that $\inf_{\nu \in \mathfrak{M}_1(\mathbb{Z})} F_1(\nu)$ has a minimizer for all ρ, which is based on a tightness argument. In what remains we will only demonstrate the link between (VIII.9) and (VIII.12).

It is not hard to show that every minimizer of F_1 is strictly positive. This fact allows us to do a standard variational argument, which shows that for every minimizer ν there exists a constant λ such that

$$-\sqrt{\frac{\nu(z+1)}{\nu(z)}} - \sqrt{\frac{\nu(z-1)}{\nu(z)}} + 2 - \rho \log \nu(z) = \lambda, \qquad z \in \mathbb{Z}.$$

EXERCISE VIII.19 *Prove the last two claims.*

The last display is just the Euler-Lagrange equation associated with (VIII.9). Now put

$$v(z) = e^{\lambda/2\rho}\sqrt{\nu(z)}. \tag{VIII.20}$$

Then the Euler-Lagrange equation transforms into

$$\frac{v(z+1)}{v(z)} + \frac{v(z-1)}{v(z)} - 2 + 2\rho \log v(z) = 0, \qquad z \in \mathbb{Z}.$$

However, this is precisely (VIII.12). Moreover,

$$\begin{aligned} \chi(\rho) &= F_1(\nu) = I_1(\nu) + \rho J_1(\nu) \\ &= \sum_z \left(\sqrt{\nu(z+1)} - \sqrt{\nu(z)}\right)^2 - \rho \sum_z \nu(z) \log \nu(z) \\ &= \lambda \sum_z \nu(z) = \lambda = 2\rho \log \|v\|_2, \end{aligned}$$

which completes the proof of Theorem VIII.11. Note that the ground states of (VIII.12) are in a one-to-one correspondence with the minimizers of (VIII.9) via the relation in (VIII.20). □

VIII.7 Concluding remarks

(1) The non-linearity and the discrete nature of (VIII.12) make it into a rather hard object to analyze mathematically. There seem to be no good functional analytic tools around to settle the uniqueness issue. For instance, we know that all ground states are unimodal (i.e., monotone left and right of their maximal value), but not whether they are all symmetric modulo translations.

(2) Theorem VIII.8 describes the growth rate of $u(0,t)$ averaged over ξ, which is called the *annealed* model. Gärtner and Molchanov [D30] compute the growth rate conditioned on ξ, called the *quenched* model. This growth rate is different from the one in the annealed model, and is determined by a different collection of dominant peaks, but the χ-function again plays a major role.

(3) What happens when we modify the scaling assumption in (VIII.5)? A glance at the proof in Section VIII.5 shows that much of the argument carries over, with the result that the second term in the expansion in Theorem VIII.8 changes (and is described by a different variational problem). However, the mathematical details are tricky. The most interesting situations occur when the right tail of the distribution of $\xi(0)$ decays *faster* than the double exponential. In that case the islands carrying the dominant peaks in the u-field grow with time, and one can study how the islands and the peaks scale as $t \to \infty$. Work in this direction and relevant references can be found in the monograph by Sznitman [B10]. When the right tail of $\xi(0)$ decays *slower* than the double exponential, the discrete-space model is trivial (because the islands are single sites). Interesting results for the continuous-space model are derived in Gärtner and König [D27] and Gärtner, König and Molchanov [D28].

CHAPTER IX

POLYMER CHAINS

In this chapter we give an application of large deviation theory in chemistry. Namely, we show how Sanov's theorem for the pair empirical measure of a Markov chain can be used to study the spatial distribution of a one-dimensional polymer. In particular, we derive the LDP *for the speed of the polymer* and show that the rate function exhibits a somewhat unexpected behavior. The full argument will take up quite a bit of space, but it is worth the effort because we will arrive at a complete analytical description.

IX.1 A polymer model: self-repellent random walk

A polymer is a long chain of monomers connected by chemical bonds. Two characteristic properties of a polymer are:

(1) *spatial irregularity*, because the chain tends to be intertwined;
(2) *self-repellence*, because monomers tend to avoid being close to each other ("excluded-volume effect").

A complete description of the spatial properties of a polymer turns out to be very difficult, so one resorts to simplified models. In this chapter we consider the so-called *weakly self-avoiding walk*, which is a model for "soft polymers".

To incorporate property (1), we think of the polymer chain as a discrete-time simple random walk (SRW) on $\mathbb{Z}^d$, starting at the origin. Let

$$\begin{aligned} S_n &= \text{the position of the walker at time } n \\ &= \text{the location of the } n\text{-th monomer,} \end{aligned}$$

and let $\mathbb{P}, \mathbb{E}$ denote probability and expectation for $(S_n)_{n \in \mathbb{N}_0}$. To also incorporate property (2), we fix n, pick a parameter $\beta \in (0, \infty]$, and define a new law $\mathbb{Q}_n^\beta$ on n-step paths by setting

$$\frac{d\,\mathbb{Q}_n^\beta}{d\,\mathbb{P}_n}[\,\cdot\,] = \frac{1}{Z_n^\beta} e^{-\beta I_n[\,\cdot\,]}, \tag{IX.1}$$

where $\mathbb{P}_n$ is the projection of $\mathbb{P}$ onto the first n steps,

$$\begin{aligned} I_n[(S_i)_{i=0}^n] &= \sum_{\substack{i,j=0 \\ i<j}}^{n} 1_{\{S_i = S_j\}} \\ &= \text{the total number of self-intersections in the path,} \end{aligned}$$

and Z_n^β is the normalizing constant. In words, under the law $\mathbb{Q}_n^\beta$ every self-intersection is penalized by a factor $e^{-\beta}$, and so self-intersections are effectively suppressed. We may therefore think of $\mathbb{Q}_n^\beta$ as the law of a polymer chain of length n that models both (1) and (2). The parameter β is called the ***strength of self-repellence.***

The definition of $\mathbb{Q}_n^\beta$ in (IX.1) shows that, in order to understand the behavior of the polymer for large n, we must study the large deviation properties of I_n under the law $\mathbb{P}$ of SRW.

There are lots of interesting questions one may ask about this model. The question that will be addressed here is the asymptotic behavior of

$$|S_n| = \text{the distance between the endpoints of a polymer chain of length } n$$

under the law $\mathbb{Q}_n^\beta$ in the limit as $n \to \infty$ for fixed β. We briefly summarize the current state of knowledge.

- For $d \geq 5$, S_n is of order $\sqrt{n}$ for all $\beta \in (0, \infty]$ (Hara and Slade [D36]). The proof of this fact is based on a perturbation technique, called the "lace expansion", which hinges on the fact that in high dimensions self-intersections of the SRW (with $\beta = 0$) predominantly occur in short loops and are therefore not capable of qualitatively changing the global asymptotic behavior when $\beta \in (0, \infty]$.

- Little is known mathematically for dimensions $d = 2$, 3 and 4, although the literature is full of conjectures. It is believed that S_n grows faster than $\sqrt{n}$ (superdiffusive) but slower than n (subballistic). Especially for $d = 2$ there are precise conjectures on the scaling behavior, e.g. that S_n is of order $n^{3/4}$.

- In $d = 1$, S_n is of order n for all $\beta \in (0, \infty]$ (Greven and den Hollander [D31]). The proof of this fact is based on large deviation theory and will be given below.

Note that $(\mathbb{Q}_n^\beta)_{n \in \mathbb{N}_0}$ is not a consistent family of probability laws, in the sense that $\mathbb{Q}_n^\beta$ does not coincide with the projection of $\mathbb{Q}_{n+1}^\beta$ obtained after summing out the $(n+1)$-st step. This fact comes from the way in which the model is set up and actually complicates our analysis somewhat. The law $\mathbb{Q}_n^\beta$ describes a polymer of a fixed length n. There are different models for describing polymers that grow with time (see e.g. Tóth [D60]).

For further background we refer the reader to the monographs by Madras and Slade [C5] and van der Hofstad [B5].

IX.2 Linear speed

Henceforth we consider the case $d = 1$ with $\beta \in (0, \infty)$. Intuitively, we expect that typical paths under the measure $\mathbb{Q}_n^\beta$ hang around the origin for a while and then wander off to infinity at a strictly positive speed because of the self-repellence. There is a trivial symmetry between left and right. The main theorems that we will prove are the following.

THEOREM IX.2 *For every $\beta \in (0, \infty)$ there exists a $\theta^*(\beta) \in (0, 1)$ such that*

$$\lim_{n \to \infty} \mathbb{Q}_n^\beta \left(\left| \frac{1}{n} S_n - \theta^*(\beta) \right| \leq \epsilon \;\middle|\; S_n \geq 0 \right) = 1 \textit{ for all } \epsilon > 0.$$

THEOREM IX.3 (a) *The function $\beta \mapsto \theta^*(\beta)$ can be computed in terms of a variational problem.*

(b) *It follows from the solution of this variational problem that*

$$\beta \mapsto \theta^*(\beta) \text{ is analytic on } (0,\infty),$$
$$\lim_{\beta\downarrow 0} \theta^*(\beta) = 0,$$
$$\lim_{\beta\to\infty} \theta^*(\beta) = 1.$$

The quantity $\theta^*(\beta)$ is the speed of the polymer with strength of repellence β. We will eventually show that $\mathbb{Q}_n^\beta(\frac{1}{n}S_n \in \cdot\,|S_n \geq 0)$ satisfies an LDP on $[0,1]$ with rate n and with a rate function having $\theta^*(\beta)$ as its unique zero (see Section IX.9), which explains what is behind Theorem IX.2. In Section IX.10 we will see that $\beta \mapsto \theta^*(\beta)$ looks like:

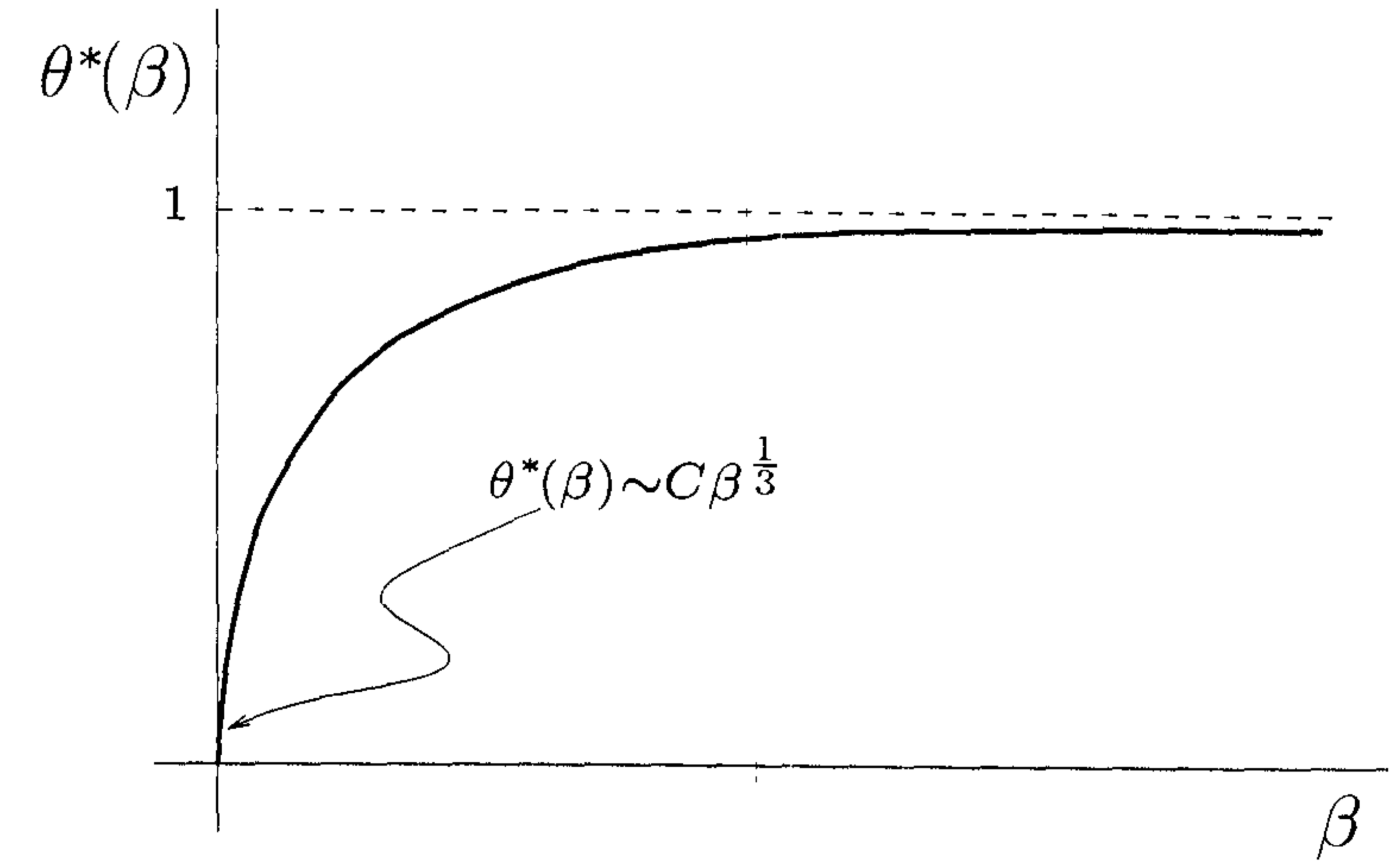

FIG. 17. The linear speed of the polymer

Before we get going, we first rewrite the definition of $\mathbb{Q}_n^\beta$ in a way more convenient for the argument. Let

$$\widehat{I}_n[(S_i)_{i=0}^n] = \sum_{i,j=0}^n 1_{\{S_i=S_j\}}.$$

Then clearly $\widehat{I}_n[\cdot] = 2I_n[\cdot]+(n+1)$. Hence we may as well put $\widehat{I}_n$ in the exponential weight factor (which only changes β to 2β). The following object is of paramount importance for the argument. Define

$$\ell_n(x) = \sum_{i=0}^n 1_{\{S_i=x\}}, \qquad x \in \mathbb{Z}, n \in \mathbb{N}_0,$$

i.e., the *local time at site x up to time n*. We can then write

$$\widehat{I}_n[(S_i)_{i=0}^n] = \sum_{x\in\mathbb{Z}} \sum_{i,j=0}^n 1_{\{S_i=S_j=x\}} = \sum_{x\in\mathbb{Z}} \ell_n^2(x).$$

We thus see that the polymer problem really amounts to understanding the large deviation properties of the random sequence $\{\ell_n(x)\}_{x\in\mathbb{Z}}$ under the law $\mathbb{P}$ of SRW.

IX.3 The LDP for bridges

In order to obtain the desired LDP for $\mathbb{Q}_n^\beta(\frac{1}{n}S_n \in \cdot \,|S_n \geq 0)$, we begin by deriving an LDP under the restriction that the path be a *bridge*, i.e., that it lies between its endpoints. This restriction will be crucial for our analysis, and will only be removed in Section IX.9.

In what follows, $\simeq$ denotes logarithmic equivalence (see (I.1)). Our first lemma shows that the bridge condition does not change the normalizing constant (see (IX.1)).

LEMMA IX.4 *For $n \to \infty$,*

$$\mathbb{E}\,(e^{-\beta \widehat{I}_n} 1_{S_n \geq 0}) \simeq \mathbb{E}\left(e^{-\beta \widehat{I}_n} 1_{\circledast_n}\right),$$

with $\widehat{I}_n = \widehat{I}_n[(S_i)_{i=0}^n]$ and

$$\circledast_n = \{S_0 \leq S_i \leq S_n \;\forall\, 0 \leq i \leq n\}.$$

PROOF. The proof requires a geometrical argument. Fix n. First, suppose that the path is a half-bridge, i.e., $S_i > S_0 \;\forall 0 < i \leq n$. We can then do a reflection procedure starting from the left endpoint of the path, as follows. Put $i_0 = 0$ and, for $j = 1, 2, \ldots$, define (R_j, i_j) recursively as

$$\begin{aligned} R_j &= \max\nolimits_{i_{j-1} < i \leq n} (-1)^j (S_{i_{j-1}} - S_i), \\ i_j &= \text{the largest } i \text{ where the maximum is attained.} \end{aligned}$$

The recursion is stopped at the smallest integer k such that $i_k = n$. What this definition says is that R_j is the span of the subwalk $(S_{i_{j-1}}, \ldots, S_n)$. Each subwalk $(S_{i_{j-1}}, \ldots, S_{i_j})$ lies strictly on one side of the point $S_{i_{j-1}}$, and

$$R_1 + \cdots + R_k \leq n \;\text{ and }\; R_1 > R_2 > \cdots > R_k \geq 1. \tag{IX.5}$$

If for $j = 1, 2, \ldots, k-1$ we reflect $(S_{i_j}, \ldots, S_n)$ around the point S_{i_j}, then we end up with a bridge, i.e., a path satisfying $S_0 < S_i \leq S_n \;\forall 0 < i \leq n$. Moreover, this bridge is less penalized than the original path because obviously it has less self-intersections. Next, we drop the assumption that the path be a half-bridge and only suppose that $S_n \geq 0$. Let

$$\begin{aligned} i_- &= \min\left\{0 \leq i \leq n\colon\, S_i = \min_{0 \leq j \leq n} S_j\right\}, \\ i_+ &= \max\left\{0 \leq i \leq n\colon\, S_i = \max_{0 \leq j \leq n} S_j\right\}. \end{aligned}$$

Then both $(S_0, \ldots, S_{i_-})$ and $(S_{i_+}, \ldots, S_n)$ are half-bridges, and the above reflection procedure applies. If we fold both pieces outwards after the reflection procedure is through, then we end up with a bridge. Hence, we conclude that

$$\mathbb{E}\,(e^{-\beta \widehat{I}_n} 1_{S_n \geq 0}) \leq N_n^2\, \mathbb{E}\,(e^{-\beta \widehat{I}_n} 1_{\circledast_n}),$$

with N_n the number of solutions $(R_1, \ldots, R_k)$ of (IX.5) summed over k. However, it is known that $N_n = \exp[O(\sqrt{n})]$ (see Madras and Slade [C5] Theorem 3.1.4), so this factor is harmless and the claim follows. □

Our main result until Section IX.9 is the following LDP for the speed of the bridge polymer.

THEOREM IX.6 *For every $\beta \in (0,\infty)$ the family $(Q_n^{\beta,\mathrm{bridge}})$ defined by*

$$Q_n^{\beta,\mathrm{bridge}}(\cdot) = \mathbb{Q}_n^{\beta}\Big(\frac{1}{n}S_n \in \cdot \;\Big|\; \circledast_n\Big)$$

satisfies the LDP *on* $(0,1]$ *with rate* n *and with rate function* J_β *identified in* (IX.9) *and Lemma* IX.25 *below (see Fig. 18 below). The unique zero of* J_β *is* $\theta^*(\beta)$ *in Theorem* IX.2.

To prove Theorem IX.6 we will carry out the following *program*:

(I) Pick $\theta \in (0,1]$ and consider the quantity

$$\mathbb{Q}_n^{\beta}(S_n = \lceil \theta n \rceil | \circledast_n) = \frac{\widehat{K}_n(\theta)}{\int_{\theta\in(0,1]} d(\theta n)\widehat{K}_n(\theta)},$$

where

$$\widehat{K}_n(\theta) = \mathbb{E}\left(e^{-\beta \widehat{I}_n} 1_{\{S_n = \lceil \theta n \rceil\}} 1_{\circledast_n}\right) \tag{IX.7}$$

($\lceil \theta n \rceil$ and n must have the same parity).

(II) Show that there exists a function $\widehat{J}_\beta\colon (0,1] \to (0,\infty)$ such that

$$\lim_{n\to\infty} \frac{1}{n} \log \widehat{K}_n(\theta) = -\widehat{J}_\beta(\theta), \tag{IX.8}$$

with the property that $\theta \mapsto \widehat{J}_\beta(\theta)$ is continuous, strictly convex and minimal at $\theta^*(\beta)$. Identify $\widehat{J}_\beta$ in terms of a variational problem.

(III) Combine (I) and (II), to obtain

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{Q}_n^{\beta}(S_n = \lceil \theta n \rceil | \circledast_n) = -J_\beta(\theta),$$

with

$$J_\beta(\theta) = \widehat{J}_\beta(\theta) - \inf_{\theta\in(0,1]} \widehat{J}_\beta(\theta). \tag{IX.9}$$

Evidently $\theta \mapsto J_\beta(\theta)$ is also continuous, strictly convex and minimal at $\theta^*(\beta)$, which is its unique zero (see Fig. 18).

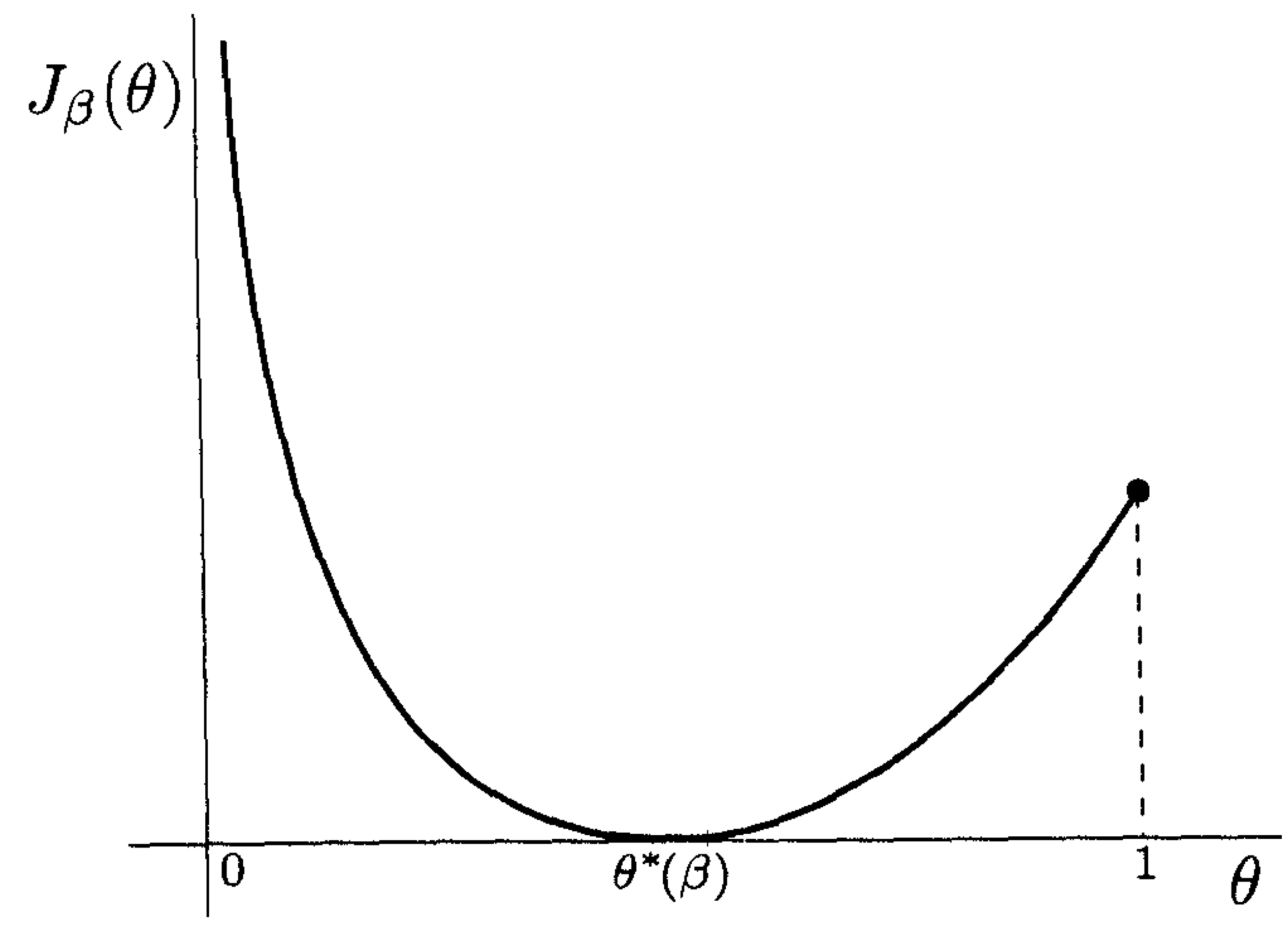

FIG. 18. The rate function J_β for bridge polymers

The argument below will show that the same results as above apply when θ is replaced by $\theta_n \to \theta$ as $n \to \infty$, which is why we get Theorem IX.6.

EXERCISE IX.10 *Show that* $\widehat{J}_\beta(1) = \log 2$.

EXERCISE IX.11 *(i) Use Lemma* III.29 *to show that the limit in* (IX.8) *exists for all* $\theta \in (0,1]$. *(ii) Show that the limit is a convex function of* θ.

The above program will be carried out in five steps, organized as Sections IX.4–IX.8. The first two steps are a preparation that is needed to get the key quantities in the right format for applying large deviation theory. The actual application of large deviation theory and the analysis of the ensuing variational problem are carried out in the last three steps.

After we have completed the proof of Theorem IX.6 we will show how to remove the bridge condition. This is done in Section IX.9 and leads to the LDP we are actually after. It will turn out that the associated rate function is different from J_β but still has $\theta^*(\beta)$ as its unique zero (see Fig. 20 below). Hence Theorem IX.2 will follow via the Borel-Cantelli lemma.

IX.4 Step 1: Adding drift

Fix $\theta \in (0,1)$. Let $\mathbb{P}_\theta, \mathbb{E}_\theta$ denote probability and expectation for the random walk with drift θ. Then we can write (IX.7) as

$$\widehat{K}_n(\theta) = (1-\theta)^{-\frac{n-\lceil\theta n\rceil}{2}}(1+\theta)^{-\frac{n+\lceil\theta n\rceil}{2}}\widetilde{K}_n(\theta), \qquad \text{(IX.12)}$$

with

$$\widetilde{K}_n(\theta) = \mathbb{E}_\theta\big(e^{-\beta\widehat{I}_n}1_{\{S_n=\lceil\theta n\rceil\}}1_{\circledast_n}\big). \qquad \text{(IX.13)}$$

Indeed, every path from 0 to $\lceil\theta n\rceil$ makes the same number of steps to the left and to the right, so we pick up a simple Radon-Nikodym factor. Thus it suffices to study the asymptotics of $\widetilde{K}_n(\theta)$, i.e., our task now is to relate the polymer with drift θ to the random walk with drift θ.

The advantage of this reformulation is that the path does not care to return to $[0,S_n]$ after time n.

LEMMA IX.14 *For all* $n \in \mathbb{N}_0$,

$$\mathbb{E}_\theta\big(e^{-\beta\widehat{I}_n}1_{\{S_n=\lceil\theta n\rceil\}}1_{\circledast_n}\big) = \frac{1}{\theta}\,\mathbb{E}_\theta\big(e^{-\beta\widehat{I}_n}1_{\{S_n=\lceil\theta n\rceil\}}1_{\circledast_n\cap\circledcirc_n}\big),$$

with

$$\circledcirc_n = \big\{S_i > S_n \ \forall i > n\big\}.$$

PROOF. Simply use that $\widehat{I}_n$ does not depend on S_i for $i > n$, and that $\mathbb{E}_\theta(\circledcirc_n) = \theta$ for all n. □

An important consequence of Lemma IX.14 is that on the event $\{S_n = \lceil\theta n\rceil\} \cap \circledast_n \cap \circledcirc_n$ we may write

$$\widehat{I}_n = \sum_{x=0}^{\lceil\theta n\rceil} \ell^2(x),$$

where

$$\ell(x) = \sum_{i\in\mathbb{N}_0} 1_{\{S_i=x\}}, \qquad x \in \mathbb{Z},$$

is the *total (!) local time at site* x. Indeed, this follows from the observation that on the event $\{S_n = \lceil \theta n \rceil\} \cap \circledast_n \cap \circledcirc_n$ we have

$$\ell_n(x) = \begin{cases} \ell(x) & \text{if } 0 \le x \le \lceil \theta n \rceil, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore we may rewrite (IX.13) as

$$\widetilde{K}_n(\theta) = \frac{1}{\theta} \mathbb{E}_\theta \Big(e^{-\beta \sum_{x=0}^{\lceil \theta n \rceil} \ell^2(x)} 1_{\{S_n = \lceil \theta n \rceil\}} 1_{\circledast_n \cap \circledcirc_n} \Big). \tag{IX.15}$$

The total local times turn out to have a nice structure, as we show next.

IX.5 Step 2: Markovian nature of the total local times

In this section we show that $\{\ell(x)\}_{x \in \mathbb{N}_0}$ admits a nice Markovian description, which will allow us to deduce the asymptotics of $\widetilde{K}_n(\theta)$ from an LDP for Markov chains. Let

$$m(x) = \sum_{i \in \mathbb{N}_0} 1_{\{S_i = x,\, S_{i+1} = x+1\}}, \qquad x \in \mathbb{Z},$$

the *total (!) number of jumps from* x *to* $x+1$. Then, on the event $\circledast_n \cap \circledcirc_n$, the total number of jumps from $x+1$ to x equals

$$\sum_{i \in \mathbb{N}_0} 1_{\{S_i = x+1,\, S_{i+1} = x\}} = m(x) - 1, \qquad 0 \le x \le \lceil \theta n \rceil,$$

because the net number of jumps along the bond between x and $x+1$ must be $+1$. Now, since $\ell(x)$ is the sum of the number of jumps to x coming from the left and from the right, we have

$$\ell(x) = m(x-1) + m(x) - 1, \qquad 0 \le x \le \lceil \theta n \rceil.$$

Moreover, since the total time spent between 0 and $\lceil \theta n \rceil$ is $n+1$, we have

$$\begin{aligned} &\{S_n = \lceil \theta n \rceil\} \cap \circledast_n \cap \circledcirc_n \\ &\quad = \Big\{ \sum_{x=0}^{\lceil \theta n \rceil} [m(x-1) + m(x) - 1] = n+1,\ m(-1) = 0,\ m(\lceil \theta n \rceil) = 1 \Big\}. \end{aligned}$$

Therefore we may rewrite (IX.15) as

$$\begin{aligned} \widetilde{K}_n(\theta) = \tfrac{1}{\theta} \mathbb{E}_\theta \Big(& e^{-\beta \sum_{x=0}^{\lceil \theta n \rceil} [m(x-1) + m(x) - 1]^2} \\ & \times 1_{\{\sum_{x=0}^{\lceil \theta n \rceil} [m(x-1) + m(x) - 1] = n+1\}} 1_{\{m(-1) = 0, m(\lceil \theta n \rceil) = 1\}} \Big). \end{aligned} \tag{IX.16}$$

The main reason for this reformulation is the following fact, which goes back to Knight [D42].

LEMMA IX.17 *Under the law* $\mathbb{P}_\theta$, $\{m(x)\}_{x \in \mathbb{N}_0}$ *is a Markov chain on state space* $\mathbb{N}$ *with transition kernel*

$$P_\theta(i,j) = \binom{i+j-2}{i-1} \Big(\frac{1+\theta}{2}\Big)^i \Big(\frac{1-\theta}{2}\Big)^{j-1}, \qquad i,j \in \mathbb{N}.$$

PROOF. Fix x. If $m(x) = i$, then the bond $(x, x+1)$ receives i upcrossings and $i-1$ downcrossings. For $s = 1, \dots, i-1$, let Z_s denote the number of upcrossings of $(x+1, x+2)$ in between the s-th upcrossing and the s-th downcrossing of $(x, x+1)$. Let Z denote the number of upcrossings of $(x+1, x+2)$ after the i-th upcrossing of $(x, x+1)$, which is different from the others because no further downcrossing of $(x, x+1)$ is allowed. Since the random walk has drift θ, the probability that it makes a loop excursion to the right of $x+1$ is $\frac{1-\theta}{2}$. Hence, we have

$$\begin{aligned} \mathbb{P}_\theta(Z_s = k | m(x) = i) &= \left(\tfrac{1-\theta}{2}\right)^{k+1}, \quad k \in \mathbb{N}_0,\ s = 1, \dots, i-1, \\ \mathbb{P}_\theta(Z = k | m(x) = i) &= \tfrac{1+\theta}{1-\theta}\left(\tfrac{1-\theta}{2}\right)^k, \quad k \in \mathbb{N}. \end{aligned}$$

Since $m(x+1) = j$ means that $Z_1 + \cdots + Z_{i-1} + Z = j$, we see that our process is Markov: it is irrelevant for the outcome of $m(x+1)$ what the random walk does to the left of x, only the value of $m(x)$ matters. Moreover, $Z_s + 1$ ($s = 1, \dots, i-1$) have the same law as Z modulo the normalizing prefactor $\frac{1+\theta}{1-\theta}$. Since $(Z_1 + 1) + \cdots + (Z_{i-1} + 1) + Z = i + j - 1$ and since the number of ways $i + j - 1$ can be divided into i pieces of length ≥ 1 equals the binomial factor, we obtain the above formula for $P_\theta(i, j)$. □

Note that $\{m(x)\}_{x \in \mathbb{N}_0}$ is a branching process with one immigrant and with an offspring distribution that has mean smaller than 1. Therefore our Markov chain is positive recurrent.

IX.6 Step 3: Key variational problem

In this section we derive the key variational problem underlying the LDP for bridges in Theorem IX.6. This can be done along fairly standard lines. However, in order not to get lost in too many technicalities, the reader is asked to make a few "small leaps of faith".

The nice fact about the representation in (IX.16) is that $\widetilde{K}_n(\theta)$ can be expressed in terms of the pair empirical measure associated with $\{m(x)\}_{x \in \mathbb{N}_0}$. To that end, define

$$L_N^2 = \frac{1}{N} \sum_{x=0}^{N-1} \delta_{(m(x-1), m(x))}, \qquad N \in \mathbb{N},$$

with periodic boundary conditions, and let

$$\begin{aligned} F_\beta(\nu) &= -\beta \sum_{i,j \in \mathbb{N}} (i+j-1)^2 \nu(i,j), \\ A_\theta &= \Big\{ \nu \in \widetilde{\mathfrak{M}}_1(\mathbb{N} \times \mathbb{N}) \colon \sum_{i,j \in \mathbb{N}} (i+j-1) \nu(i,j) = \tfrac{1}{\theta} \Big\}. \end{aligned}$$

Then (IX.16) becomes

$$\widetilde{K}_n(\theta) \simeq \mathbb{E}_\theta\Big(e^{N F_\beta(L_N^2)} 1_{\{L_N^2 \in A_\theta\}} \Big) \text{ with } N = \lceil \theta n \rceil + 1. \tag{IX.18}$$

Indeed,

1. The exponential factor in (IX.16) equals the one in (IX.18), with a negligible error arising from forcing the periodic boundary condition in the definition of L_N^2.
2. The first constraint in (IX.16) is asymptotically the same as the constraint in (IX.18), because we replaced $\frac{n+1}{\lceil \theta n \rceil + 1}$ by $\frac{1}{\theta}$, which will *a posteriori* be justified by the continuity of the function $\theta \mapsto \widetilde{J}_\beta(\theta)$ appearing in Lemma IX.19 below (see Sections IX.7–IX.8).

3. The second contraint in (IX.16) is negligible as $n \to \infty$.

The reason for introducing the representation in (IX.18) is of course that it allows us to use the LDP for L_N^2, as formulated in Theorem IV.3, based on the Markov property established in Lemma IX.17.

LEMMA IX.19 *For every* $\theta \in (0,1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \widetilde{K}_n(\theta) = -\widetilde{J}_\beta(\theta),$$

with

$$\widetilde{J}_\beta(\theta) = \theta \inf_{\nu\in A_\theta} \big[- F_\beta(\nu) + I^2_{P_\theta}(\nu)\big],$$

where

$$I^2_{P_\theta}(\nu) = \sum_{i,j\in\mathbb{N}} \nu(i,j) \log\Big(\frac{\nu(i,j)}{\bar{\nu}(i)P_\theta(i,j)}\Big).$$

PROOF. From Lemma IV.3 we see that $I^2_{P_\theta}(\nu)$ is the weak rate function in the weak LDP for (L_N^2). In order to apply Varadhan's Lemma we need the LDP, i.e., we need to overcome the fact that the state space $\mathbb{N}$ is infinite (recall Remark III.24.3). This can be handled via a truncation argument because, as was observed at the end of Section IX.5, the Markov chain $\{m(x)\}_{x\in\mathbb{N}_0}$ has strong recurrence properties (see Greven and den Hollander [D31]).

Next we apply Theorem III.13 (Varadhan's Lemma) to (IX.18), which is an exponential integral restricted to the set A_θ. Here another technical difficulty arises: before Theorem III.13 can be applied, the weak LDP "needs to be transferred from $\widetilde{\mathfrak{M}}_1(\mathbb{N}\times\mathbb{N})$ to A_θ" (see Dembo and Zeitouni [A2] Lemma 4.1.5). The result of the usual manipulations reads

$$\begin{aligned}\widetilde{K}_n(\theta) \simeq \mathbb{E}_\theta\Big(e^{NF_\beta(L_N^2)}1_{\{L_N^2\in A_\theta\}}\Big) &= \int_{A_\theta} e^{NF_\beta(L_N^2)}\mathbb{P}_\theta(L_N^2 \in d\nu)\\ &\simeq e^{N\sup_{\nu\in A_\theta}[F_\beta(\nu)-I^2_{P_\theta}(\nu)]}, \qquad N\to\infty,\end{aligned}$$

which proves the claim because $N = \lceil \theta n\rceil + 1$. □

At this point we recall (IX.12) and (IX.13), and rewrite Lemma IX.19 as follows:

LEMMA IX.20 *For every* $\theta \in (0,1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \widehat{K}_n(\theta) = -\widehat{J}_\beta(\theta), \tag{IX.21}$$

with

$$\widehat{J}_\beta(\theta) = \theta \inf_{\nu\in A_\theta} \big[- F_\beta(\nu) + I^2_{P_0}(\nu)\big], \tag{IX.22}$$

i.e., the same variational formula as in Lemma IX.19 *but with* P_θ *replaced by* P_0, *given by (recall Lemma* IX.17*)*

$$P_0(i,j) = \binom{i+j-2}{i-1}\Big(\frac{1}{2}\Big)^{i+j-1}, \qquad i,j\in\mathbb{N}.$$

PROOF. Simply note that $\nu \in A_\theta$ implies $\sum_{i\in\mathbb{N}} i\bar{\nu}(i) = \frac{1+\theta}{2\theta}$, so that

$$\begin{aligned} I^2_{P_0}(\nu) - I^2_{P_\theta}(\nu) &= \sum_{i,j\in\mathbb{N}} \nu(i,j)\log\big[(1+\theta)^i(1-\theta)^{j-1}\big] \\ &= \tfrac{1+\theta}{2\theta}\log(1+\theta) + \tfrac{1-\theta}{2\theta}\log(1-\theta), \qquad \nu \in A_\theta, \end{aligned}$$

which makes the prefactor in (IX.12) cancel out. □

Thus we have identified $\widehat{J}_\beta(\theta)$ for $\theta \in (0,1)$, which is the function we were after in Section IX.3. In view of Exercise IX.10, the same formulas as in (IX.21) and (IX.22) apply for $\theta = 1$. Finally, (IX.9) gives us J_β, the rate function in the LDP for bridge polymers in Theorem IX.6.

IX.7 Step 4: Analysis of the rate function for bridges

We next proceed to give the solution of the variational problem in (IX.22), leading to the qualitative shape of the function $\theta \mapsto J_\beta(\theta)$ anticipated in Section VI.2 (see Fig. 18). Note that the variational problem is far from trivial, because it requires minimizing a non-linear functional under a linear constraint. Fortunately, it is possible to find the solution in terms of a certain eigenvalue problem that is well-behaved, and we will see that the outcome is relatively simple.

Fix $\beta \in (0,\infty)$ and $r \in \mathbb{R}$, and let $A_{r,\beta}$ be the $\mathbb{N}\times\mathbb{N}$-matrix with components

$$A_{r,\beta}(i,j) = e^{r(i+j-1)-\beta(i+j-1)^2} P_0(i,j), \qquad i,j \in \mathbb{N}.$$

The parameter r will be seen to play the role of a Lagrange multiplier needed to handle the constraint in (IX.22).

LEMMA IX.23 *Fix $\beta \in (0,\infty)$. For every $r \in \mathbb{R}$, $A_{r,\beta}$ is a self-adjoint operator on $l^2(\mathbb{N})$ having a unique largest eigenvalue $\lambda_{r,\beta}$ and corresponding eigenvector $\tau_{r,\beta}$ (normalized as $\|\tau_{r,\beta}\|_2 = 1$).*

PROOF. Since $A_{r,\beta}$ is strictly positive and has rapidly decaying tails, the assertion follows from standard Perron-Frobenius theory. In fact, $A_{r,\beta}$ has the so-called Hilbert-Schmidt property $\sum_{i,j\in\mathbb{N}} A_{r,\beta}(i,j)^2 < \infty$ and, consequently, is a compact operator (see Dunford and Schwartz [C1] Section XI.6). □

The eigenvalue $\lambda_{r,\beta}$ has the following properties:

LEMMA IX.24 (i) *$(r,\beta) \mapsto \lambda_{r,\beta}$ is analytic on $\mathbb{R}\times(0,\infty)$.*
(ii) *$\lim_{r\to-\infty} \frac{\partial}{\partial r}\log\lambda_{r,\beta} = 1$ and $\lim_{r\to\infty} \frac{\partial}{\partial r}\log\lambda_{r,\beta} = \infty$ for all $\beta \in (0,\infty)$.*
(iii) *$r \mapsto \log\lambda_{r,\beta}$ is stricly convex for all $\beta \in (0,\infty)$.*

PROOF. For details we refer to Greven and den Hollander [D31].
(i) Analyticity holds because $\lambda_{r,\beta}$ has multiplicity 1 and all elements of the matrix $A_{r,\beta}$ are analytic.
(ii) This follows from straightforward estimates on the eigenvector $\tau_{r,\beta}$ for $r \to -\infty$ and $r \to \infty$, respectively.
(iii) Convexity follows from the observations:

1. $\lambda_{r,\beta} = \sup_{x\in l^2(\mathbb{N}):\, x>0,\, \|x\|_2=1} \sum_{i,j\in\mathbb{N}} x(i)A_{r,\beta}(i,j)x(j)$;
2. $r \mapsto \log A_{r,\beta}(i,j)$ is linear for all $i,j \in \mathbb{N}$ and $\beta \in (0,\infty)$;
3. log-convexity is preserved under taking sums and suprema.

Strict convexity follows from convexity in combination with (i) and (ii). □

With the help of Lemma IX.24 we can express $\widehat{J}_\beta(\theta)$ in terms of the eigenvalue $\lambda_{r,\beta}$ for some r depending on θ.

LEMMA IX.25 *Fix $\beta \in (0,\infty)$. Then, for every $\theta \in (0,1)$,*

$$\widehat{J}_\beta(\theta) = r - \theta \log \lambda_{r,\beta}\Big|_{r=r_\beta(\theta)},$$

where $r_\beta(\theta) \in \mathbb{R}$ is the unique solution of the equation

$$\frac{1}{\theta} = \frac{\partial}{\partial r} \log \lambda_{r,\beta}. \tag{IX.26}$$

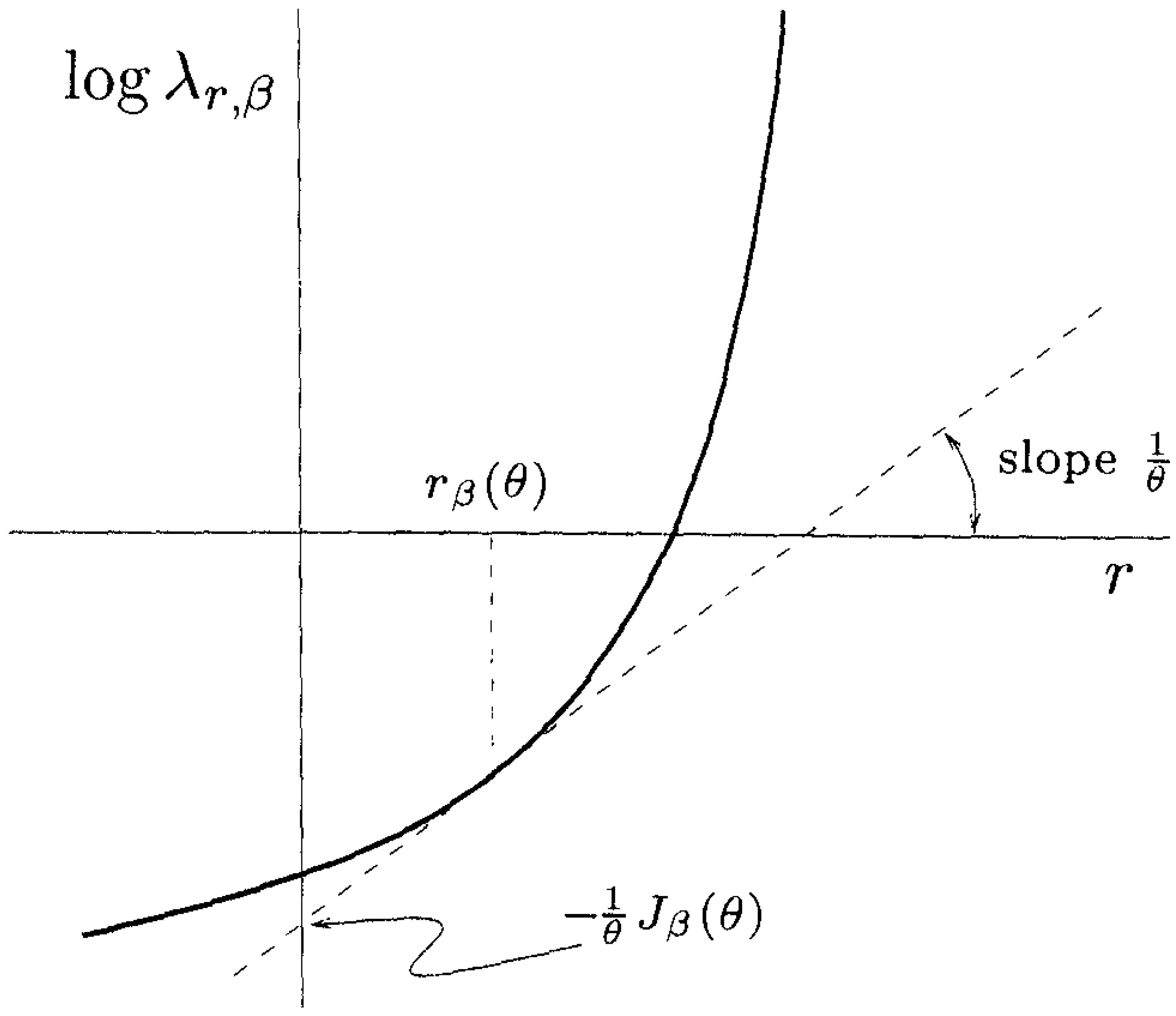

FIG. 19. Identification of $\widehat{J}_\beta(\theta)$

PROOF. The fact that (IX.26) has a solution for all $\theta \in (0,1)$ and that this solution is unique follows from Lemma IX.24 (see Fig. 19).

Consider the following family of pair probability measures:

$$\nu_{r,\beta}(i,j) = \frac{1}{\lambda_{r,\beta}} \tau_{r,\beta}(i) A_{r,\beta}(i,j) \tau_{r,\beta}(j).$$

One easily checks that $\nu_{r,\beta} \in \widetilde{\mathfrak{M}}_1(\mathbb{N}\times\mathbb{N})$. Compute

$$\begin{aligned} I_P^2(\nu_{r,\beta}) &= \sum_{i,j} \nu_{r,\beta}(i,j) \log\Big(\tfrac{\nu_{r,\beta}(i,j)}{\bar{\nu}_{r,\beta}(i) P(i,j)}\Big) \\ &= \sum_{i,j} \nu_{r,\beta}(i,j) \Big[r(i+j-1) - \beta(i+j-1)^2 - \log \lambda_{r,\beta} + \log \tfrac{\tau_{r,\beta}(j)}{\tau_{r,\beta}(i)} \Big], \end{aligned}$$

where we use that $\bar{\nu}_{r,\beta}(i) = \tau_{r,\beta}^2(i)$. Since $\nu_{r,\beta}$ has identical marginals, the last term vanishes and we end up with the simple expression

$$I_P^2(\nu_{r,\beta}) = \frac{r}{\theta} + F_\beta(\nu_{r,\beta}) - \log \lambda_{r,\beta},$$

provided (!) $\nu_{r,\beta} \in A_\theta$. This expression says that

$$-\theta\big[F_\beta(\nu_{r,\beta}) - I_P^2(\nu_{r,\beta})\big] = r - \theta \log \lambda_{r,\beta}.$$

We thus see from Lemma IX.20 that the claim in Lemma IX.25 is correct provided (!) we can prove the following two properties:

(i) $r = r_\beta(\theta)$ implies $\nu_{r,\beta} \in A_\theta$;

(ii) $\nu_{r_\beta(\theta),\beta}$ is a minimizer of the variational problem in (IX.22).

Property (i) : Compute

$$\begin{aligned}\sum_{i,j}(i+j-1)\nu_{r,\beta}(i,j) &= \tfrac{1}{\lambda_{r,\beta}}\sum_{i,j}\tau_{r,\beta}(i)\Big[\tfrac{\partial}{\partial r}A_{r,\beta}(i,j)\Big]\tau_{r,\beta}(j)\\ &= \tfrac{1}{\lambda_{r,\beta}}\tfrac{\partial}{\partial r}\Big[\sum_{i,j}\tau_{r,\beta}(i)A_{r,\beta}(i,j)\tau_{r,\beta}(j)\Big] = \tfrac{1}{\lambda_{r,\beta}}\tfrac{\partial}{\partial r}\lambda_{r,\beta} = \tfrac{\partial}{\partial r}\log\lambda_{r,\beta}.\end{aligned}$$

The second equality uses that $A_{r,\beta}\tau_{r,\beta} = \lambda_{r,\beta}\tau_{r,\beta}$ and $\|\tau_{r,\beta}\|_2 = 1$ for all r,β. Hence (IX.26) indeed guarantees (i).

Property (ii) : If $\nu \in A_\theta$, then we can write

$$\begin{aligned}-\theta\big[F_\beta(\nu) - I_P^2(\nu)\big] &= [r - \theta\log\lambda_{r,\beta}] + \theta\sum_{i,j}\nu(i,j)\log\Big(\frac{\nu(i,j)}{\bar\nu(i)\frac{1}{\lambda_{r,\beta}}A_{r,\beta}(i,j)}\Big)\\ &= [r - \theta\log\lambda_{r,\beta}] + \theta\sum_{i,j}\nu(i,j)\log\Big(\frac{\nu(i,j)}{\bar\nu(i)}\frac{\sqrt{\bar\nu_{r,\beta}(i)\bar\nu_{r,\beta}(j)}}{\nu_{r,\beta}(i,j)}\Big),\end{aligned}$$

where we once again use that $\bar\nu_{r,\beta}(i) = \tau^2_{r,\beta}(i)$. The first term is precisely the value we found when $\nu = \nu_{r,\beta}$. Moreover, because ν has identical marginals the second term simplifies further to

$$\theta\sum_{i,j}\nu(i,j)\log\Big(\frac{\nu(i,j)}{\bar\nu(i)}\frac{\bar\nu_{r,\beta}(i)}{\nu_{r,\beta}(i,j)}\Big) = \theta\sum_i\bar\nu(i)H\big([\nu(i)]\,|\,[\nu_{r,\beta}(i)]\big),$$

where $[\nu(i)], [\nu_{r,\beta}(i)] \in \mathfrak{M}_1(\mathbb{N})$ are defined by $[\nu(i)](j) = \nu(i,j)/\nu(i)$, $[\nu_{r,\beta}(i)](j) = \nu_{r,\beta}(i,j)/\nu_{r,\beta}(i)$, and H denotes relative entropy. Clearly, this is ≥ 0 with equality when $\nu = \nu_{r,\beta}$. □

IX.8 Step 5: Identification of the speed

Lemma IX.25 gives us a nice representation of $\widehat{J}_\beta(\theta)$, $\theta \in (0,1)$, in terms of the family of eigenvalues $\lambda_{r,\beta}$, $r \in \mathbb{R}$. No explicit formula is known for the latter quantity, but we can certainly manipulate it analytically. As we saw in Section IX.2, $\theta^*(\beta)$ is to be identified as the unique minimum of $\theta \mapsto \widehat{J}_\beta(\theta)$.

LEMMA IX.27 *Fix $\beta \in (0,\infty)$. Then*

$$\frac{1}{\theta^*(\beta)} = \frac{\partial}{\partial r}\log\lambda_{r,\beta}\Big|_{r=r^*(\beta)},$$

with $r^*(\beta) \in (0,\infty)$ the unique solution of the equation

$$\lambda_{r,\beta} = 1. \tag{IX.28}$$

PROOF. The fact that (IX.28) has a solution for all $\beta \in (0,\infty)$ and that this solution is unique follows from Lemma IX.24 (see Fig. 19).

Differentiate $\widehat{J}_\beta(\theta)$ with respect to θ to obtain

$$\frac{\partial}{\partial\theta}\widehat{J}_\beta(\theta) = \frac{\partial}{\partial\theta}r_\beta(\theta) - \log\lambda_{r_\beta(\theta),\beta} - \theta\Big[\frac{\partial}{\partial\theta}r_\beta(\theta)\Big]\Big[\frac{\partial}{\partial r}\log\lambda_{r,\beta}\Big]_{r=r_\beta(\theta)}.$$

However, the first and the third term cancel out because of (IX.26), so we get

$$\frac{\partial}{\partial\theta}\widehat{J}_\beta(\theta) = -\log\lambda_{r_\beta(\theta),\beta}. \tag{IX.29}$$

This is zero if and only if θ is such that $\lambda_{r_\beta(\theta),\beta} = 1$, i.e., the minimum $\theta^*(\beta)$ of $\theta \mapsto \widehat{J}_\beta(\theta)$ is found by solving (IX.28). After that we put

$$r_\beta(\theta^*(\beta)) = r^*(\beta) \tag{IX.30}$$

and use (IX.26). Note that, by Lemma IX.25,

$$\frac{\partial^2}{\partial\theta^2}\widehat{J}_\beta(\theta) = -\frac{1}{\theta}\frac{\partial}{\partial\theta}r_\beta(\theta) > 0 \tag{IX.31}$$

(see Fig. 19 and note that $\theta \mapsto r_\beta(\theta)$ has a negative slope), so that $r_\beta(\theta^*(\beta))$ is indeed the unique minimizer of $\theta \mapsto \widehat{J}_\beta(\theta)$. □

Lemmas IX.24, IX.25 and IX.27 yield Fig. 18. This finishes our analysis of the rate function J_β for bridge polymers, and the proof of Theorem IX.6 is now complete.

IX.9 The LDP without the bridge condition

In Sections IX.4–IX.8 we have proved Theorem IX.6, the LDP for bridge polymers. In this section we give a quick sketch of how to obtain the LDP without the bridge condition, which is the result we want in view of Theorems IX.2 and IX.3. Remarkably, it turns out that the rate function in this LDP has a linear piece between 0 and a *critical speed* $\theta^{**}(\beta)$ that is strictly smaller than $\theta^*(\beta)$ (see Fig. 20). The argument that follows developed from discussions with Wolfgang König.

THEOREM IX.32 *For every $\beta \in (0,\infty)$ the family (Q_n^β) defined by*

$$Q_n^\beta(\cdot) = \mathbb{Q}_n^\beta\Big(\frac{1}{n}S_n \in \cdot \;\Big|\; S_n \geq 0\Big)$$

satisfies the LDP *on* $[0,1]$ *with rate n and with rate function I_β given by*

$$I_\beta(\theta) = \begin{cases} J_\beta(\theta) & \text{if } \theta \geq \theta^{**}(\beta), \\ I_\beta(0) + \frac{\theta}{\theta^{**}(\beta)}[J_\beta(\theta^{**}(\beta)) - I_\beta(0)] & \text{if } \theta \leq \theta^{**}(\beta), \end{cases} \tag{IX.33}$$

where $\theta^{**}(\beta)$ and $I_\beta(0)$ are identified in (IX.37) *and Lemma* IX.41 *below. Moreover,* $\theta^{**}(\beta) \in (0, \theta^*(\beta))$.

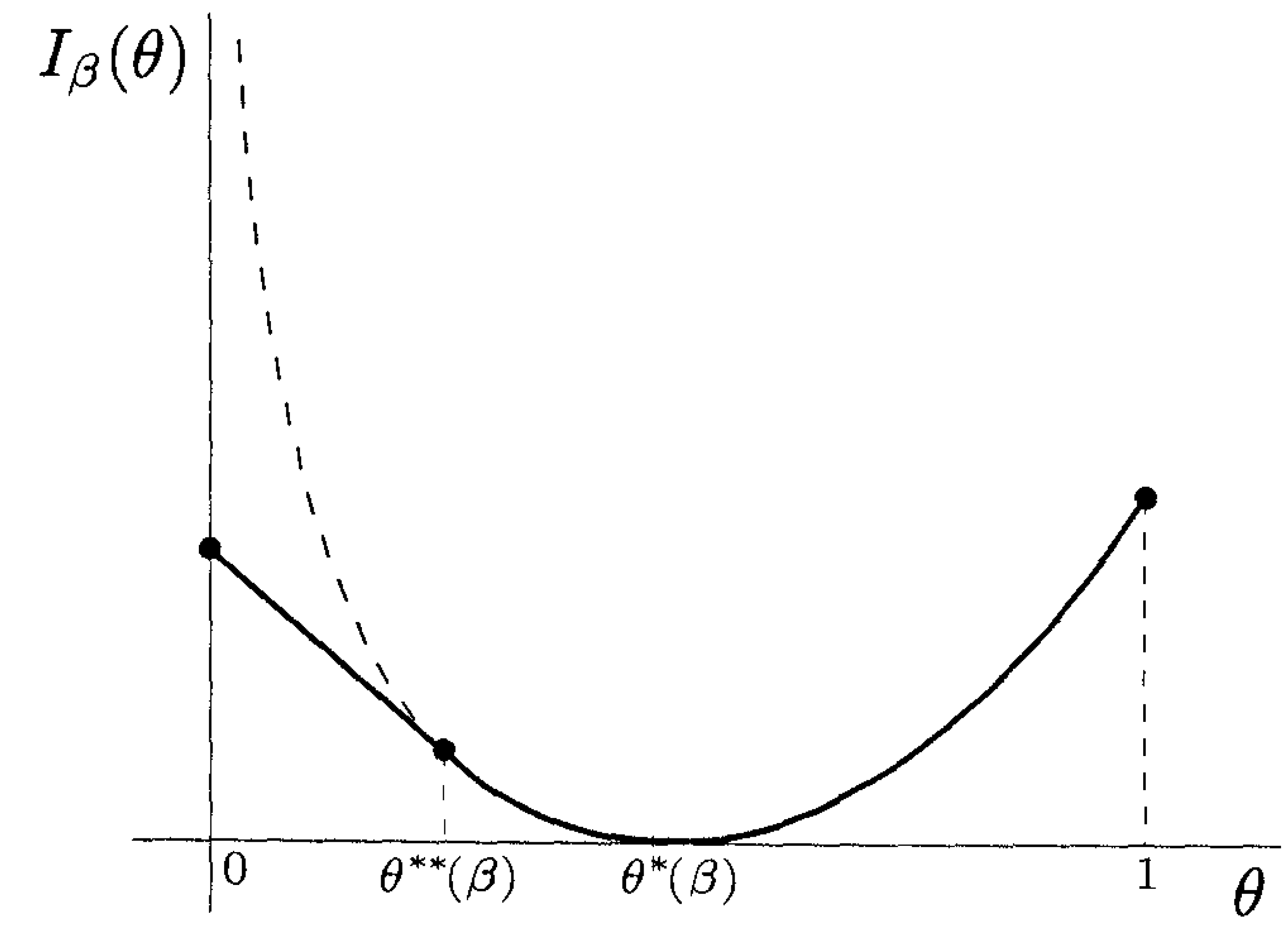

FIG. 20. The rate function I_β (compare with Fig. 18)

The linear piece can be understood as follows. If the polymer is required to move at a speed $\theta < \theta^{**}(\beta)$, then it prefers to violate the bridge condition by moving at speed $\theta^{**}(\beta)$ between 0 and $\lceil \theta n \rceil$, and making two loops, one below 0 and one above $\lceil \theta n \rceil$. The penalty for making these loops is less than the penalty for staying locked up like a bridge. The total length of these loops is proportional to $\theta^{**}(\beta) - \theta$, i.e., the penalty for not behaving like a bridge grows linearly with $\theta^{**}(\beta) - \theta$.

Let us now sketch how Theorem IX.32 comes about. In order not to get buried in another long calculation, we will omit the technical details. These can all be worked out in complete analogy with what was done for bridges.

For $\theta \in (0,1]$, define

$$-J_\beta^{\circlearrowleft}(\theta) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{Q}_n^\beta\Big(\max_{0\le i\le n} S_i = \lceil \theta n \rceil,\ S_i \ge 0\ \forall\, 0 \le i \le n \ \Big|\ S_n = 0\Big) \quad \text{(IX.34)}$$

(with the limit along even integers). This is the rate function for performing a loop that grows at speed θ. Let

$$j_\beta^{\circlearrowleft} = \inf_{\theta\in(0,1]} J_\beta^{\circlearrowleft}(\theta).$$

LEMMA IX.35 *For all $\theta \in (0,1]$,*

$$I_\beta(\theta) = \inf_{t_1,t_2\ge 0,\, t_1+t_2\le 1} \Big[(1-t_1-t_2) J_\beta\Big(\frac{\theta}{1-t_1-t_2}\Big) + (t_1+t_2) j_\beta^{\circlearrowleft}\Big].$$

PROOF. Fix t_1, t_2. The path can decide to first spend a time $t_1 n$ to the left of 0, next spend a time $(1-t_1-t_2)n$ between 0 and $\lceil \theta n \rceil$, and finally spend a time $t_2 n$ to the right of $\lceil \theta n \rceil$. The cost for doing so is given by the three terms in the RHS of the last display. After optimizing over t_1, t_2 we get the claim. Of course, we must rule out that the path prefers a more complicated strategy. However, this can be done via a counting argument similar to that in the proof of Lemma IX.4. □

To simplify the variational problem in Lemma IX.35, we make an important observation.

LEMMA IX.36 $I_\beta(0) = j_\beta^{\circlearrowleft}$.

PROOF. Suppose that the polymer forms a loop. Then this loop lies either left or right of 0, or it consists of several smaller loops. By showing that the latter alternative is more costly, we get the claim. Again, the details can be filled in via a counting argument similar to that in the proof of Lemma IX.4. □

Combining Lemmas IX.35 and IX.36 we arrive at the variational problem

$$I_\beta(\theta) = \inf_{0\le t\le 1}\Big[(1-t)J_\beta\Big(\frac{\theta}{1-t}\Big) + tI_\beta(0)\Big].$$

Rewrite this formula as

$$I_\beta(\theta) - I_\beta(0) = \inf_{0\le s\le 1}\Big[s\Big\{J_\beta\Big(\frac{\theta}{s}\Big) - I_\beta(0)\Big\}\Big].$$

Because $\theta \mapsto J_\beta(\theta)$ is strictly convex (recall (IX.9) and (IX.29)), it follows that the supremum is attained at

$$s = \begin{cases} 1 & \text{if } \theta \ge \theta^{**}(\beta), \\ \frac{\theta}{\theta^{**}(\beta)} & \text{if } \theta \le \theta^{**}(\beta), \end{cases}$$

where $\theta^{**}(\beta)$ is the unique solution of the equation

$$J_\beta(\theta) - I_\beta(0) = \theta\frac{\partial}{\partial\theta}J_\beta(\theta). \tag{IX.37}$$

This completes the proof of Theorem IX.32. Note that (IX.37) says that the linear piece in I_β is tangent to J_β at $\theta = \theta^{**}(\beta)$. Hence we have $\theta^{**}(\beta) \in (0, \theta^*(\beta))$ as soon as $I_\beta(0) > 0$.

To complete the computation of I_β, it remains to identify $I_\beta(0)$ and show that $I_\beta(0) > 0$. This can be done with the help of a spectral analysis similar as in Sections IX.7 and IX.8. Let

$$-\widehat{J}_\beta^{\circlearrowleft}(\theta) = \lim_{n\to\infty}\frac{1}{n}\log \mathbb{E}\left(e^{-\beta \widehat{I}_n} 1_{\{\max_{0\le i\le n} S_i = \lceil \theta n\rceil,\ S_i \ge 0\ \forall 0\le i\le n\}}\ \Big|\ S_n = 0\right), \tag{IX.38}$$

i.e., the same as in (IX.34) but without the normalizing constant Z_n^β (recall (IX.1)). Since $\lim_{n\to\infty}\frac{1}{n}\log Z_n^\beta = -r^*(\beta) = -\inf_{\theta\in(0,1]}\widehat{J}_\beta(\theta)$ by the results in Sections IX.7 and IX.8 (see Fig. 18 and Lemmas IX.25 and IX.27), we have

$$J_\beta^{\circlearrowleft}(\theta) = \widehat{J}_\beta^{\circlearrowleft}(\theta) - r^*(\beta). \tag{IX.39}$$

LEMMA IX.40 *For every* $\theta \in (0,1)$,

$$\widehat{J}_\beta^{\circlearrowleft}(\theta) = \theta \inf_{\nu\in A_\theta^{\circlearrowleft}}[-F_\beta^{\circlearrowleft}(\nu) + I^2_{P_0^{\circlearrowleft}}(\nu)],$$

where

$$F_\beta^{\circlearrowleft}(\nu) = -\beta \sum_{i,j\in\mathbb{N}_0}(i+j)^2\nu(i,j),$$

$$A_\theta^{\circlearrowleft} = \Big\{\nu \in \widetilde{\mathfrak{M}}_1(\mathbb{N}_0\times\mathbb{N}_0)\colon \sum_{i,j\in\mathbb{N}_0}(i+j)\nu(i,j) = \tfrac{1}{\theta},\ \nu(\mathbb{N}\times\mathbb{N}) = 1\Big\},$$

and $I^2_{P_0^{\circlearrowleft}}(\nu)$ *is the same as* $I^2_{P_0}$ *in Lemma* IX.22 *but with* P_0 *replaced by (recall Lemma* IX.17*)*

$$P_0^{\circlearrowleft}(i,j) = 1_{\{i\ne 0\}}P_0(i,j+1) + 1_{\{i=j=0\}}, \qquad i,j\in\mathbb{N}_0.$$

PROOF. These are the same formulas as for $\widehat{J}_\beta(\theta)$ in Lemma IX.20, except that everywhere (i,j) is replaced by $(i,j-1)$. This can be traced back to the fact that now we are dealing with looped paths rather than paths that move to the right (see the second display in Section IX.5). We have the extra condition $\nu(\mathbb{N}\times\mathbb{N}) = 1$, because the Markov chain with kernel $P_0^{\circlearrowleft}$ has 0 as an absorbing state. This absorbing state must be avoided during $\lceil\theta n\rceil$ steps of the Markov chain in order to get a loop between 0 and $\lceil\theta n\rceil$. □

The analogue of Lemmas IX.25 and IX.27 reads as follows:

LEMMA IX.41 *Fix $\beta \in (0,\infty)$. For $r \in \mathbb{R}$, let $A_{r,\beta}^{\circlearrowleft}$ be the $\mathbb{N}\times\mathbb{N}$-matrix with components*

$$A_{r,\beta}^{\circlearrowleft}(i,j) = e^{r(i+j)-\beta(i+j)^2} P_0^{\circlearrowleft}(i,j), \qquad i,j \in \mathbb{N}.$$

Let $\lambda_{r,\beta}^{\circlearrowleft}$ be the unique largest eigenvalue of $A_{r,\beta}^{\circlearrowleft}$ acting as an operator on $l^2(\mathbb{N})$. Then

$$I_\beta(0) = r^{**}(\beta) - r^*(\beta), \tag{IX.42}$$

with $r^{**}(\beta) \in (0,\infty)$ the unique solution of the equation

$$\lambda_{r,\beta}^{\circlearrowleft} = 1. \tag{IX.43}$$

PROOF. By Lemma IX.36 and (IX.39), we have

$$I_\beta(0) = \inf_{\theta\in(0,1]} J_\beta^{\circlearrowleft}(\theta) = \inf_{\theta\in(0,1]} \widehat{J}_\beta^{\circlearrowleft}(\theta) - r^*(\beta).$$

However, the last infimum equals $r^{**}(\beta)$. The proof of this fact follows the same line of argument as in Sections IX.7 and IX.8. Even though $A_{r,\beta}$ is not symmetric, its largest eigenvalue is real because all its components are positive (Seneta [C8] Chapter 6). Note that we throw out $i,j=0$. □

EXERCISE IX.44 *Show that $r^{**}(\beta) > r^*(\beta)$ for all $\beta \in (0,\infty)$.*

EXERCISE IX.45 *Explain why (IX.37) says that $r^{**}(\beta) = r_\beta(\theta^{**}(\beta))$, with $\theta \to r_\beta(\theta)$ defined in Lemma IX.25. Combine this relation with Exercise IX.44 to prove that $\theta^{**}(\beta) < \theta^*(\beta)$ for all $\beta \in (0,\infty)$.*

Theorem IX.32 implies Theorems IX.2 and IX.3(a).

EXERCISE IX.46 *Show that $\theta^*(\beta) \in (0,1)$ for all $\beta \in (0,\infty)$.*

EXERCISE IX.47 *Prove Theorem IX.3(b).*

Thus, after a fairly long and complicated series of steps, we have managed to prove the existence of the speed $\theta^*(\beta)$ of the polymer and to identify it in terms of an eigenvalue problem that is sufficiently tractable to allow for some interesting statements about its β-dependence.

IX.10 Concluding remarks

(1) Lemmas IX.25 and IX.27 state that

$$\widehat{J}_\beta(\theta) = \theta \sup_{r\in\mathbb{R}} \left[\frac{r}{\theta} - \log \lambda_{r,\beta}\right],$$

which reminds us of Lemma VII.10(i). This is no surprise, because of the link we saw in Section V.4, Part 3. In the present chapter we did not follow the route of the Gärtner-Ellis Theorem, contrary to what was done in Chapter VII. The reason is that the polymer problem is not so well adapted to this route.

(2) It is natural to conjecture that $\beta \mapsto \theta^*(\beta)$ is non-decreasing (see Fig. 17). Even though this property seems intuitively plausible, it is actually rather deep (see Greven and den Hollander [D31]) and unfortunately remains open. It is not hard to compute $\lambda_{r,\beta}$ numerically and get support for the monotonicity.

EXERCISE IX.48 *Use Lemma IX.27 to see what* $\frac{d}{d\beta}\theta^*(\beta) \geq 0$ *amounts to in terms of the family* $\{\lambda_{r,\beta}\colon r \in \mathbb{R}, \beta \in (0,\infty)\}$.

(3) It is proved in van der Hofstad and den Hollander [D38] that $\theta^*(\beta) \sim C\beta^{\frac{1}{3}}$, $\beta \downarrow 0$, for some constant $C > 0$ (see Fig. 17). This asymptotics shows that, even in the limit of weak self-repellence, the behavior of the polymer cannot be understood via a perturbation argument around the non-repellent SRW. In other words, large deviation theory is vital for the whole description.

(4) Theorem IX.2 has been extended to a CLT by König [D43]. The standard deviation, denoted by $\sigma^*(\beta)$, turns out to be given by the formula

$$\frac{1}{\sigma^{*2}(\beta)} = \frac{\partial^2}{\partial\theta^2} J_\beta(\theta)\Big|_{\theta=\theta^*(\beta)} = \frac{\partial^2}{\partial\theta^2} I_\beta(\theta)\Big|_{\theta=\theta^*(\beta)},$$

which is in harmony with Lemma I.14(vi). Numerical calculation gives the following picture:

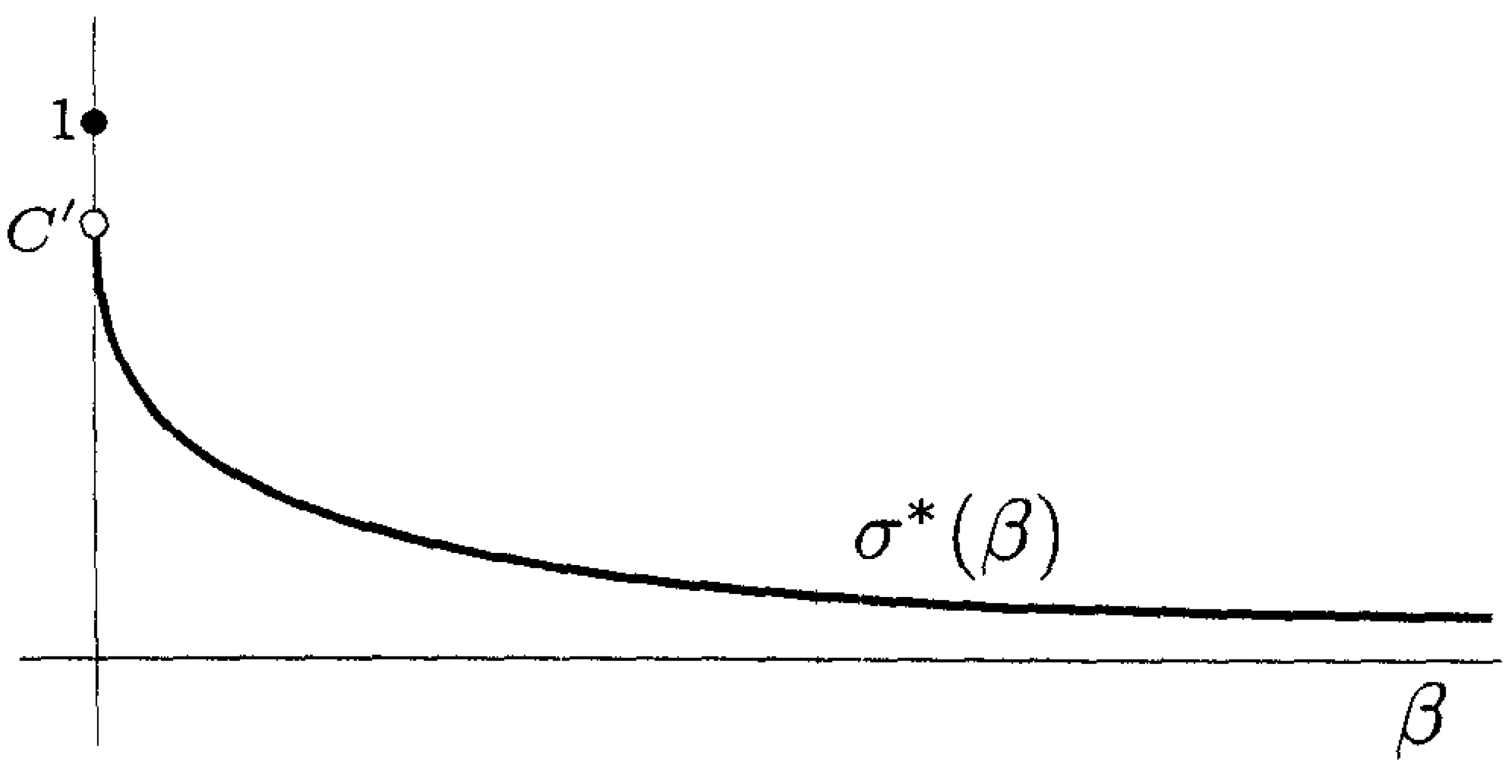

FIG. 21. The spread of the polymer

In van der Hofstad, den Hollander and König [D39] it is proved that $\lim_{\beta\downarrow 0} \sigma^*(\beta) = C'$ for some constant $C' \neq 1$. Hence there is a discontinuity at $\beta = 0$, which is another manifestation of the fact that perturbation arguments are doomed to fail.

(5) Van der Hofstad [D37] derives rigorous bounds on C and C'. Numerically, $C \approx 1.1$ and $C' \approx 0.63$. These constants are related to a Brownian version of the polymer model (see van der Hofstad [B5]).

(6) It is possible to prove that, under the law $\mathbb{Q}_n^\beta(\,\cdot\,|S_n \geq 0)$,

$$\left\{\frac{1}{\sigma^*(\beta)\sqrt{n}}\Big(S_{\lfloor tn\rfloor} - \lfloor tn\rfloor\theta^*(\beta)\Big)\right\}_{0\leq t\leq 1}$$

converges weakly to standard Brownian motion as $n \to \infty$.

CHAPTER X

INTERACTING DIFFUSIONS

In this chapter we give an application of large deviation theory in statistical physics. Namely, we show how to apply Sanov's theorem to study *the evolution of a system of diffusions with a mean field interaction.* It turns out that the typical evolution of the system can be related to the unique zero of a rate function for the empirical measure on path space. The results to be described can be found in Dawson and Gärtner [D14] and in Ben Arous and Brunaud [D4] for the case of a *non-random* interaction, and in Dai Pra and den Hollander [D13] for the case of a *random* interaction. We follow the line of reasoning in the last reference.

X.1 The interaction Hamiltonian

Let $H_N \colon \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be the N-particle random Hamiltonian given by

$$H_N(\underline{x}, \underline{\omega}) = \frac{1}{2N} \sum_{i,j=1}^{N} f(x_j - x_i; \omega_i, \omega_j) + \sum_{i=1}^{N} g(x_i; \omega_i), \tag{X.1}$$

where

$$\underline{x} = (x_i)_{i=1}^N \qquad \text{and} \qquad \underline{\omega} = (\omega_i)_{i=1}^N$$

are the *state* variable and the *medium* variable, respectively. Assume that

$$\underline{\omega} \text{ has distribution } \mu^N \tag{X.2}$$

for some probability law μ on $\mathbb{R}$. For a given realization of $\underline{\omega}$, think of $H_N(\underline{x}, \underline{\omega})$ as the energy of the state $\underline{x}$ in the medium $\underline{\omega}$. The functions f and g play the role of a pair potential and an external field, respectively, and are assumed to satisfy:

1. f, f', f'' and g, g', g'' are bounded and jointly continuous in all variables, where $'$ denotes the derivative with respect to the first argument.
2. f is symmetric in its first argument.

For fixed $\underline{\omega}$, let

$$\underline{x}[0,T] = \{\underline{x}(t) \colon t \in [0,T]\} = \left\{ (x_i(t))_{i=1}^N \colon t \in [0,T] \right\},$$

be N diffusions evolving according to the system of coupled Itô stochastic differential equations

$$dx_i(t) = -\frac{\partial H_N}{\partial x_i} \big(\underline{x}(t); \underline{\omega}\big)\, dt + db_i(t), \qquad i = 1, \dots, N;\ t \in [0,T], \tag{X.3}$$

where

$$\underline{b}[0,T] = \{\underline{b}(t) \colon t \in [0,T]\} = \left\{ (b_i(t))_{i=1}^N \colon t \in [0,T] \right\}$$

are N i.i.d. standard Brownian motions on $\mathbb{R}$, and $T > 0$ is a fixed but arbitrary time horizon. Thus, the components of the state variable feel a drift in the direction where the energy decreases and are subject to local white noise. Because f', g' are globally Lipschitz, (X.3) has a unique strong solution with continuous trajectories (see e.g. Karatzas and Shreve [C4] Theorem 2.9). As initial condition we pick

$$\underline{x}(0) \text{ has distribution } \lambda^N \tag{X.4}$$

for some probability law λ on $\mathbb{R}$.

For given $\underline{\omega}$, we write $\mathbb{P}_N^{\underline{\omega}}$ to denote the law of $\underline{x}[0,T]$, which is the *quenched* model. We write $\mathbb{W}_N = W^N$ to denote the law of $\underline{b}[0,T]$ with initial distribution λ^N. In other words, $\mathbb{W}_N$ is the law of our system in the absence of interaction.

The system defined by (X.3) and (X.4) will be our object of study. We will identify its large deviation behavior in the limit as $N \to \infty$ for fixed T. For given $\underline{\omega}$, the (infinite) Gibbs measure with density

$$\underline{x} \mapsto e^{-H_N(\underline{x};\underline{\omega})}$$

is its reversible equilibrium as $T \to \infty$ for fixed N (for the definition of Gibbs measure, see Georgii [C3] Chapter 2).

X.2 Radon-Nikodym formula

Define the *double layer empirical measure*

$$L_N\big[\underline{x}[0,T],\underline{\omega}\big](\,\cdot\,) = \frac{1}{N}\sum_{i=1}^{N} \delta_{(x_i[0,T],\omega_i)}(\,\cdot\,).$$

This is a random element of $\mathfrak{M}_1(C[0,T] \times \mathbb{R})$, equipped with the weak topology, where $C[0,T]$ is the space of continuous functions on $[0,T]$, playing the role of the path space. We begin by comparing $\mathbb{P}_N^{\underline{\omega}}$ with $\mathbb{W}_N$ in terms of a Radon-Nikodym formula involving L_N.

LEMMA X.5 *For given* $\underline{\omega}$,

$$\frac{d\mathbb{P}_N^{\underline{\omega}}}{d\mathbb{W}_N}\big(\underline{x}[0,T]\big) = e^{NF(L_N[\underline{x}[0,T],\underline{\omega}])}, \tag{X.6}$$

where, for $Q \in \mathfrak{M}_1(C[0,T] \times \mathbb{R})$,

$$F(Q) = \int Q(dy[0,T],d\pi)\Big[F_1^Q\big(y[0,T];\pi\big) + F_2^Q\big(y[0,T];\pi\big) + F_3^Q\big(y(0),y(T);\pi\big)\Big],$$

with

$$\begin{aligned}
&F_1^Q\big(y[0,T];\pi\big)\\
&\qquad = -\tfrac{1}{2}\textstyle\int_0^T dt\,\Big\{\int Q\big(d\tilde{y}[0,T],d\tilde{\pi}\big)f'\big(\tilde{y}(t)-y(t);\pi,\tilde{\pi}\big) + g'\big(y(t);\pi\big)\Big\}^2,\\
&F_2^Q\big(y[0,T];\pi\big)\\
&\qquad = \tfrac{1}{2}\textstyle\int_0^T dt\,\Big\{\int Q\big(d\tilde{y}[0,T],d\tilde{\pi}\big)f''\big(\tilde{y}(t)-y(t);\pi,\tilde{\pi}\big) + g''\big(y(t);\pi\big)\Big\},\\
&F_3^Q\big(y(0),y(T);\pi\big)\\
&\qquad = -\tfrac{1}{2}\textstyle\int Q\big(d\tilde{y}[0,T],d\tilde{\pi}\big)\,\Big[f\big(\tilde{y}(T)-y(T);\pi,\tilde{\pi}\big) - f\big(\tilde{y}(0)-y(0);\pi,\tilde{\pi}\big)\Big]\\
&\qquad\qquad -\Big[g\big(y(T);\pi\big) - g\big(y(0);\pi\big)\Big].
\end{aligned} \tag{X.7}$$

PROOF. The proof uses two basic tools in stochastic calculus, namely, Girsanov's formula and Itô's rule (see Karatzas and Shreve [C4] Theorems 3.3.3 and 3.5.1). Girsanov's formula applied to (X.3) yields

$$\frac{d\mathbb{P}_N^{\underline{\omega}}}{d\mathbb{W}_N}(\underline{x}[0,T])$$
$$= \exp\Big[-\tfrac{1}{2}\sum_{i=1}^{N}\int_0^T dt\,\Big(\tfrac{\partial H_N}{\partial x_i}(\underline{x}(t);\underline{\omega})\Big)^2 - \sum_{i=1}^{N}\int_0^T dx_i(t)\,\tfrac{\partial H_N}{\partial x_i}(\underline{x}(t);\underline{\omega})\Big].$$

Under the law $\mathbb{W}_N$, the process $\underline{x}[0,T]$ is N-dimensional Brownian motion. Therefore Itô's rule yields

$$\sum_{i=1}^{N}\int_0^T dx_i(t)\,\tfrac{\partial H_N}{\partial x_i}(\underline{x}(t);\underline{\omega})$$
$$= H_N(\underline{x}(T);\underline{\omega}) - H_N(\underline{x}(0);\underline{\omega}) - \tfrac{1}{2}\sum_{i=1}^{N}\int_0^T dt\,\tfrac{\partial^2 H_N}{\partial x_i^2}(\underline{x}(t);\underline{\omega}).$$

Combining the last two displays, inserting (X.1) and rewriting the resulting expression in terms of the empirical measure $L_N[\underline{x}[0,T],\underline{\omega}]$, we get the statement. □

The expression for $F(Q)$ in Lemma X.5 is a bit baroque. However, in Section X.4 we will write down a more compact formula. Note that $Q \mapsto F(Q)$ is quadratic.

X.3 The LDP for the double layer empirical measure

Lemma X.5 allows us to deduce the LDP for $(L_N)_{N\in\mathbb{N}}$ in a way that is similar to the argument used in Sections IV.1 and IV.2. Define

$$\mathbb{P}_N(\cdot) = \int \mu^N(d\underline{\omega})\,\mathbb{P}_N^{\underline{\omega}}(L_N \in \cdot\,), \tag{X.8}$$

which is the law of L_N under the joint distribution of process and medium, i.e., the *annealed* model. Note that $\mathbb{P}_N$ is a probability law on $\mathfrak{M}_1(C[0,T]\times\mathbb{R})$.

THEOREM X.9 $(\mathbb{P}_N)_{N\in\mathbb{N}}$ *satisfies the* LDP *on* $\mathfrak{M}_1(C[0,T]\times\mathbb{R})$ *with rate* N *and with rate function*

$$I(Q) = H(Q|W\times\mu) - F(Q), \tag{X.10}$$

where H *denotes the relative entropy*

$$H(Q|W\times\mu) = \begin{cases} \int dQ\,\log\frac{dQ}{d(W\times\mu)} & \text{if } Q \ll W\times\mu, \\ \infty & \text{otherwise.}\end{cases}$$

PROOF. Let $\mathbb{S}_N$ be the law of L_N under the probability measure $(W\times\mu)^N$, i.e., the annealed law in the absence of interaction. Under $\mathbb{S}_N$, the $(x_i[0,T],\omega_i)$'s are i.i.d. random variables on $C[0,T]\times\mathbb{R}$. Hence, it follows from Sanov's Theorem that $(\mathbb{S}_N)_{N\in\mathbb{N}}$ satisfies the LDP with rate N and with rate function $H(Q|W\times\mu)$. (Here we use the version of Sanov's Theorem that was mentioned at the end of Section II.7.) Now, using Lemma X.5, we may write

$$\begin{aligned}\mathbb{P}_N(\cdot) &= \int W^N(d\underline{x}[0,T])\int\mu^N(d\underline{\omega})\,e^{NF(L_N[\underline{x}[0,T],\underline{\omega}])}\,1_{\{L_N[\underline{x}[0,T],\underline{\omega}]\in\,\cdot\,\}}\\ &= \int\mathbb{S}_N(dQ)\,e^{NF(Q)}\,1_{\{Q\in\,\cdot\,\}}.\end{aligned}$$

This identity says that

$$\frac{d\mathbb{P}_N}{d\mathbb{S}_N}(Q) = e^{NF(Q)}.$$

Our assumptions on f, g in Section X.1 imply that F in Lemma X.5 is bounded and continuous with respect to the weak topology in $\mathfrak{M}_1(C[0,T] \times \mathbb{R})$, so we can use Theorem III.17 to transfer the LDP for $(\mathbb{S}_N)_{n\in\mathbb{N}}$ to the LDP for $(\mathbb{P}_N)_{N\in\mathbb{N}}$. This leads us to the rate function in (X.10). □

X.4 A simpler representation for the rate function

To analyze the rate function in Theorem X.9, we first give an alternative representation for $F(Q)$ that turns out to be more convenient. For $\omega \in \mathbb{R}$ and $q \in \mathfrak{M}_1(\mathbb{R} \times \mathbb{R})$, define

$$\beta^{\omega,q}(x) = -\int q(dy, d\pi) f'(y - x; \omega, \pi) - g'(x; \omega), \qquad x \in \mathbb{R}. \tag{X.11}$$

For $\omega \in \mathbb{R}$ and $Q \in \mathfrak{M}_1(C[0,T] \times \mathbb{R})$, let $P^{\omega,Q}$ denote the law of the unique strong solution of the 1-dimensional Itô stochastic differential equation

$$\begin{aligned} &dx(t) = \beta^{\omega,\pi_t Q}\big(x(t)\big)\, dt + db(t), \\ &x(0) \text{ has distribution } \lambda, \end{aligned} \tag{X.12}$$

where $b(t)$ is a standard Brownian motion on $\mathbb{R}$, and $\pi_t Q$ is the evaluation of Q at time t, i.e.,

$$(\pi_t Q)(A \times B) = Q\Big(\Big\{(x[0,T], \omega) : x(t) \in A,\, \omega \in B\Big\}\Big), \qquad A, B \subset \mathbb{R}.$$

With this notation we can now write the following compact formula for the function F in Lemma X.5:

LEMMA X.13 $F(Q) = \int Q(dx[0,T], d\omega) \log \frac{dP^{\omega,Q}}{dW}\big(x[0,T]\big)$.

PROOF. By applying Girsanov's formula to (X.11), we get

$$\log \frac{dP^{\omega,Q}}{dW}\big(x[0,T]\big) = -\frac{1}{2}\int_0^T dt\, \Big(\beta^{\omega,\pi_t Q}\big(x(t)\big)\Big)^2 + \int_0^T dx(t)\, \beta^{\omega,\pi_t Q}\big(x(t)\big). \tag{X.14}$$

We want to show that the RHS yields $F(Q)$ when integrated over $Q\big(dx[0,T], d\omega\big)$. This goes as follows.

From (X.11) we see that the first term in the RHS of (X.14) gives rise to the first term in the RHS of (X.7). Consider the second term in the RHS of (X.14). By (X.11), we have

$$\begin{aligned} &\textstyle\int Q\big(dx[0,T], d\omega\big) \int_0^T dx(t)\, \beta^{\omega,\pi_t Q}\big(x(t)\big) \\ &\quad = -\textstyle\int Q\big(dx[0,T], d\omega\big) \int_0^T dx(t) \\ &\qquad \times \Big[\textstyle\int Q\big(dy[0,T], d\pi\big) f'\big(y(t) - x(t); \omega, \pi\big) + g'\big(x(t); \omega\big)\Big]. \end{aligned} \tag{X.15}$$

(The stochastic integral is well defined because $Q \ll W \times \mu$ implies that $x[0,T]$ is a Q-semimartingale.) Take the first term in the RHS of (X.15). Since f' is an odd function of its first argument, this term equals

$$-\frac{1}{2}\int Q\big(dx[0,T],d\omega\big)\int Q\big(dy[0,T],d\pi\big)\int_0^T [dx(t)-dy(t)]\; f'\big(y(t)-x(t);\omega,\pi\big).$$

We can apply Itô's rule to the two-dimensional semimartingale $(x,y)[0,T]$ and write

$$\begin{aligned} &df\big(y(t)-x(t);\omega,\pi\big)\\ &\qquad= f''\big(y(t)-x(t);\omega,\pi\big)dt - f'\big(y(t)-x(t);\omega,\pi\big)[dx(t)-dy(t)]. \end{aligned}$$

Substitution into the previous display gives

$$\begin{aligned} &-\tfrac{1}{2}\textstyle\int Q\big(dx[0,T],d\omega\big)\int Q\big(dy[0,T],d\pi\big)\\ &\times\Big[\textstyle\int_0^T dt\; f''\big(y(t)-x(t);\omega;\pi\big) - f\big(y(T)-x(T);\omega,\pi\big) + f\big(y(0)-x(0);\omega,\pi\big)\Big]. \end{aligned}$$

Next take the second term in the RHS of (X.15). Itô's rule yields that this term equals

$$-\int Q\big(dx[0,T],d\omega\big)\Big[-\frac{1}{2}\int_0^T dt\; g''\big(x(t);\omega\big) + g\big(x(T);\omega\big) - g\big(x(0);\omega\big)\Big].$$

From the last two displays we get the second and the third term in the RHS of (X.7). □

The above reformulation leads us to the following simpler representation for the rate function:

THEOREM X.16 *For all* $Q\in\mathfrak{M}_1(C[0,T]\times\mathbb{R})$,

$$I(Q)=H(Q|P^Q),$$

where $P^Q\in\mathfrak{M}_1(C[0,T]\times\mathbb{R})$ *is defined by*

$$P^Q\big(dx[0,T],d\omega\big)=\mu(d\omega)P^{\omega,Q}\big(dx[0,T]\big).$$

PROOF. Combine Theorem X.9 and Lemma X.13. □

With Theorems X.9 and X.16 we have obtained the LDP for $(L_N)_{N\in\mathbb{N}}$ with a rate function $I(Q)$ that is the relative entropy of Q w.r.t. a measure P^Q, which depends on Q itself and which is given by the 1-dimensional Itô stochastic differential equation in (X.12).

X.5 McKean-Vlasov equation

As a corollary of Theorems X.9 and X.16 we obtain that, as $N\to\infty$, the probability law $\mathbb{P}_N$ in (X.8) tends to concentrate around the zeroes of the rate function, i.e., the solutions of the equation

$$Q=P^Q,\qquad Q\in\mathfrak{M}_1(C[0,T]\times\mathbb{R}). \tag{X.17}$$

Our next theorem states that (X.17) has a unique solution that is described by a certain partial differential equation, called the McKean-Vlasov equation.

Let $\nu^Q\in\mathfrak{M}_1(\mathbb{R})$ denote the projection of Q on the medium coordinate, i.e.,

$$\nu^Q(B)=Q\Big(\big\{(x[0,T],\omega):\omega\in B\big\}\Big),\qquad B\subset\mathbb{R}.$$

Let $Q^\omega \in \mathfrak{M}_1(C[0,T])$ denote the regular conditional probability measure obtained from Q after conditioning on ω, i.e.,

$$Q(dx[0,T],\omega) = \nu^Q(d\omega)Q^\omega(dx[0,T]).$$

We will make the following assumptions on the law λ for the initial condition in (X.4):

$$\begin{array}{l}(1)\ \lambda \text{ has a density with respect to Lebesgue measure.}\\ (2)\ \lambda \text{ has a finite } p\text{-th moment for some } p>1.\end{array} \tag{X.18}$$

THEOREM X.19 *Assume* (X.18). *Then* (X.17) *has a unique solution* Q_* *with the following properties:*
(a) $\nu^{Q_*} = \mu$.
(b) For μ*-a.s. all* ω, Q_*^ω *is the law of a (time-inhomogeneous) Markov diffusion process on* $\mathbb{R}$.
(c) Let $q_t^\omega = \pi_t Q_*^\omega$. *Then* q_t^ω *is the weak solution of the McKean-Vlasov equation*

$$\begin{array}{rcll}\frac{\partial}{\partial t} q_t^\omega &=& \mathcal{L}^\omega q_t^\omega, & \\ q_0^\omega &=& \lambda, & t\in[0,T],\ \omega\in\mathbb{R},\end{array} \tag{X.20}$$

where $\mathcal{L}^\omega$ *is the diffusion operator*

$$\mathcal{L}^\omega q_t^\omega = -\frac{\partial}{\partial x}[\beta^{\omega,q_t} q_t^\omega] + \frac{1}{2}\frac{\partial^2}{\partial x^2} q_t^\omega, \qquad \omega \in \mathbb{R}, \tag{X.21}$$

with $q_t \in \mathfrak{M}_1(\mathbb{R}\times\mathbb{R})$ *defined by* $q_t(dx,d\omega) = \mu(d\omega)q_t^\omega(dx)$.
(d) The diffusion process in (b) has a (time-dependent) generator L_t^ω *given by*

$$L_t^\omega = \beta^{\omega,q_t}\frac{\partial}{\partial x} + \frac{1}{2}\frac{\partial^2}{\partial x^2}. \tag{X.22}$$

PROOF. Note that $\nu^Q = \nu^{P^Q} = \mu$ (recall (X.2)) and that $P^{\omega,Q}$ is the law of the solution of (X.12), i.e., the Markov diffusion process on $\mathbb{R}$ with generator as in (X.21). It is therefore easy to see that (a–d) are satisfied for any solution of (X.17). The existence of a solution of (X.17) is guaranteed by the fact that a rate function has at least one zero (recall Comment (6) in Section III.2). The uniqueness of the solution comes from certain analytical estimates based on the assumption in (X.18). See e.g. Sznitman [D58] and Gärtner [D25] for proofs of uniqueness in a much more general context. □

Equations (X.20), (X.21) and (X.22) are to be read as the statement that

$$\frac{d}{dt}\int q_t^\omega(dx)\phi(x) = \int q_t^\omega(dx)\beta^{\omega,q_t}(x)\phi'(x) + \frac{1}{2}\int q_t^\omega(dx)\phi''(x)$$

for every $\phi \in C_c^\infty(\mathbb{R})$, the infinitely differentiable functions with compact support. By standard arguments the latter implies that q_t^ω for all $t>0$ has a density that is a classical solution.

Equation (X.20) is a *coupled* family of equations labelled by ω. The coupling comes from the fact that $\beta^{\omega,q_t}(x)$ depends not only on q_t^ω but on the whole family $(q_t^\pi)_{\pi\in\mathbb{R}}$ (see (X.11)). Thus, it is a non-linear equation, typical for mean-field models. We should think of q_t as:

> *q_t is the joint probability law of a typical state component and medium component at time t.*

The McKean-Vlasov equation tells us how this law evolves with time. Finding the solution of this equation is in general a highly non-trival problem. An example will be given in Section X.6.

As a corollary of Theorems X.9 and X.19 we get the law of large numbers for the annealed model:

$$\mathbb{P}_N \underset{N\to\infty}{\longrightarrow} \delta_{Q_*} \qquad \text{weakly.}$$

REMARK X.23 With the help of the Contraction Principle it is possible to derive the LDP for the *double layer empirical flow*

$$\ell_N = \Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{(x_i(t),\omega_i)}\Big)_{t\in[0,T]},$$

which is a random variable taking values in $\mathfrak{M}_1(\mathbb{R}\times\mathbb{R})$. However, the rate function for ℓ_N is a bit more involved than that for L_N. In fact, a good part of the analysis in the references cited at the beginning of this chapter revolves around this rate function. For more details, see Dai Pra and den Hollander [D13].

X.6 The Kuramoto model

We close this chapter by describing the *stationary* solutions of (X.20) and (X.21) for an interesting example, called the Kuramoto model. This model corresponds to the choice

$$\begin{aligned} f(x;\omega,\pi) &= -K\cos x, \\ g(x;\omega) &= -x\omega, \qquad x\in[0,2\pi);\ \omega,\pi\in\mathbb{R}, \end{aligned}$$

where $K\in(0,\infty)$ is the *coupling strength*. With this choice, our system in (X.3) becomes a system of mean-field nonlinearly coupled oscillators, each with its own random frequency and white noise. The two terms in the Hamiltonian have competing effects: f tends to point the oscillators in the same direction, g tends to make each oscillator rotate at its own local frequency. We will see that this system has a phase transition as a function of the parameter K. Note that the components of the state variable take values in $[0,2\pi)$, rather than in $\mathbb{R}$ as in Sections X.1–X.5, i.e., $\mathbb{R}$ is wrapped around the unit circle.

Let $q_t^\omega(x)$ denote the probability density at time t that a typical oscillator has angle x and rotation frequency ω (in the McKean-Vlasov limit), normalized as

$$\int_0^{2\pi} dx\ q_t^\omega(x) = 1 \qquad \forall t\in[0,T],\ \omega\in\mathbb{R} \tag{X.24}$$

and respecting the periodic boundary condition

$$q_t^\omega(0) = q_t^\omega(2\pi) \qquad \forall t\in[0,T],\ \omega\in\mathbb{R}. \tag{X.25}$$

The appropriate "order parameter" of the system is the complex number

$$r_t e^{i\psi_t} = \int_{\mathbb{R}}\mu(d\omega)\int_0^{2\pi} dx\ e^{ix} q_t^\omega(x). \tag{X.26}$$

Here, $r_t\geq 0$ measures the *phase coherence* of the oscillators and ψ_t measures the *average phase*. In terms of these quantities, the McKean-Vlasov equation reads

$$\frac{\partial}{\partial t}q_t^\omega = -\frac{\partial}{\partial x}[\beta^{\omega,q_t}q_t^\omega] + \frac{1}{2}\frac{\partial^2}{\partial x^2}q_t^\omega, \tag{X.27}$$

with the drift term given by (X.11):

$$\beta^{\omega,q_t}(x) = Kr_t \sin(\psi_t - x) + \omega. \tag{X.28}$$

The stationary solutions of (X.27) and (X.28) (i.e., the solutions where the LHS of (X.27) is zero) have been studied by Sakaguchi [D51]. To summarize his results, we restrict ourselves to the situation where μ is *symmetric*. In that case the average phase can be taken zero without loss of generality.

Any stationary solution of (X.27) and (X.28) respecting the periodic boundary condition is of the form

$$q^\omega(x) = \frac{1}{Z^{\omega,r}} A^{\omega,r}(x), \tag{X.29}$$

with

$$\begin{aligned} A^{\omega,r}(x) &= B^{\omega,r}(x)\Big\{e^{4\pi\omega} \int_0^{2\pi} \tfrac{dy}{B^{\omega,r}(y)} + \big[1 - e^{4\pi\omega}\big] \int_0^x \tfrac{dy}{B^{\omega,r}(y)}\Big\}, \\ B^{\omega,r}(x) &= \exp[2Kr\cos x + 2\omega x], \end{aligned} \tag{X.30}$$

and with $Z^{\omega,r}$ the normalizing constant. The variable r must satisfy the *consistency relation*

$$r = \Phi_\mu(r), \tag{X.31}$$

where

$$\Phi_\mu(r) = \int_{\mathbb{R}} \mu(d\omega)\, \frac{\int_0^{2\pi} dx\, (\cos x)\, A^{\omega,r}(x)}{\int_0^{2\pi} dx\, A^{\omega,r}(x)}. \tag{X.32}$$

This relation is the stationary version of (X.26).

EXERCISE X.33 *Derive* (X.29) *and* (X.30).

Solutions with $r = 0$ are called *incoherent*, those with $r > 0$ are called *synchronized*. It is obvious from the above formulas that the only incoherent solution is

$$q^\omega(x) = \frac{1}{2\pi} \qquad \forall x \in [0, 2\pi),\ \omega \in \mathbb{R},$$

and that this solution exists for all choices of μ and K. The next theorem shows that if K exceeds a certain μ-dependent threshold, then a synchronized solution is possible. To state this theorem we make the additional assumption that μ is *unimodal*, i.e, has a density that is monotone left and right of its symmetry point $\omega = 0$.

THEOREM X.34 *Suppose that* μ *is unimodal. Define*

$$K_c = \Big[\int_{\mathbb{R}} \frac{\mu(d\omega)}{1 + 4\omega^2}\Big]^{-1} \in (1, \infty).$$

Then:
(a) For K *small enough there is no synchronized solution.*
(b) For $K > K_c$ *there is a synchronized solution.*

PROOF. To prove the claim, we need the following:

EXERCISE X.35 *Show that: (i)* $\Phi_\mu(0) = 0$*; (ii)* $r \mapsto \Phi_\mu(r)$ *is continuous; (iii)* $\lim_{r\to\infty} \Phi_\mu(r) = 1$.

EXERCISE X.36 *Show that: (i)* $\Phi'_\mu(0) = K/K_c$*; (ii)* $\Phi''_\mu(0) = 0$*; (iii)* $\Phi'''_\mu(0) < 0$ *when* μ *is unimodal.*

See Fig. 22. As K increases through K_c, the slope of $r \mapsto \Phi_\mu(r)$ at the origin increases through 1, so that (X.31) develops a solution for some $r > 0$. Thus, the critical value K_c is a bifurcation point. Note that, since we do not know whether $r \mapsto \Phi_\mu(r)$ is strictly concave everywhere, we do not know whether there is a synchronized solution below K_c or whether the synchronized solution above K_c is unique. □

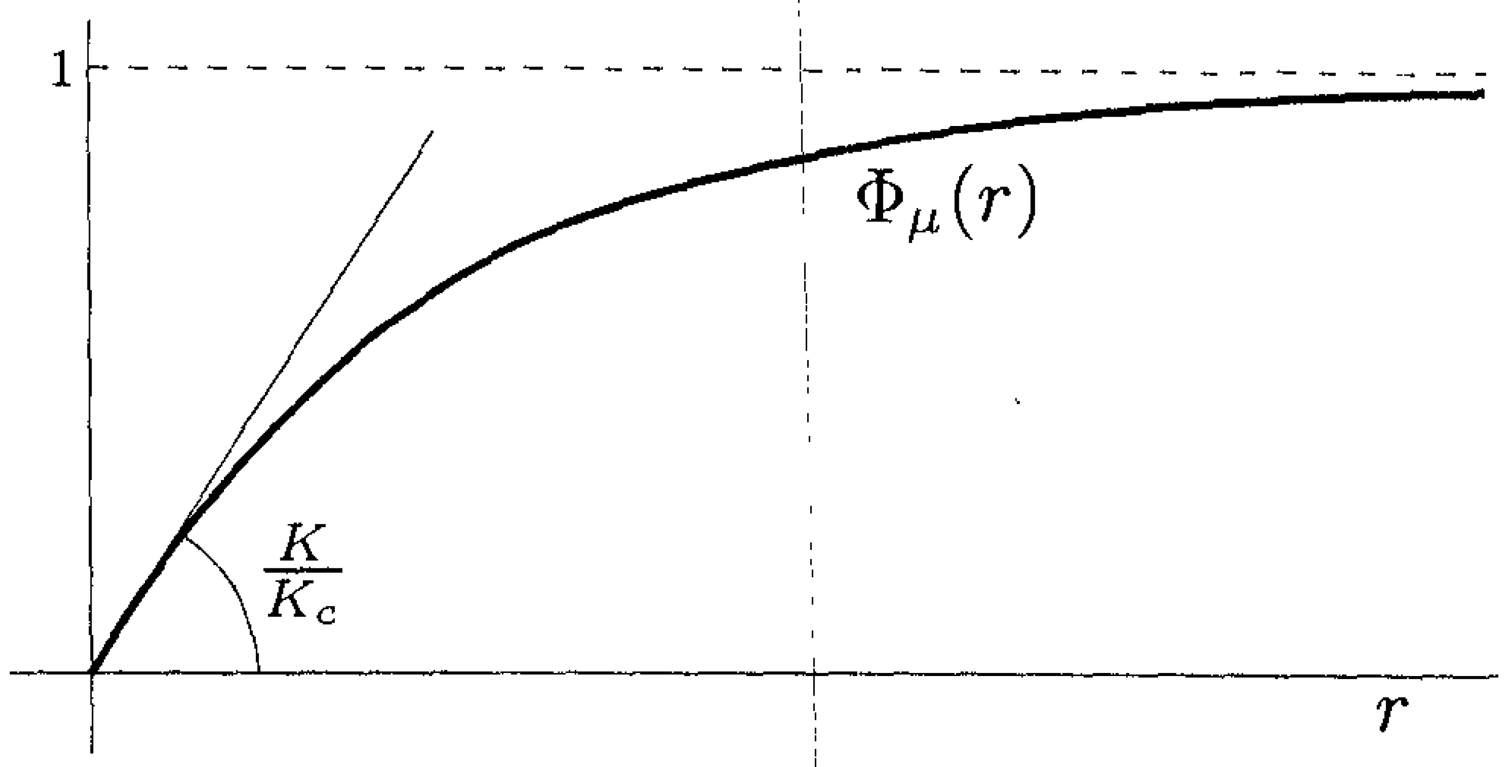

FIG. 22. Qualitative picture of $r \mapsto \Phi_\mu(r)$

Theorem X.34 shows that the system has a phase transition. Apparently, for small coupling strength most oscillators move about in an erratic manner, sticking to their local frequencies and not bothering too much about their interaction. Since the frequency distribution is symmetric, this behavior results in $r = 0$. However, for large coupling strength a positive fraction of the oscillators spontaneously locks to a common frequency, despite their different local frequencies, resulting in $r > 0$. As K increases, the fraction of oscillators involved in this locking phenomenon increases, and tends to one as K tends to infinity. The value $r(K)$ of the phase coherence as a function of K is plotted in Fig. 23. For large but finite K, a positive fraction of the oscillators moves about erratically, typically those oscillators whose local frequency is far away from the center of μ (which is $\omega = 0$).

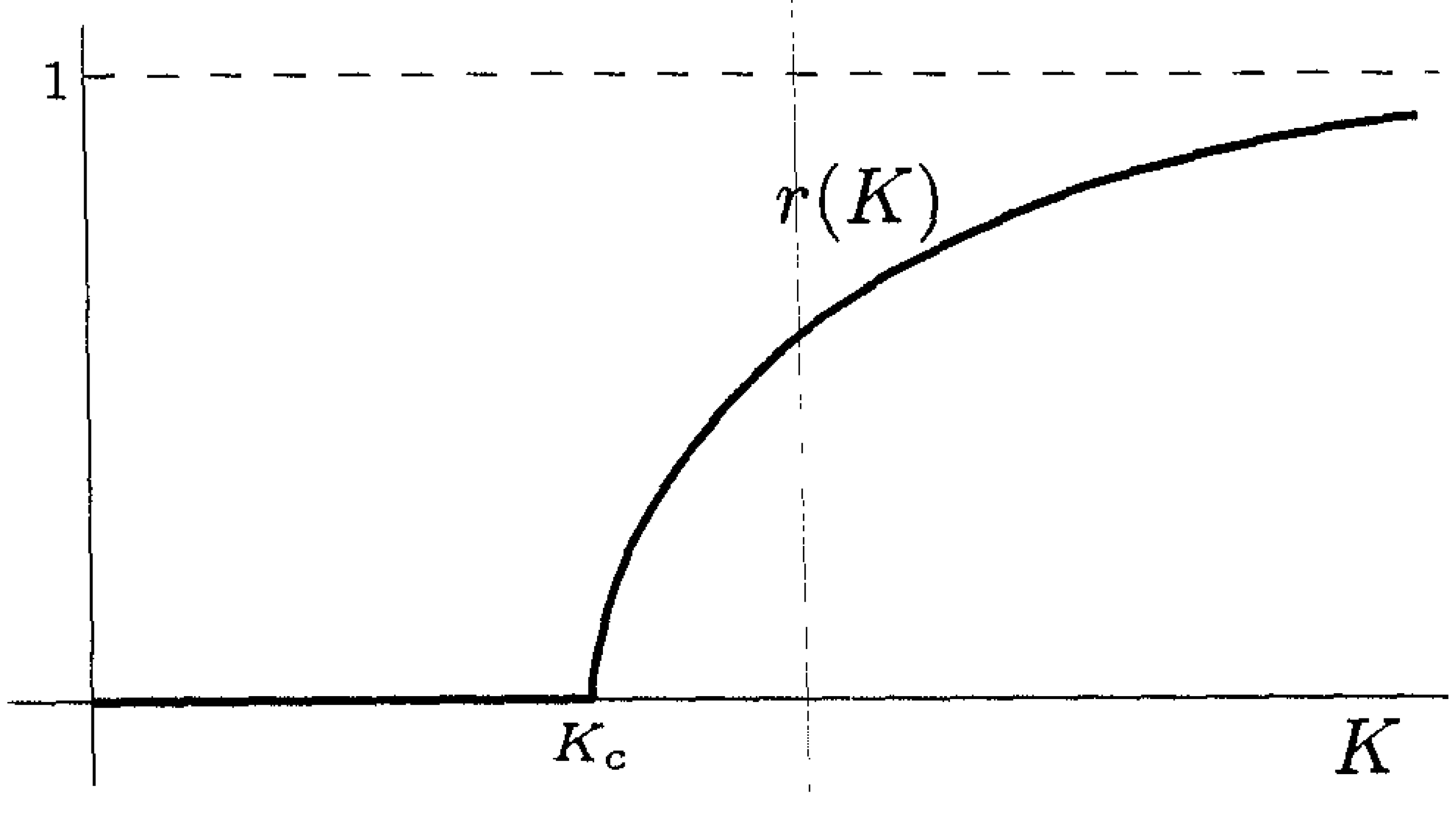

FIG. 23. Qualitative picture of $K \mapsto r(K)$

EXERCISE X.37 *Show that* $r(K) \sim C\sqrt{K - K_c}$, $K \downarrow K_c$, *for some* $C > 0$ *when* μ *is unimodal.*

X.7 Concluding remarks

(1) In order to be sure that there is no synchronized solution for $K \leq K_c$, we would need to prove that $r \mapsto \Phi_\mu(r)$ is strictly concave. This property is known to fail when μ has two bumps that are sufficiently separated. It seems to be open to show that strict concavity holds as soon as μ is unimodal. Figs. 22 and 23 therefore represent an optimistic scenario.

(2) The *stability* of the stationary solutions of (X.27) and (X.28) has been studied by Strogatz and Mirollo [D56] and Strogatz, Mirollo and Matthews [D57]. In the unimodal case, the incoherent solution is linearly stable for $K < K_c$ and linearly unstable for $K > K_c$. For the synchronized solution, however, stability has not been settled in full generality. For the bimodal case the situation is considerably more complex. For further background on the Kuramoto model and its history, the reader is referred to the review article by Strogatz [D55].

(3) For finite N, the reversible equilibrium of the Kuramoto model is the (finite) Gibbs measure

$$\rho_N^{\underline{\omega}}(d\underline{x}) = \frac{1}{Z_N^{\underline{\omega}}} e^{-H_N(\underline{x};\underline{\omega})}\, d\underline{x},$$

where $Z_N^{\underline{\omega}}$ is the normalizing constant. It can be shown that the family $(\rho_N^{\underline{\omega}} \circ \pi_1^{-1})_{N\in\mathbb{N}}$, with π_1 denoting the projection on the first state coordinate, satisfies an LDP on $\mathfrak{M}_1([0,2\pi) \times \mathbb{R})$ with rate N and with an $\underline{\omega}$-a.s. constant rate function whose critical points are the stationary solutions of the McKean-Vlasov equation.

(4) The mean-field model studied in this chapter is relatively easy because standard large deviation techniques apply. In the literature one finds interesting variations that have a much more complicated structure. For instance, the "Sherrington-Kirkpatrick spin glass model with noise" is the system where the Hamiltonian takes the form

$$H_N(\underline{x},\underline{\omega}) = K\frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} \omega_{ij} x_i x_j,$$

where the ω_{ij}'s are i.i.d. standard normal random variables, and $K > 0$ is the coupling strength. The LDP for the empirical measure on path space has been derived by Grunwald [D34], Ben Arous and Guionnet [D5], [D6] and Guionnet [D35]. The rate function $Q \mapsto I(Q)$ is very complex. It again has a unique zero Q_* but, contrary to the situation in Theorem X.19, the law Q_*^{ω} is not Markovian. For small K a good understanding of the McKean-Vlasov equation is available, but for large K much remains shrouded in mystery. With the additional restriction that $\omega_{ij} = \omega_{ji}$ and that the sum in the Hamiltonian runs over $1 \leq i < j \leq N$, the model becomes slightly easier, but it still remains ill understood.

Appendix: Solutions to the Exercises

Chapter I:

Exercise I.7 Since $\mathbb{P}(X_1 = a) = 1$ gives $\log\varphi(t) = at$ in (I.5), we get from (I.6) that $I(z) = \sup_{t\in\mathbb{R}}[(z-a)t]$. Hence $I(z) = \infty$ for all $z \neq a$, and $I(a) = 0$. So the statement in Theorem I.4 is true for this case.

Exercise I.11 Use (I.5) and (I.6).
1. Poisson(λ): $\varphi(t) = \exp[\lambda(e^t - 1)]$, $t \in \mathbb{R}$; $I(z) = z\log(\frac{z}{\lambda}) - z + \lambda$, $z \geq 0$, and $I(z) = \infty$, $z < 0$.
2. Exponential(α): $\varphi(t) = \frac{\alpha}{\alpha - t}$, $t < \alpha$, and $\varphi(t) = \infty$, $t \geq \alpha$; $I(z) = \alpha z - \log(\alpha z) - 1$, $z > 0$, and $I(z) = \infty$, $z \leq 0$.
3. Normal(μ, σ^2): $\varphi(t) = \exp[\mu t + \frac{1}{2}\sigma^2 t^2]$, $t \in \mathbb{R}$; $I(z) = \frac{(z-\mu)^2}{2\sigma^2}$, $z \in \mathbb{R}$.

Exercise I.12 For the setting in Theorem I.3 we have $\varphi(t) = \frac{1}{2}(1 + e^t)$, $t \in \mathbb{R}$. Let $z \in [0,1]$. Then $t \mapsto zt - \log\varphi(t)$ is bounded from above, continuous, and strictly concave. Hence its supremum is attained at $t = \tau$ given by the equation $z = \varphi'(\tau)/\varphi(\tau)$, or $z^{-1} = e^{-\tau} + 1$. After elimination of τ in favor of z, we obtain the formula in Theorem I.3. If $z \notin [0,1]$, then $t \mapsto zt - \log\varphi(t)$ is unbounded from above, which implies $I(z) = \infty$, as desired.

Exercise I.13 The shift $z \to z - 2$ places us in the case where $X_1 = 0, \pm 1$ with probability $\frac{1}{3}$ each. In this case $\varphi(t) = \frac{1}{3}(1 + 2\cosh t)$, $t \in \mathbb{R}$, and, for $z \in [-1,1]$, the supremum of $t \mapsto zt - \log\varphi(t)$ is attained at $t = \tau$ given by

$$\tanh(\tfrac{\tau}{2}) = \tfrac{2}{z}\left(1 - \sqrt{1 - \tfrac{3}{4}z^2}\,\right).$$

Substitution gives $I(z-2) = z\tau - \log\varphi(\tau)$.

Exercise I.16 Again use (I.5) and (I.6).
(i) Since $z \mapsto zt - \log\varphi(t)$ is continuous and since the supremum of continuous functions is lower semi-continuous, $z \mapsto I(z)$ is lower semi-continuous. Since $z \mapsto zt - \log\varphi(t)$ is linear and since the supremum of linear functions is convex, $z \mapsto I(z)$ is convex.
(ii) Since $\{z \in \mathbb{R}\colon I(z) \leq c\}$ is closed for all $c \in [0,\infty)$ (by the lower semi-continuity of I), it suffices to show that this set is bounded. Suppose the contrary. Then there exists a sequence (z_n) with $I(z_n) \leq c$ for all n and $z_n \to \infty$ as $n \to \infty$ (the case $-\infty$ is analogous). Hence $z_n t - \log\varphi(t) \leq c$ for all $t \in \mathbb{R}$ and all n. Pick $t = 2c/z_n$. Then $2c - \log\varphi(2c/z_n) \leq c$ for all n. However, this is impossible because $\varphi(2c/z_n) \to \varphi(0) = 1$ as $n \to \infty$.
(iii–iv) Note that $\mathcal{D}_I \supset \text{Range}([\log\varphi]')$ (see Fig. 3). Since φ is smooth, we have $I(z) = z\tau(z) - \log\varphi(\tau(z))$, $z \in \text{int}(\mathcal{D}_I)$, with $\tau(z)$ the unique solution of $z =$

$[\log\varphi]'(\tau(z))$. By the implicit function theorem, τ is smooth on $\mathrm{int}(\mathcal{D}_I)$ and hence so is I. Differentiating twice with respect to z, we get

$$I'(z) = \tau(z), \quad I''(z) = \tau'(z) = \frac{1}{[\log\varphi]''(\tau(z))},$$

showing that I is continuous and strictly convex on $\mathrm{int}(\mathcal{D}_I)$.
(v) $I(z) \geq -\log\varphi(0) = 0$ for all $z \in \mathbb{R}$. By Jensen's inequality, $\log\varphi(t) \geq t\mu$ for all $t \in \mathbb{R}$. Hence $I(\mu) = 0$. Part (iii) implies that μ is the only zero.
(vi) Since $\varphi(0) = 1$, $\varphi'(0) = \mu$ and $\varphi''(0) = \mu^2 + \sigma^2$, it follows from the last display (with $\tau(\mu) = 0$) that $I(\mu) = 0$, $I'(\mu) = 0$, $I''(\mu) = \frac{1}{\sigma^2}$.

Exercise I.19 If X_1 has probability density $f(x) = \frac{1}{N_\alpha}x^{-\alpha}e^{-x}1_{[1,\infty)}(x)$, $x \in \mathbb{R}, \alpha > 0$, then $\log\varphi$ is steep for $0 < \alpha \leq 2$ and not steep for $\alpha > 2$.

Exercise I.20 $\mathcal{D}_I \supset \mathrm{Range}([\log\varphi]')$, as noted in the solution of Exercise I.16.

Exercise I.21 Hölder: $\varphi(t) \leq \varphi(\frac{t}{\gamma})^\gamma \varphi(\frac{t}{1-\gamma})^{1-\gamma}$, $t \in \mathbb{R}, 0 < \gamma < 1$.
Fatou: $\liminf_{s\to t}\varphi(s) \geq \varphi(t)$, $t \in \mathbb{R}$.

Exercise I.24 Recall that $\mu = \mathbb{E}X_1$.
(i) If $A = [a,\infty)$ with $a \leq \mu$, then $\mathbb{P}(\frac{1}{n}S_n \in A)$ is bounded away from zero by the CLT (assume (I.17)). Hence the LHS of (I.23) is zero. But the RHS is zero too, because of Lemma I.14(v).
(ii) If $a \in \mathrm{int}(\mathcal{D}_I)$, then there exists a $\delta > 0$ such that $a + \delta \in \mathrm{int}(\mathcal{D}_I)$. Put $A = (a,\infty)$, $\bar{A} = [a,\infty)$, and $A_\delta = [a+\delta,\infty)$. With the help of (i), we get

$$-\inf_{z\in A_\delta} I(z) \leq \liminf_{n\to\infty}\tfrac{1}{n}\log\mathbb{P}(\tfrac{1}{n}S_n \in A) \leq \limsup_{n\to\infty}\tfrac{1}{n}\log\mathbb{P}(\tfrac{1}{n}S_n \in A) = -\inf_{z\in\bar{A}} I(z).$$

Let $\delta \downarrow 0$ and use Lemma I.14(iii), to get (I.23). If $a \notin (\mathcal{D}_I \cup \partial\mathcal{D}_I)$, then a similar argument shows that both sides of (I.23) are infinite.
(iii) Use the same argument as in (ii).

Exercise I.25 Note that $J(a) = I(a-\frac{1}{2})$ with I as in Theorem I.3. The variational problem reads $b = \sup_{a>0} K(a)$ with $K(a) = \log a - J(a)$. Compute

$$K'(a) = \frac{1}{a} - \log\frac{a-\frac{1}{2}}{\frac{3}{2}-a}, \qquad a \in (\tfrac{1}{2}, \tfrac{3}{2}).$$

In particular, $K'(1) = 1 > 0$, so $a \mapsto K(a)$ is strictly increasing in a neighborhood of $a = 1$. Consequently, $a^* \neq 1$. Since $K(1) = 0$, also $b = K(a^*) > 0$. Numerically $a^* \approx 1.2$, $b \approx 0.1$.

Chapter II:

Exercise II.5 Let $h(x) = x\log x$.
(i) Write (II.3) as $I_\rho(\nu) = \sum_s \rho_s h(\frac{\nu_s}{\rho_s})$. Then use that $x \mapsto h(x)$ is finite, continuous and strictly convex on $[0,1]$.
(ii) By Jensen's inequality,

$$I_\rho(\nu) = \sum_s \rho_s h\Big(\frac{\nu_s}{\rho_s}\Big) \geq h\Big(\sum_s \rho_s\frac{\nu_s}{\rho_s}\Big) = h(1) = 0.$$

Equality occurs if and only if $\nu_s/\rho_s = C$ for all s. But $C = 1$ because $\sum_s \rho_s = \sum_s \nu_s = 1$.

Exercise II.6 If ρ is uniform on Γ with $|\Gamma| = r$ and ν is uniform on $\Gamma' \subset \Gamma$ with $|\Gamma'| = r'$, then $I_\rho(\nu) = \log(r/r')$.

Exercise II.7 Write $L_n^2 = \frac{1}{2}L_n^{\text{even}} + \frac{1}{2}L_n^{\text{odd}}$, where

$$L_n^{\text{even}} = \lfloor \tfrac{n}{2} \rfloor^{-1} \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \delta_{(X_{2i}, X_{2i+1})}, \qquad L_n^{\text{odd}} = \lfloor \tfrac{n}{2} \rfloor^{-1} \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \delta_{(X_{2i-1}, X_{2i})}.$$

In both sums only independent terms enter. Consequently, by the SLLN, both $d(L_n^{\text{even}}, \rho \times \rho) \to 0$ and $d(L_n^{\text{odd}}, \rho \times \rho) \to 0$ as $n \to \infty$. Hence, by the triangle inequality, also $d(L_n^2, \rho \times \rho) \to 0$.

Exercise II.12 Let $h(x) = x \log x$.
(i) Same as in Exercise II.5(i). Write (II.9) as $I_\rho^2(\nu) = \sum_{s,t} \bar{\nu}_s \rho_t h(\frac{\nu_{st}}{\bar{\nu}_s \rho_t})$. From this we see that $\nu \mapsto I_\rho^2(\nu)$ is strictly convex, except along line segments where $\nu_{st}/\bar{\nu}_s$ is constant.
(ii) Same as in Exercise II.5(ii). By Jensen's inequality,

$$I_\rho^2(\nu) \geq h\Big(\sum_{s,t} \bar{\nu}_s \rho_t \frac{\nu_{st}}{\bar{\nu}_s \rho_t} \Big) = h(1) = 0.$$

Equality occurs if and only if $\nu_{st}/\bar{\nu}_s \rho_t = C$ for all s, t. Since ν has equal marginals, it follows that $\nu_{st} = \rho_s \rho_t$. The origin of the affine part of I_ρ^2 can be understood as follows. Let $\Gamma = \{1, 2\}$ and $\rho = (\frac{1}{2}, \frac{1}{2})$, i.e., $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 2) = \frac{1}{2}$. Pick $\alpha \in [0, 1]$ such that αn is integer, and consider the cylinder where $X_i = 1$ for $1 \leq i \leq \alpha n$ and $X_i = 2$ for $\alpha n < i \leq n$. The probability of this cylinder is $(\frac{1}{2})^n$, while the pair empirical measure associated with it is

$$\frac{\alpha n - 1}{n} \delta_{11} + \frac{(1-\alpha)n - 1}{n} \delta_{22} + \frac{1}{n} \delta_{12} + \frac{1}{n} \delta_{21}.$$

Letting $n \to \infty$, we thus see why

$$I^2_{(\frac{1}{2}, \frac{1}{2})}(\alpha \delta_{11} + (1-\alpha)\delta_{22}) = -\log(\tfrac{1}{2}) \qquad \forall \alpha \in [0, 1].$$

Exercise II.13 Use the recursive definition of (Y_i).
(i) If $p = \frac{1}{2}$, then the probability of a flip is equal to $\frac{1}{2}$ and flips occur independently. Hence $\mathbb{E}\, Y_{n+1} = \mathbb{E}\, (\mathbb{E}\, (Y_{n+1} | Y_n)) = \mathbb{E}\, (\frac{1}{2}(1+2)Y_n) = \frac{3}{2}\mathbb{E}\, Y_n$. Since $Y_0 = 1$, this gives $\mathbb{E}(Y_n) = (\frac{3}{2})^n$.
(ii) In the variational expression, parametrize $\nu = (\nu_{11}, \nu_{12}, \nu_{21}, \nu_{22}) = (a, b, b, 1 - a - 2b)$, with $a, b \geq 0$ and $a + 2b \leq 1$. A little calculation shows that the supremum is attained at $a = \frac{1}{6}$, $b = \frac{1}{3}$ and equals $\log(\frac{3}{2})$.

Exercise II.14 Pick $a = (1-p)^2 - \delta$, $b = p(1-p) + \delta$ in the parametrization of ν. Then, for small δ, the functional under the supremum equals $[2p(1-p) + \delta] \log 2 + O(\delta^2)$. Hence $\delta = 0$ is not optimal.

Exercise II.17 According to (II.16), we need to compute $\inf_{\nu \in \mathfrak{M}_1(\Gamma):\, m_\nu = z} I_\rho(\nu)$. The infimum is subject to the conditions $\sum_s \nu_s = 1$ and $\sum_s s\nu_s = z$. Using Lagrange multipliers c_1 and c_2, we thus need to compute

$$\inf_{(\nu_s):\, \nu_s \geq 0} \sum_s \Big[\nu_s \log\Big(\frac{\nu_s}{\rho_s}\Big) - c_1 \nu_s - c_2 s \nu_s \Big].$$

Since $x \mapsto x \log x$ is steep at $x = 0$ and $x = \infty$, the infimum is attained in $(0, \infty)^r$. Differentiate w.r.t. ν_s and use that $\sum_s \nu_s = 1$, to get $\nu_s = \rho_s e^{cs} / \sum_s \rho_s e^{cs}$ for an appropriate $c = c_2 = c_2(z) \in \mathbb{R}$. Use $\sum_s s\nu_s = z$ to get $I_\rho(\nu) = \sum_s \nu_s \log(\nu_s/\rho_s) = cz - \log(\sum_s \rho_s e^{cs})$. Optimize over c to get the result.

Exercise II.21 Assume that $\nu \in \widetilde{\mathfrak{M}}_1(\Gamma^N)$. Then, for all $s_1, \dots, s_{N-2}$,

$$\begin{aligned}\sum_{s_{N-1}} \bar{\nu}_{s_1,\dots,s_{N-2},s_{N-1}} &= \sum_{s_{N-1},s_N} \nu_{s_1,\dots,s_{N-2},s_{N-1},s_N} \\ &= \sum_{s_{N-1},s_N} \nu_{s_N,s_1,\dots,s_{N-2},s_{N-1}} = \sum_{s_N} \bar{\nu}_{s_N,s_1,\dots,s_{N-2}}.\end{aligned}$$

Exercise II.22 Assume that $\mu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}})$. Then, for all $s_1, \dots, s_{N-1}$,

$$\begin{aligned}\sum_{s_N}(\pi_N\mu)_{s_1,\dots,s_{N-1},s_N} &= \mu\big(\{x \in \Gamma^{\mathbb{N}}: x_1 = s_1, \dots, x_{N-1} = s_{N-1}\}\big) \\ &= \mu\big(\{x \in \Gamma^{\mathbb{N}}: x_2 = s_1, \dots, x_N = s_{N-1}\}\big) \\ &= \sum_{s_N}(\pi_N\mu)_{s_N,s_1,\dots,s_{N-1}}.\end{aligned}$$

Exercise II.25 Recall that the RHS of (II.19) is $H(\pi_N\nu|\pi_{N-1}\nu \times \rho)$.
Property 1: Write $\nu_{s_1,\dots,s_N} = (\pi_N\nu)_{s_1,\dots,s_N}$. By applying Jensen's inequality to (II.19), we get

$$\begin{aligned}H(\pi_N\nu|\pi_{N-1}\nu \times \rho) &= -\sum_{s_1,\dots,s_N} \nu_{s_1,\dots,s_N} \log\Big(\frac{\nu_{s_1,\dots,s_{N-1}}\rho_{s_N}}{\nu_{s_1,\dots,s_N}}\Big) \\ &= -\sum_{s_2,\dots,s_N} \nu_{s_2,\dots,s_N} \sum_{s_1} \frac{\nu_{s_1,\dots,s_N}}{\nu_{s_2,\dots,s_N}} \log\Big(\frac{\nu_{s_1,\dots,s_{N-1}}\rho_{s_N}}{\nu_{s_1,\dots,s_N}}\Big) \\ &\geq -\sum_{s_2,\dots,s_N} \nu_{s_2,\dots,s_N} \log\Big(\sum_{s_1} \frac{\nu_{s_1,\dots,s_{N-1}}\rho_{s_N}}{\nu_{s_2,\dots,s_N}}\Big) \\ &= H(\pi_{N-1}\nu|\pi_{N-2}\nu \times \rho).\end{aligned}$$

Property 2: It is obvious that $a \mapsto J(a)$ is non-decreasing. To prove that it is right-continuous, we use that $B_a^c(\rho^{\mathbb{N}})$ is open. Indeed, for every $\epsilon > 0$ there exists a $\nu_\epsilon \in B_a^c(\rho^{\mathbb{N}})$ such that $I_\rho^\infty(\nu_\epsilon) \leq J(a) + \epsilon$. However, for all $\delta > 0$ sufficiently small we have $\nu_\epsilon \in B_{a+\delta}^c(\rho^{\mathbb{N}})$ and hence $J(a+\delta) \leq I_\rho^\infty(\nu_\epsilon)$. Therefore $\lim_{\delta\downarrow 0} J(a+\delta) \leq J(a) + \epsilon$. Let $\epsilon \downarrow 0$ to get $\lim_{\delta\downarrow 0} J(a+\delta) = J(a)$.

Exercise II.30 Use (II.24).
(i) A non-decreasing sequence of continuous functions has a lower-semicontinuous limit. The affineness property is explained in Comment (2) in Section II.6. From (II.32) it is clear that $I_\rho^\infty(\nu) \leq \max_s \log(1/\rho_s)$ uniformly in ν.
(ii) It follows from (II.24) that $I_\rho^\infty(\nu) = 0$ if and only if $I_\rho^N(\pi_N\nu) = 0$ for all $N \geq 2$. But then $\pi_N\nu = \rho^N$ for all $N \geq 2$, and so $\nu = \rho^{\mathbb{N}}$.

Exercise II.35 For all $\nu \in \widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{Z}})$, we have the recursion

$$\begin{aligned}h(\pi_N\nu) &= -\sum_{s_1,\dots,s_N} \nu_{s_1,\dots,s_N} \log \nu_{s_1,\dots,s_N} \\ &= -\sum_{s_1,\dots,s_{N-1}} \nu_{s_1,\dots,s_{N-1}} \log \nu_{s_1,\dots,s_{N-1}} \\ &\quad + \sum_{s_1,\dots,s_{N-1}} \nu_{s_1,\dots,s_{N-1}} \Big[-\sum_{s_N} \frac{\nu_{s_1,\dots,s_N}}{\nu_{s_1,\dots,s_{N-1}}} \log\Big(\frac{\nu_{s_1,\dots,s_N}}{\nu_{s_1,\dots,s_{N-1}}}\Big)\Big] \\ &= h(\pi_{N-1}\nu) + \mathbb{E}\Big(h\big(\pi_{\{N\}}\nu(\mathcal{F}_{\{1,\dots,N-1\}})\big)\Big), \qquad N \geq 2,\end{aligned}$$

where $\pi_{\{N\}}\nu(\mathcal{F}_{\{1,\dots,N-1\}}) \in \mathfrak{M}_1(\Gamma)$ is the conditional probability law of the N-th coordinate given the sigma-field $\mathcal{F}_{\{1,\dots,N-1\}}$ generated by the coordinates labelled $1,\dots,N-1$. Iterate the recursion and use the shift-invariance of ν, to obtain

$$h(\pi_N\nu) = \sum_{M=0}^{N-1} \mathbb{E}\left(h\big(\pi_{\{0\}}\nu(\mathcal{F}_{\{-M,\dots,-1\}})\big)\right), \qquad N \geq 1.$$

Now use that

$$h\big(\pi_{\{0\}}\nu(\mathcal{F}_{\{-M,\dots,-1\}})\big) \underset{M\to\infty}{\longrightarrow} h\big(\pi_{\{0\}}\nu(\mathcal{F}_{-\mathbb{N}})\big) \qquad \mathbb{P}\text{-a.s.}$$

by the Backward Martingale Convergence Theorem (see P. Billingsley, *Probability and Measure* (2nd. ed.), Wiley, New York, 1986, Section 35.

Exercise II.38 Use the definition of the truncation π_N.
Property 1: The lower bound is trivial. Let $K_N = \{s \geq N\colon \nu_s \geq \rho_s\}$, $A_N = \frac{1}{2}\sum_{s\in K_N}(\nu_s - \rho_s) \geq 0$ and $B_N = -\frac{1}{2}\sum_{s\notin K_N, s\geq N}(\nu_s - \rho_s) \geq 0$. Then $0 \leq d(\nu,\rho) - d(\pi_N\nu, \pi_N\rho) = A_N + B_N - |A_N - B_N| \leq 2B_N$. But $2B_N \leq \sum_{s\notin K_N, s\geq N}\rho_s \leq \sum_{s\geq N}\rho_s$.
Property 2: Same as in Exercise II.25.
Property 3: Write (II.37) as $I_\rho(\nu) = \sum_{s\in\mathbb{N}}\rho_s h(\nu_s/\rho_s)$ with $h(x) = x\log x$. Since h is bounded from below, the sum is convergent (possibly infinite). Jensen's inequality gives $I_\rho(\nu) \geq 0$. To prove the last assertion, put $\nu_{[N]} = \sum_{s>N}\nu_s$ and $\rho_{[N]} = \sum_{s>N}\rho_s$. Let $\Delta = I_{\pi_{N+1}\rho}(\pi_{N+1}\nu) - I_{\pi_N\rho}(\pi_N\nu)$. Then an easy calculation gives

$$\Delta = \rho_N h\Big(\frac{\nu_N}{\rho_N}\Big) + \rho_{[N]} h\Big(\frac{\nu_{[N]}}{\rho_{[N]}}\Big) - (\rho_N + \rho_{[N]}) h\Big(\frac{\nu_N + \nu_{[N]}}{\rho_N + \rho_{[N]}}\Big).$$

By the convexity of h, the RHS is ≥ 0, with equality if and only if $\nu_N/\nu_{[N]} = \rho_N/\rho_{[N]}$.

Exercise II.40 Use property 3.
(i) A non-decreasing sequence of continuous convex functions has a lower-semicontinuous convex limit. Strict convexity follows from Jensen's inequality like in Exercise II.5(i).
(ii) Using the notation of Exercise II.38, we have $I_\rho(\nu) \geq I_{\pi_{N+1}\rho}(\pi_{N+1}\nu) \geq R_N(\nu)$, with

$$R_N(\nu) = \nu_{[N]}\log\frac{\nu_{[N]}}{\rho_{[N]}} + (1-\nu_{[N]})\log\frac{1-\nu_{[N]}}{1-\rho_{[N]}}.$$

Hence, for all $K \geq 0$,

$$\{\nu \in \mathfrak{M}_1(\mathbb{N})\colon I_\rho(\nu) \leq K\} \subseteq \bigcap_{N\in\mathbb{N}} \{\nu \in \mathfrak{M}_1(\mathbb{N})\colon R_N(\nu) \leq K\}.$$

Since $\lim_{N\to\infty}\rho_{[N]} = 0$, the RHS is a compact subset of $\mathfrak{M}_1(\mathbb{N})$. The LHS is a closed subset of $\mathfrak{M}_1(\mathbb{N})$, because I_ρ is lower semi-continuous. Hence it is compact too.
(iii) By property 3, $I_\rho(\nu) \geq I_{\pi_N\rho}(\pi_N\nu) \geq 0$ for all $N \geq 1$, with the last inequality being an equality if and only if $\nu_s = \rho_s$ for all $s < N$ and $\nu_{[N]} = \rho_{[N]}$. Since N is arbitrary, the latter gives $\nu = \rho$.

Chapter III:

Exercise III.2 (i)$\Rightarrow$(ii): Suppose that (ii) fails, i.e., $\lim_{\epsilon\downarrow 0}\inf_{y\in B_\epsilon(x)} f(y) < f(x)$. Then there exists a sequence (x_n), with $x_n \to x$ as $n \to \infty$, such that $f(x_n) \leq f(x) - \delta$ for some $\delta > 0$ and n large enough. This contradicts (i).

(ii)⇒(iii): Fix $c \geq 0$. The set $\{x \in \mathcal{X}: f(x) > c\}$ is open, because (ii) and $f(x) > c$ imply $f(y) > c$ for all $y \in B_\epsilon(x)$ with $\epsilon > 0$ small enough.
(iii)⇒(i): Suppose that (i) fails, i.e., there exists a sequence (x_n) with $x_n \to x$ and $f(x_n) \to c < f(x)$ as $n \to \infty$. Pick $0 < \delta < f(x) - c$. Then $x_n \in \{x \in \mathcal{X}: f(x) \leq c + \delta\}$ for n large enough. By compactness of this level set, it follows that $f(x) \leq c + \delta$, which contradicts the choice of δ.

Exercise III.4 Let K be a compact subset of $\mathcal{X}$, and let $c = \inf_{x \in K} f(x)$. Then there exists a sequence (x_n) in K such that $f(x_n) \to c$ as $n \to \infty$. Suppose that $c > -\infty$. Then $\{x \in \mathcal{X}: f(x) \leq c + \frac{1}{N}\} \cap K$ is non-empty and compact for all $N \in \mathbb{N}$. Hence there exist a sequence (n_k) such that $x_{n_k} \in \{f \leq a + \frac{1}{N}\} \cap K$ for all $k \geq N$ and $x_{n_k} \to x \in K$ as $k \to \infty$. By the lower semi-continuity of f, we have $f(x) \leq c$, so the infimum is attained at x. If $c = -\infty$, then we use the same argument with the sets $\{x \in \mathcal{X}: f(x) \leq -N\} \cap K$.

Exercise III.9 We must verify Definitions III.5 and III.6.
1. $P_n(S) = \frac{1}{2n}|S \cap [-n, n]|$. For any bounded interval $S \subset \mathbb{R}$, we have $\lim_{n\to\infty} \frac{1}{n} \log P_n(S) = 0$. If the LDP holds with rate n, then Definition III.6(D2′) implies that $I(C) = 0$ for any closed interval $C \subset \mathbb{R}$. But then necessarily $I \equiv 0$, which is ruled out by Definition III.5(D1).
2. $P_n(S) = 2n|S \cap [-\frac{1}{n}, \frac{1}{n}]|$. For any closed interval $C \subset \mathbb{R}$ containing a neighborhood of 0, we have $\lim_{n\to\infty} \frac{1}{n} \log P_n(C) = 0$. Similarly, for any open interval $O \subset \mathbb{R}$ outside a neighborhood of 0, we have $\lim_{n\to\infty} \frac{1}{n} \log P_n(O) = -\infty$. If the LDP holds with rate n, then Definitions III.6(D2′) and III.6(D3′) yield that $I(C) = 0$, respectively, $I(O) = \infty$. Hence $I(x) = \infty$, $x \neq 0$, and $I(0) = 0$. It is easy to check that, for this I, Definition III.6(D2′) holds for all C closed and Definition III.6(D3′) holds for all O open.
3. $P_n(S) = \frac{1}{2}|S \cap [-1, 1]|$. As in 1, we find $I(x) = 0$, $x \notin \{-1, 1\}$. By lower semi-continuity, this gives $I \equiv 0$. Since $[-1, 1]$ is compact, there is now no contradiction with Definition III.5(D1).

Exercise III.10 A direct calculation with Stirling's formula gives, for all $z > 0$,

$$\begin{aligned} \mathbb{P}(Z_n = \lceil znp_n \rceil) &= \binom{n}{\lceil znp_n \rceil} p_n^{\lceil znp_n \rceil} (1 - p_n)^{n - \lceil znp_n \rceil} \\ &= \left[\frac{1}{z^z} \left(\frac{1 - p_n}{1 - zp_n} \right)^{\frac{1 - zp_n}{p_n}} \right]^{np_n} e^{o(np_n)}. \end{aligned}$$

Since $\lim_{n\to\infty} p_n = 0$ and $\lim_{n\to\infty} np_n = \infty$, it follows that

$$\lim_{n\to\infty} \frac{1}{np_n} \log \mathbb{P}(Z_n = \lceil znp_n \rceil) = -I(z) \qquad \text{with} \qquad I(z) = z \log z - z + 1.$$

From this the LDP with rate np_n and with rate function I easily follows by considering closed and open sets, respectively. Note that I is the same as Cramér's rate function for Poisson(1) random variables (see Exercise I.11). This is, of course, related to the well-known fact that the Bernoulli distribution with parameters n and p_n satisfying $np_n \equiv \lambda$ converges to the Poisson distribution with parameter λ as $n \to \infty$, for any $\lambda > 0$.

Exercise III.12 The only non-continuity points of $z \mapsto I(z)$ are $z = 0, 1$. The I-continuous sets are those S such that either $\mathrm{cl}(S) \subset [0, 1]^c$ or $\mathrm{int}(S) \cap [0, 1] \neq \emptyset$.

Exercise III.19 Use (III.18). Both lower semi-continuity of I^F and $I^F \not\equiv \infty$ are inherited from the corresponding properties of I (because F is continuous). Let

$a = \sup_{x\in\mathcal{X}} F(x)$ and $b = \sup_{x\in\mathcal{X}}[F(x) - I(x)]$. Then I^F has compact level sets because, for all $c \geq 0$,

$$\begin{aligned}\{x \in \mathcal{X}\colon I^F(x) \leq c\} &= \{x \in \mathcal{X}\colon I(x) - F(x) \leq c - b\} \\ &\subseteq \{x \in \mathcal{X}\colon I(x) \leq c - b + a\},\end{aligned}$$

with the latter being compact.

Exercise III.33 Let $y_1, y_2 \in \mathcal{D}_J$. Then it follows from (III.21) that, because I has compact level sets and T is continuous, there exist $x_1, x_2 \in \mathcal{D}_I$ such that $T(x_1) = y_1$, $T(x_2) = y_2$ and $J(y_1) = I(x_1)$, $J(y_2) = I(x_2)$. Hence, for any $\alpha \in [0,1]$,

$$\begin{aligned}\alpha J(y_1) + (1-\alpha)J(y_2) &= \alpha I(x_1) + (1-\alpha)I(x_2) > I\big(\alpha x_1 + (1-\alpha)x_2\big) \\ &\geq \inf_{x\in\mathcal{X}\colon T(x)=\alpha y_1+(1-\alpha)y_2} I(x) = J\big(\alpha y_1 + (1-\alpha)y_2\big).\end{aligned}$$

Chapter IV:

Exercise IV.9 For $\mu > 0$ the maximizer u^* is unique because, for all t,

$$\left(\frac{\partial^2}{\partial u_t^2} i_\mu\right)(u^*) = \frac{1}{(u_t^*)^2}\sum_s \mu_s[-(Q_{st}^{u^*})^2 + \delta_{st}] > \frac{1}{(u_t^*)^2}\sum_s \mu_s[-Q_{st}^{u^*} + \delta_{st}] = 0.$$

Exercise IV.11 Use (IV.8).
Finiteness: Observe that $(Pu)_s \geq P_{ss}u_s$, so the function under the supremum in (IV.8) is bounded above by $\max_s \log(1/P_{ss})$. Consequently, I_P is also bounded above by this number.
Lower semi-continuity and convexity: Note that I_P is a supremum of affine functions on $\mathfrak{M}_1(\Gamma)$.
Continuity: This follows from finiteness, lower semi-continuity and convexity.
Strict convexity: It follows from (IV.8) and Exercise IV.9 that $I_P(\alpha\mu_1 + (1-\alpha)\mu_2) = \alpha I_P(\mu_1) + (1-\alpha)I_P(\mu_2)$ for $\mu_1, \mu_2 > 0$ is possible if and only if the suprema for $I_P(\mu_1)$ and $I_P(\mu_2)$ are attained at the same minimizer, say u^*. Then, however, both μ_1 and μ_2 are the stationary distribution for the same stochastic matrix Q^{u^*}, and hence $\mu_1 = \mu_2$.
Positivity: The choice u in (IV.8) yields $I_P(\mu) \geq 0$ because $P1 = 1$. If $I_P(\mu) = 0$, then μ is the stationary distribution of the matrix $Q_{st}^{u^*} = P_{st}u_t^*/(Pu^*)_s = P_{st}$ with $u^* \equiv 1$, which is P_{st}. Hence $\mu = \pi$.

Exercise IV.12 Put $v = u_1/u_2$ and write (IV.8) as

$$I_P(\mu) = \sup_{v>0}\big[-\mu_1 \log\big(p + (1-p)v^{-1}\big) - \mu_2 \log\big(p + (1-p)v\big)\big].$$

The supremum is attained at v solving the quadratic equation $p\mu_2 v^2 - (1-p)(\mu_1 - \mu_2)v - p\mu_1 = 0$. The solution is

$$v = \frac{1}{2p\mu_2}\Big[(1-p)(\mu_1 - \mu_2) + \sqrt{(1-p)^2(\mu_1-\mu_2)^2 + 4p^2\mu_1\mu_2}\Big].$$

When $p = \frac{1}{2}$, this gives $v = \mu_1/\mu_2$, and we end up with $I_P(\mu) = H(\mu|\rho)$ with $\rho = (\frac{1}{2}, \frac{1}{2})$, in agreement with (II.3).

Exercise IV.23 The proof is similar to Exercise IV.11. The function under the supremum in (IV.15) is continuous, differentiable and bounded on any compact subset of $\{u > 0\}$. The choice $u \equiv 1$ in (IV.15) yields $\widetilde{I}_G \geq 0$ because $G1 = 0$. The supremum is attained at u^* satisfying $\sum_i \nu_i G_{ij}/u_i^* = \nu_j (Gu^*)_j/u_j^{*2}$ for all j. In particular, $u^* \equiv 1$ is a maximizer if and only if $\nu G \equiv 0$, i.e., $\nu = \pi$.

Exercise IV.24 Rewrite (IV.15) as

$$\widetilde{I}_G(\nu) = \sup_{u>0} \Big[- \sum_{i=1}^{r} \frac{\nu_i}{\pi_i} \frac{(G_\pi u)_i}{u_i} \Big],$$

with $(G_\pi)_{ij} = \pi_i G_{ij}$. Since G_π is symmetric, the claim follows from (IV.16).

Exercise IV.25 According to Theorems III.20 and IV.14(b), the rate function for $\frac{1}{t} S_t$ is

$$I(z) = \inf_{\nu \in \mathfrak{M}_1(\Gamma):\ \sum_i i\nu_i = z} \langle \sqrt{\nu}, (-G)\sqrt{\nu} \rangle.$$

The infimum is empty when $z \notin [1, r]$. Since $G_{ij} = 1$ for all $i, j \in \Gamma$ with $i \neq j$, we thus need to compute (compare with Exercise II.17)

$$\inf_{(\nu_i):\ \nu_i \geq 0} \Big[- \sum_{i,j:\ i \neq j} \sqrt{\nu_i \nu_j} + (r-1) \sum_i \nu_i - c_1 \sum_i \nu_i - c_2 \sum_i i\nu_i \Big],$$

where c_1, c_2 are Lagrange multipliers. The minimizer ν^* satisfies the equation

$$\sqrt{\nu_i^*} = \tfrac{1}{2} \zeta^* \big[(r - \tfrac{1}{2}) - c_1 - c_2 i \big]^{-1}, \qquad i \in \Gamma,$$

where $\zeta^* = \sum_i \sqrt{\nu_i^*}$ and c_1, c_2 are to be chosen such that $\sum_i \nu_i^* = 1$, $\sum_i i\nu_i^* = z$. The minimum is $I(z) = r - \zeta^{*2}$. For $r = 2$, the solution is: $\zeta^* = \sqrt{z-1} + \sqrt{2-z}$, $\nu_1^* = 2 - z$, $\nu_2^* = z - 1$.

Chapter V:

Exercise V.10 Clearly, $\varphi_n(t) = \mathbb{E}\, e^{\langle t, Z_n \rangle}$ with $t = (t_1, t_2)$ is a function of t_2 only, and is finite for all $t_2 \in \mathbb{R}$ (recall Exercise I.11(3)). Since $\Lambda^*(x) = \sup_{t_1,t_2} [x_1 t_1 + x_2 t_2 - \Lambda(t)]$, it follows that $\Lambda^*(x) = \infty$ unless $x_1 = 0$. Thus, $\mathcal{D}_{\Lambda^*}$ is the vertical axis in $\mathbb{R}^2$, which has empty interior (in $\mathbb{R}^2$) but has full relative interior.

Exercise V.11 Since $\mathbb{P}(Z_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}$, $k = 0, \ldots, n$, we find that $\varphi_n(t) = \mathbb{E}\, e^{t(Z_n/np_n)} = [p_n e^{t/np_n} + (1-p_n)]^n$. From this it follows that $\Lambda(t) = \lim_{n\to\infty} \frac{1}{np_n} \log \varphi_n(np_n t) = e^t - 1$. Applying Theorem V.6(c) with np_n replacing n, we get the LDP with rate np_n and with rate function $\Lambda^*(x) = x \log x - x + 1$ (recall Exercise I.11(1)).

Exercise V.12 Under the weaker condition $\sum_{i\in\mathbb{Z}} C_i^2 < \infty$, the process $(X_i)_{i\in\mathbb{Z}}$ can be represented as

$$X_i = \sum_{j\in\mathbb{Z}} a_j Y_{i-j},$$

with $(Y_i)_{i\in\mathbb{Z}}$ i.i.d. standard normal and $\sum_{i\in\mathbb{Z}} a_i^2 < \infty$ (see I.A. Ibragimov and Yu.V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, Groningen, 1971, Sections 16.4 and 16.7). In terms of this representation the covariance function reads $C_i = \mathbb{E}\, X_0 X_i = \sum_{j\in\mathbb{Z}} a_j a_{i+j}$. Compute

$$\begin{aligned}\varphi_n(nt) &= \mathbb{E}\, e^{t\sum_{i=1}^n X_i} = \mathbb{E}\,\exp\Big[t\sum_{k\in\mathbb{Z}}\Big(\sum_{i=1}^n a_{i-k}\Big)Y_k\Big]\\ &= \exp\Big[\tfrac{1}{2}t^2\sum_{k\in\mathbb{Z}}\Big(\sum_{i=1}^n a_{i-k}\Big)^2\Big]\\ &= \exp\Big[\tfrac{1}{2}t^2\sum_{i,j=1}^n C_{j-i}\Big]\\ &= \exp\Big[\tfrac{1}{2}t^2\sum_{i=-(n-1)}^{n-1}(n-i)C_i\Big].\end{aligned}$$

It follows from (V.1) that $\Lambda(t) = \lim_{n\to\infty}\frac{1}{n}\log\varphi_n(nt) = \frac{1}{2}t^2 C$, with $C = \sum_{i\in\mathbb{Z}} C_i$. Hence, (V.3) gives $\Lambda^*(x) = x^2/2C$.

Exercise V.13 Let $\psi(t) = \mathbb{E}\, e^{\langle t, X_1\rangle}$. Then $\varphi_n(nt) = e^{n(\psi(t)-1)}$. Hence $\Lambda(t) = \psi(t) - 1$. From the solution to Exercise I.11(3) we know that $\psi(t) = e^{|t|^2/2}$. Hence $\Lambda^*(x)$ is a function of $|x|$ only, and a straightforward calculation gives the claim.

Exercise V.14 We use the representation $\lambda(t) = \sup_{x,y}\langle x, Q(t)y\rangle/\langle x,y\rangle$ (see Seneta [C8] Chapter 1). Let $e \in \mathbb{R}^r \times \mathbb{R}^r$ be given by $e_{11} = 1$, $e_{rr} = -1$ and $e_{ij} = 0$ otherwise. For any $u \in \mathbb{R}$, we have $\lambda(ue) \geq Q_{11}(ue) = P_{11}e^u$ and $\lambda(ue) \geq Q_{rr}(ue) = P_{rr}e^{-u}$. Hence, $\lim_{u\to\infty}\log\lambda(ue) = \lim_{u\to-\infty}\log\lambda(ue) = \infty$. Therefore $t \mapsto \log\lambda(t)$ is not linear. But then, by analyticity, it must be strictly convex.

Chapter VI:

Exercise VI.3 See e.g. W.R. Pestman, *Mathematical Statistics*, de Gruyter, Berlin, 1998, Section III.1.

Exercise VI.6 Let $\gamma = \frac{\delta_1}{\delta_0} > 1$. Then $\varphi_0(t) = \frac{\gamma^t}{1+(\gamma-1)t}$, $t > -\frac{1}{\gamma-1}$, and infinite otherwise (recall Exercise I.11(2)). Hence $t \mapsto \log\varphi_0$ assumes a minimum in $t = \frac{1}{\log\gamma} - \frac{1}{\gamma-1} > 0$, and

$$I_0(0) = [x - \log x - 1]_{x=\frac{\log\gamma}{\gamma-1}}.$$

Note that the latter is invariant under the transformation $\gamma \to \gamma^{-1}$, i.e., is symmetric in δ_0 and δ_1.

Chapter VII:

Exercise VII.5 The LDP in Theorem VII.3 requires that (pick $C = \{1\}$ in Definition III.6(D2′))

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\omega(X_n = n) \leq -I(1).$$

Since $\mathbb{P}_\omega(X_n = n) = \prod_{x=0}^{n-1}\omega_x$, this gives us the upper bound $I(1) \leq -\langle\log\omega_0\rangle$. Pick $\epsilon > 0$. The LDP also requires that (pick $O = (1-\epsilon, 1)$ in Definition III.6(D3′))

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\omega\big((1-\epsilon)n < X_n < n\big) \geq -\inf_{x\in(1-\epsilon,1)} I(x).$$

We have $\mathbb{P}_\omega((1-\epsilon)n < X_n < n) \leq N_n(\epsilon)\prod_{x=0}^{\lceil(1-\epsilon)n\rceil-1}\omega_x$, with $N_n(\epsilon)$ the number of n-step paths that start at 0 and end between $(1-\epsilon)n$ and n. We know that

$\lim_{n\to\infty}\frac{1}{n}\log N_n(\epsilon) = -\frac{\epsilon}{2}\log\frac{\epsilon}{2} - (1-\frac{\epsilon}{2})\log(1-\frac{\epsilon}{2})$. Hence we get the lower bound

$$\inf_{x\in(1-\epsilon,1)} I(x) \geq -(1-\epsilon)\langle\log\omega_0\rangle + \tfrac{\epsilon}{2}\log\tfrac{\epsilon}{2} + (1-\tfrac{\epsilon}{2})\log(1-\tfrac{\epsilon}{2}).$$

Thus, $\liminf_{x\uparrow 1} I(x) \geq -\langle\log\omega_0\rangle$. But $I(1) = \lim_{x\uparrow 1} I(x)$ (by Definition III.5(D2)), and $-\langle\log\omega_0\rangle = \langle\log(1+\rho)\rangle$, so we get the claim. The computation for $I(-1)$ is analogous.

Exercise VII.11 Let $\omega_{\min} = \operatorname{essinf}[\operatorname{supp}(\alpha)]$ and $\omega_{\max} = \operatorname{esssup}[\operatorname{supp}(\alpha)]$. We have $0 < \omega_{\min} < \omega_{\max} < 1$, because we have assumed that α is non-degenerate with support bounded away from 0 and 1. For all $\theta \in (0,1]$,

$$\begin{aligned}\mathbb{P}_\omega(X_n = \lceil\theta n\rceil) &\geq \mathbb{P}_\omega(X_{n-\lceil\theta n\rceil} = 0)\,\omega_{\min}^{\lceil\theta n\rceil},\\ \mathbb{P}_\omega(X_n = 0) &\geq \mathbb{P}_\omega(X_{n-\lceil\theta n\rceil} = \lceil\theta n\rceil)\,(1-\omega_{\max})^{\lceil\theta n\rceil}.\end{aligned}$$

Since $-I(\theta) = \lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\omega(X_n = \lceil\theta n\rceil)$ by Theorem VII.3 (already proved for $\theta \neq 0$), it follows that

$$\begin{aligned}-I(\theta) &\geq -(1-\theta)I(0) + \theta\log\omega_{\min},\\ -I(0) &\geq -(1-\theta)I(\tfrac{\theta}{1-\theta}) + \theta\log(1-\omega_{\max}),\end{aligned}$$

where we use that $\theta \mapsto I(\theta)$ is continuous on $(0,1]$ (see Fig. 13). Let $\theta \downarrow 0$, to get continuity in $\theta = 0$ from the right. The same argument gives continuity from the left.

Exercise VII.17 For $r \to -\infty$, we have $\varphi(r,\omega) = \mathbb{E}_\omega(e^{r\tau_1}) = e^r\mathbb{P}_\omega(\tau_1 = 1) + O(e^{2r})$, with the error term uniform in ω. But $\mathbb{P}_\omega(\tau_1 = 1) = \omega_0$, so (VII.8) gives $\log\lambda(r) = r + \langle\log\omega_0\rangle + O(e^r)$. Hence, $J(1) = -\langle\log\omega_0\rangle = \langle\log(1+\rho)\rangle$ by (VII.7).

Exercise VII.23 From (VII.18) and (VII.19) we know that $I''(\theta) = -\frac{1}{\theta}r'(\theta)$ and $[\log\lambda]'(r(\theta)) = \frac{1}{\theta}$. Differentiate the latter w.r.t. θ, to get $r'(\theta)[\log\lambda]''(r(\theta)) = -\frac{1}{\theta^2}$. Eliminating θ and $r'(\theta)$ from these three equations, we end up with

$$I''(\theta) = \frac{\{[\log\lambda]'(r(\theta))\}^3}{[\log\lambda]''(r(\theta))}.$$

This yields the desired expression because $\theta \downarrow 0$ corresponds to $r \uparrow 0$.

Exercise VII.24 Using Jensen's inequality, we have for the annealed model with $\mathbb{P}(\cdot) = \langle\mathbb{P}_\omega(\cdot)\rangle$ that

$$\mathbb{P}(X_n = \lfloor\theta n\rfloor) \geq e^{\langle\log\mathbb{P}_\omega(X_n = \lfloor\theta n\rfloor)\rangle}.$$

This implies that

$$-\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(X_n \geq \lfloor\theta n\rfloor) \leq \left\langle -\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\omega(X_n \geq \lfloor\theta n\rfloor)\right\rangle = \langle I(\theta)\rangle = I(\theta),$$

the rate function for the quenched model.

Chapter VIII:

Exercise VIII.6 Use (VIII.2) to compute

$$e^{H(t)} = \int_{-\infty}^{\infty} ds\; e^{ts}\frac{d}{ds}\Big[-e^{-e^{s/\rho}}\Big] = \int_0^\infty dx\; x^{t\rho}\frac{d}{dx}\Big[-e^{-x}\Big] = \Gamma(\rho t + 1),$$

where we substitute $s = \rho\log x$ to get the second equality. Stirling's formula gives $\log\Gamma(\rho t + 1) = \rho t[\log(\rho t) - 1 + o(1)]$, $t \to \infty$, which yields (VIII.5).

Exercise VIII.7 Let $F(s) = \mathbb{P}(\xi(0) \leq s)$, $s \in \mathbb{R}$, and $M = \min\{s \in \mathbb{R}\colon F(s) = 1\}$. Then (VIII.2) gives $cM \geq \frac{1}{t}H(ct) \geq cM + \frac{1}{t}\log\int_{-\infty}^{M} e^{ct(s-M)}dF(s)$. The last integral can be bounded from below as follows. Pick $N < M$, and estimate

$$\begin{aligned}\int_{-\infty}^{M} e^{ct(s-M)}dF(s) &\geq \int_{N}^{M} e^{ct(s-M)}dF(s)\\ &\geq [1-F(N)]\exp\Big[ct\int_N^M (s-M)\tfrac{1}{1-F(N)}dF(s)\Big],\end{aligned}$$

where we use Jensen's inequality. It follows that

$$\liminf_{t\to\infty}\frac{1}{t}\log\int_{-\infty}^{M} e^{ct(s-M)}dF(s) \geq c\int_N^M (s-M)\frac{1}{1-F(N)}dF(s).$$

The RHS tends to zero as $N \uparrow M$, and so the conclusion is that $\lim_{t\to\infty}\frac{1}{t}H(ct) = cM$, which makes the RHS of (VIII.5) vanish.

Exercise VIII.14 Use $H(0) = 0$ to write $\frac{1}{t}[H(ct) - cH(t)] = \frac{c}{t}\int_0^t ds\int_s^{cs} du\, H''(u)$. Substitution of $H''(u) \sim \frac{\rho}{u}[1+o(1)]$, $u \to \infty$, gives $\lim_{t\to\infty}\frac{1}{t}[H(ct) - cH(t)] = \rho c\log c$.

Exercise VIII.18 Pick $N, N' \in \mathbb{N}$ with $N' > N$. Any $\nu \in \mathfrak{M}_1(T_N)$ with $\nu(z) = 0$ $\forall z \in \partial T_N$ can be extended to a $\nu' \in \mathfrak{M}_1(T_{N'})$ with $\nu'(z) = 0$ $\forall z \in \partial T_{N'}$ by putting $\nu' = \nu$ on T_N and $\nu' = 0$ and $T_{N'}\backslash T_N$. Since $I^{N'}(\nu') = I^N(\nu)$ and $J^{N'}(\nu') = J^N(\nu)$, it follows that $\widehat{\chi}^{N'}(\rho) \leq \widehat{\chi}^N(\rho)$ (recall (VIII.9) and (VIII.10)).

Exercise VIII.19 Use that $F_1(\nu) = I_1(\nu) + \rho J_1(\nu)$ for $\nu \in \mathfrak{M}_1(\mathbb{Z})$.
(1) Let ν be any minimizer of (VIII.9) such that $\nu(z) = 0$ and $\nu(z-1)+\nu(z+1) > 0$. Let $\nu_\alpha = (1-\alpha)\nu + \alpha\delta_z$, $\alpha \in [0,1]$, where δ_z is the point-measure at z. Then an easy calculation based on (VIII.10) gives

$$\begin{aligned}I_1(\nu_\alpha) - I_1(\nu) &= -\alpha I_1(\nu) + 2\alpha - 2\sqrt{\alpha(1-\alpha)}\big[\sqrt{\nu(z-1)} + \sqrt{\nu(z+1)}\big],\\ J_1(\nu_\alpha) - J_1(\nu) &= -\alpha J_1(\nu) - (1-\alpha)\log(1-\alpha) - \alpha\log\alpha.\end{aligned}$$

The terms $\alpha\log\alpha$ and $\sqrt{\alpha(1-\alpha)}$ have an infinite derivative at $\alpha = 0$. Hence, $F_1(\nu_\alpha) < F_1(\nu)$ for α small enough, which is a contradiction. Therefore any minimizer ν is strictly positive.
(2) Let ν be any minimizer. Since it is strictly positive, we can do a standard variational argument. Pick any $h\colon \mathbb{Z} \mapsto \mathbb{R}$ with finite support and $\sum_z h(z) = 0$. Since $\nu + \epsilon h \in \mathfrak{M}_1(\mathbb{Z})$ for ϵ small enough, we may compute

$$\begin{aligned}0 &\leq \lim_{\epsilon\to 0}\tfrac{1}{\epsilon}\big[F_1(\nu+\epsilon h) - F_1(\nu)\big]\\ &= \sum_z\Big(\sqrt{\nu(z+1)} - \sqrt{\nu(z)}\Big)\Big(\tfrac{h(z+1)}{\sqrt{\nu(z+1)}} - \tfrac{h(z)}{\sqrt{\nu(z)}}\Big) - \rho\sum_z h(z)\big[1+\log\nu(z)\big]\\ &= \sum_z h(z)\Big(-\sqrt{\tfrac{\nu(z+1)}{\nu(z)}} - \sqrt{\tfrac{\nu(z-1)}{\nu(z)}} + 2 - \rho\log\nu(z)\Big).\end{aligned}$$

Since h is arbitrary, we get the Euler-Lagrange equation as stated.

Chapter IX:

Exercise IX.10 It is clear from (VII.4) that $K_n(1) = 2^{-n}$, because speed 1 corresponds to the walk that only steps to the right. From this the claim follows via a simple perturbation argument, in the spirit of Exercise VII.5.

Exercise IX.11 Recall (IX.7).
(i) Let θ be a rational number in $(0,1)$. Pick m,n such that $\theta m, \theta n$ are integer. Then, by forcing the path to cross the bond between θm and $\theta m+1$ only once, we get

$$\begin{aligned}
&\widehat{K}_{m+n+1}(\theta)\\
&= \mathbb{E}\Big(e^{-\beta \widehat{I}_{m+n+1}[(S_i)_{i=0}^{m+n+1}]}\\
&\qquad 1_{\{S_{m+n+1}=\lceil\theta(m+n+1)\rceil\}}\\
&\qquad 1_{\{0\le S_i\le S_{m+n+1}\forall\ 0\le i\le m+n+1\}}\Big)\\
&\ge \mathbb{E}\Big(e^{-\beta \widehat{I}_m[(S_i)_{i=0}^{m}]}\, e^{-\beta \widehat{I}_n[(S_i)_{i=m+1}^{m+n+1}]}\\
&\qquad 1_{\{S_m=\theta m\}}\, 1_{\{S_{m+1}=\theta m+1\}}\, 1_{\{S_{m+n+1}=S_{m+1}+\theta n\}}\\
&\qquad 1_{\{0\le S_i\le S_m \forall\, 0\le i\le m\}}\, 1_{\{S_{m+1}\le S_i\le S_{m+n+1}\forall\, m+1\le i\le m+n+1\}}\Big)\\
&= \tfrac{1}{2}\widehat{K}_m(\theta)\widehat{K}_n(\theta),
\end{aligned}$$

where we take advantage of the bridge restriction to separate the exponents. Thus, $n \mapsto -\log \frac{1}{2}\widehat{K}_{n-1}(\theta)$ is a subadditive sequence, and so Lemma III.29 tells us that $\lim_{n\to\infty} \frac{1}{n}\log\frac{1}{2}\widehat{K}_{n-1}(\theta)$ exists along n-values for which θn is integer. An easy perturbation argument (in the spirit of Exercise VII.11) shows that the limit actually exists along all n-values, and also for irrational θ.
(ii) The same type of estimate gives $\widehat{K}_{2n+1}(\frac{1}{2}(\theta+\theta')) \ge \frac{1}{2}\widehat{K}_n(\theta)\widehat{K}_n(\theta')$ for all rational θ,θ' such that $\theta n, \theta' n$ are integer. Again, the result can be extended via perturbation.

Exercise IX.44 Note that $A^{\circlearrowleft}_{r,\beta} = A_{r,\beta}S$, where S is the matrix with components $S(i,j) = 1_{\{i=j+1\}}$, $i,j\in\mathbb{N}$. Hence

$$\lambda^{\circlearrowleft}_{r,\beta} = \|A^{\circlearrowleft}_{r,\beta}\| < \|A_{r,\beta}\|\,\|S\| = \|A_{r,\beta}\| = \lambda_{r,\beta},$$

where $\|\cdot\|$ is the operator norm in $\ell^2(\mathbb{N})$. Since both eigenvalues are strictly increasing in r and since $r^{**}(\beta), r^*(\beta)$ are the values where they are equal to 1, it must be that $r^{**}(\beta) > r^*(\beta)$.

Exercise IX.45 Lemmas IX.25 and IX.27 give

$$\widehat{J}_\beta(\theta) - \theta\frac{\partial}{\partial\theta}\widehat{J}_\beta(\theta) = r_\beta(\theta), \qquad \inf_{\theta\in(0,1]}\widehat{J}_\beta(\theta) = r^*(\beta).$$

Since $J_\beta(\theta) = \widehat{J}_\beta(\theta) - r^*(\beta)$, we get $J_\beta(\theta) - \theta\frac{\partial}{\partial\theta}J_\beta(\theta) = r_\beta(\theta) - r^*(\beta)$. Since $I_\beta(0) = r^{**}(\beta) - r^*(\beta)$ according to (IX.42), we see that $\theta^{**}(\beta)$ is the unique solution of the equation $r_\beta(\theta) = r^{**}(\beta)$. Since $\theta \to r_\beta(\theta)$ is strictly decreasing for all $\beta\in(0,\infty)$ (see Fig. 19), it follows that $\theta^{**}(\beta) < \theta^*(\beta)$.

Exercise IX.46 Use Lemma IX.24 and Lemma IX.27 (see also Fig. 19).

Exercise IX.47 The analyticity of $\beta \to \theta^*(\beta)$ follows from Lemma IX.24(i), Lemma IX.27 and the implicit function theorem. For the limits $\beta \downarrow 0$ and $\beta \to \infty$, we refer the reader to Greven and den Hollander [D31] Section 3.5.

Exercise IX.48 Write $\lambda(r,\beta) = \lambda_{r,\beta}$. By Lemma IX.27, $1 = \lambda(r^*(\beta),\beta)$ for all β. Using bracketed lower indices to denote partial derivatives, we have

$$0 = \lambda_{(\beta)}(r^*(\beta),\beta) + r^*_{(\beta)}(\beta)\lambda_{(r)}(r^*(\beta),\beta).$$

By Lemma IX.25, $1/\theta^*(\beta) = \lambda_{(r)}(r^*(\beta),\beta)$ for all β. Hence

$$\left(\frac{1}{\theta^*(\beta)}\right)_{(\beta)} = \lambda_{(\beta,r)}(r^*(\beta),\beta) + r^*_{(\beta)}(\beta)\lambda_{(r,r)}(r^*(\beta),\beta).$$

Combining the last two displays, and eliminating $r^*_{(\beta)}(\beta)$, we see that $\frac{\partial}{\partial\beta}\theta^*(\beta) \geq 0$ is equivalent to

$$\left[\lambda_{(r,r)}\lambda_{(\beta)} - \lambda_{(\beta,r)}\lambda_{(r)}\right](r^*(\beta),\beta) \geq 0.$$

Chapter X:

Exercise X.33 The stationary solutions of (X.27) and (X.28) satisfy the equation

$$\frac{\partial}{\partial x}\left[\beta^{\omega,q}(x)q^\omega(x)\right] = \frac{1}{2}\frac{\partial^2}{\partial x^2}q^\omega(x).$$

Integrating once w.r.t. x, we get

$$\beta^{\omega,q}(x)q^\omega(x) = \frac{1}{2}\frac{\partial}{\partial x}q^\omega(x) + C^\omega$$

for some constant C^ω. The function $x \mapsto B_*^{\omega,r}(x) = \exp[2\int_0^x dy\, \beta^{\omega,q}(x)]$ is a homogeneous solution of the last display (i.e., with $C^\omega = 0$), while the function $x \mapsto B_*^{\omega,r}(x)\int_0^x dy/B_*^{\omega,r}(y)$ is a particular solution. Hence, the general solution is of the form

$$q^\omega(x) = C_1^\omega B_*^{\omega,r}(x) + C_2^\omega B_*^{\omega,r}(x)\int_0^x \frac{dy}{B_*^{\omega,r}(y)}.$$

The constants C_1^ω, C_2^ω are to be chosen such that q^ω is normalized and periodic. This yields

$$\begin{aligned} C_1^\omega &= \tfrac{1}{Z^{\omega,r}} B_*^{\omega,r}(2\pi)\int_0^{2\pi} \tfrac{dy}{B_*^{\omega,r}(y)}, \\ C_2^\omega &= \tfrac{1}{Z^{\omega,r}}\left[B_*^{\omega,r}(0) - B_*^{\omega,r}(2\pi)\right]. \end{aligned}$$

Since $\beta^{\omega,q}(x) = -Kr\sin x + \omega$, we have $B_*^{\omega,r}(x) = e^{-2Kr}B^{\omega,r}(x)$, with $B^{\omega,r}(x)$ as defined in (X.30). Hence, we indeed find (X.29).

Exercise X.35 Use (X.30) and (X.32).
(i) An easy computation gives $\int_0^{2\pi} dx\,(\cos x)A^{\omega,0}(x) = 0$ for all $\omega \in \mathbb{R}$.
(ii) Note that $r \mapsto A^{\omega,r}(x)$ is continuous for all x, ω.
(iii) As $r \to \infty$, the function $x \mapsto B^{\omega,r}(x)$ becomes sharply peaked around $x = 0$. Hence, so does $x \mapsto A^{\omega,r}(x)$. Consequently, $\lim_{r\to\infty}\Phi_\mu(r) = \int_{\mathbb{R}}\mu(d\omega) = 1$.

Exercise X.36 (i) Differentiate (X.32) once w.r.t. r, to get

$$\Phi'_\mu(r) = \int_{\mathbb{R}}\mu(d\omega)\left[\frac{D^\omega(r)(\frac{\partial}{\partial r}N^\omega)(r) - N^\omega(r)(\frac{\partial}{\partial r}D^\omega)(r)}{D^\omega(r)^2}\right],$$

with

$$N^\omega(r) = \int_0^{2\pi} dx\,(\cos x)A^{\omega,r}(x),$$

$$D^\omega(r) = \int_0^{2\pi} dx\,A^{\omega,r}(x).$$

A straightforward computation yields

$$N^\omega(0)=0,\quad D^\omega(0)=\frac{\pi}{\omega}\left[e^{4\pi\omega}-1\right],\quad \left(\frac{\partial}{\partial r}N^\omega\right)(0)=\frac{K}{1+4\omega^2}\,\frac{\pi}{\omega}\left[e^{4\pi\omega}-1\right].$$

Therefore $\Phi'_\mu(0) = K\int_{\mathbb{R}} \frac{\mu(d\omega)}{1+4\omega^2} = K/K_c$.

(ii–iii) Taylor expansion of $\Phi_\mu(r)$ for small Kr gives, after a little calculation,

$$\Phi_\mu(r) = \frac{1}{K_c}Kr - L(Kr)^3 + O\big((Kr)^5\big),$$

with

$$L = \int_{\mathbb{R}} \mu(d\omega)\left\{\frac{1}{2(1+\omega^2)} - \frac{8\omega^2}{1+4\omega^2}\right\}.$$

It is easily checked that if μ is unimodal, then $L>0$.

Exercise X.37 Let $K > K_c$. If $K-K_c$ is small, then $r=r(K)$ is small. It then follows, after substitution of the Taylor expansion in Exercise X.36(ii,iii) into (X.31), that

$$r(K) = \frac{1}{K_c}Kr(K) - L\big(Kr(K)\big)^3 + O\big(\big(Kr(K)\big)^5\big).$$

Hence, $r(K) \sim \sqrt{(K-K_c)/LK_c^4}$ as $K \downarrow K_c$.

Bibliography

[A] **Monographs on large deviations:**

[A1] R. Azencott, Grandes déviations et applications, in: *Ecole d'Eté de Probabilités de Saint-Flour VIII-1978* (ed. P.L. Hennequin), Lecture Notes in Math. 774, pp. 1–176, Springer, Berlin, 1980.

[A2] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (2nd. ed.), Springer, New York, 1998.

[A3] J.-D. Deuschel and D.W. Stroock, *Large Deviations*, Academic Press, London, 1989.

[A4] P. Dupuis and R.S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, Wiley, New York, 1997.

[A5] R.S. Ellis, *Entropy, Large Deviations and Statistical Mechanics*, Springer, New York, 1985.

[A6] L. Saulis and V. Statulevičius, *Limit Theorems for Large Deviations*, Kluwer, Dordrecht, 1991.

[A7] D.W. Stroock, *An Introduction to the Theory of Large Deviations*, Springer, Berlin, 1984.

[A8] S.R.S. Varadhan, *Large Deviations and Applications*, SIAM, Philadelphia, 1984.

[A9] V. Vinogradov, *Refined Large Deviation Limit Theorems*, Pitman Research Notes in Mathematics Series 315, Longman, Essex, 1994.

[A10] A.D. Wentzell, *Limit Theorems on Large Deviations for Markov Stochastic Processes*, Kluwer, Dordrecht, 1990.

[B] **Monographs on specific applications of large deviations, and related topics:**

[B1] R.R. Bahadur, *Some Limit theorems in Statistics*, vol. 4 of CBMS-NSF regional conference series in applied mathematics, SIAM, Philadelphia, 1971.

[B2] J.A. Bucklew, *Large Deviations Techniques in Decision, Simulation, and Estimation*, Wiley, New York, 1990.

[B3] R.A. Carmona and S.A. Molchanov, *Parabolic Anderson Problem and Intermittency*, AMS Memoir 518, American Mathematical Society, Providence RI, 1994.

[B4] M.I. Freidlin and A.D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer, New York, 1984.

[B5] R. van der Hofstad, *One-Dimensional Random Polymers*, CWI Tract 123, Stichting Mathematisch Centrum, Amsterdam, 1998.

[B6] O.E. Lanford, *Entropy and Equilibrium States in Classical Mechanics*, in: *Statistical Mechanics and Mathematical Problems*, Lecture Notes in Physics 20, pp. 1-113, Springer, New York, 1973.

[B7] J.T. Lewis, The large deviation principle and its application to the Boson gas, in: *Mark Kac seminar on Probability and Physics, Syllabus 1985-1989* (eds. F. den Hollander and H. Maassen), CWI Syllabus 17, pp. 85–146, CWI, Amsterdam, 1988.

[B8] S.A. Molchanov, Lectures on random media, in: *Ecole d'Eté de Probabilités de Saint-Flour XXII-1992*, Lecture Notes in Math. 1581, pp. 242–411, Springer, Heidelberg, 1994.

[B9] A. Shwartz and A. Weiss, *Large Deviations for Performance Analysis, Queues, Communications and Computing*, Chapman & Hall, London, 1995.

[B10] A.-S. Sznitman, *Brownian Motion, Obstacles and Random Media*, Springer, Berlin, 1998.

[C] **Monographs on other topics:**

[C1] N. Dunford and J.T. Schwartz, *Linear Operators, Part II: Spectral Theory*, Interscience Publishers, New York, 1963.

[C2] M. Fukushima, *Dirichlet Forms and Markov Processes*, North-Holland, Amsterdam, 1980.

[C3] H.-O. Georgii, *Gibbs Measures and Phase Transtions*, de Gruyter Studies in Mathematics 9, Walter de Gruyter, Berlin, 1988.

[C4] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, New York, 1988.

[C5] N. Madras and G. Slade, *The Self-Avoiding Walk*, Birkhäuser, Boston, 1993.

[C6] K. Petersen, *Ergodic Theory*, Cambridge University Press, Cambridge, 1983.

[C7] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[C8] E. Seneta, *Non-Negative Matrices: An Introduction to Theory and Applications*, George Allen & Unwin Ltd., London, 1973.

[D] **Research papers:**

[D1] A. de Acosta, Large deviations for empirical measures of Markov chains, J. Theor. Prob. 3 (1990) 395–431.

[D2] R.R. Bahadur and R. Ranga Rao, On deviations of the sample mean, Ann. Math. Statis. 31 (1960) 1015–1027.

[D3] R.R. Bahadur and S.L. Zabell, Large deviations of the sample mean in general vector spaces, Ann. Probab. 7 (1979) 587–621.

[D4] G. Ben Arous and M. Brunaud, Méthode de Laplace: étude variationnelle des fluctuations de diffusions de type "champs moyens', Stochastics and Stochastic Reports 31 (1990) 79–144.

[D5] G. Ben Arous and A. Guionnet, Large deviations for Langevin spin glasses, Probab. Theory Relat. Fields 102 (1995) 455–509.

[D6] G. Ben Arous and A. Guionnet, Symmetric Langevin spin glass dynamics, Ann. Probab. 25 (1997) 1367–1422.

[D7] E. Bolthausen and U. Schmock, On the maximum entropy principle for uniformly ergodic Markov chains, Stoch. Proc. Appl. 33 (1989) 1–27.

[D8] W. Bryc, Large deviations by the asymptotic value method, in: *Diffusion Processes and Related Problems in Analysis* (ed. M. Pinsky), Vol. 1, pp. 447–472, Birkhäuser, Boston, 1990.

[D9] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Ann. Math. Statist. 23 (1952) 493–507.

[D10] F. Comets, Large deviations estimates for a conditional probability distribution. Applications to random interacting Gibbs measures, Probab. Theory Relat. Fields 80 (1989) 407–432.

[D11] F. Comets, N. Gantert and O. Zeitouni, Quenched, annealed and functional large deviations for one-dimensional random walk in random environment, to appear in Probab. Theory Relat. Fields.

[D12] H. Cramér, Sur un nouveau théorème-limite de la théorie des probabilités, in: Actualités Scientifiques et Industrielles 736, Hermann, Paris, 1938, pp. 5-23.

[D13] P. Dai Pra and F. den Hollander, McKean-Vlasov limit for interacting random processes in random media, J. Stat. Phys. 84 (1996) 735-772.

[D14] D.A. Dawson and J. Gärtner, Large deviations from the McKean-Vlasov limit for weakly interacting diffusions, Stochastics 20 (1987) 247–308.

[D15] A. Dembo, Y. Peres and O. Zeitouni, Tail estimates for one-dimensional random walk in random environment, Comm. Math. Phys. 181 (1996) 667–683.

[D16] J.D. Deuschel, D.W. Stroock and H. Zessin, Microcanonical distributions for lattice gases, Comm. Math. Phys. 139 (1991) 83–101.

[D17] M.D. Donsker and S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, Comm. Pure Appl. Math. (I) 28 (1975) 1–47; (II) 28 (1975) 279–301; (III) 29 (1976) 389–461; (IV) 36 (1983) 183–212.

[D18] R.S. Ellis, Large deviations for a general class of random vectors, Ann. Probab. 12 (1984) 1–12.

[D19] R.S. Ellis, Large deviations for the empirical measure of a Markov chain with an application to the multivariate empirical measure, Ann. Probab. 16 (1988) 1496–1508.

[D20] R.S. Ellis and A.D. Wyner, Uniform large deviation property of the empirical process of a Markov chain, Ann. Probab. 17 (1989) 1147–1151.

[D21] H. Föllmer and S. Orey, Large deviations for the empirical field of a Gibbs measure, Ann. Probab. 16 (1988) 961–977.

[D22] N. Gantert and O. Zeitouni, Quenched sub-exponential tail estimates for one-dimensional random walk in random environment, Commun. Math. Phys. 194 (1998) 177-190.

[D23] N. Gantert and O. Zeitouni, Recent results on large deviations for one-dimensional random walk in random environment, in: *Random Walks* (eds. P. Révész and B. Tóth), Bolyai Mathematical Studies 8, to appear.

[D24] J. Gärtner, On large deviations from the invariant measure, Th. Prob. Appl. 22 (1977) 24–39.

[D25] J. Gärtner, On the McKean-Vlasov limit for interacting diffusions, Math. Nachr. 137 (1988) 197–248.

[D26] J. Gärtner and F. den Hollander, Correlation structure of intermittency in the parabolic Anderson model, Probab. Theory Relat. Fields 114 (1999) 1–54.

[D27] J. Gärtner and W. König, Moment asymptotics for the continuous parabolic Anderson model, to appear in Ann. Appl. Prob.

[D28] J. Gärtner, W. König and S.A. Molchanov, Almost sure asymptotics for the continuous parabolic Anderson model, preprint 1999.

[D29] J. Gärtner and S.A. Molchanov, Parabolic problems for the Anderson Hamiltonian. I. Intermittency and related topics, Commun. Math. Phys. 132 (1990) 613–655.

[D30] J. Gärtner and S.A. Molchanov, Parabolic problems for the Anderson Hamiltonian. II. Second-order asymptotics and structure of high peaks, Probab. Theory Relat. Fields 111 (1998) 17–55.

[D31] A. Greven and F. den Hollander, A variational characterization of the speed of a one-dimensional self-repellent random walk, Ann. Appl. Probab. 3 (1993) 1067–1099.

[D32] A. Greven and F. den Hollander, Large deviations for a random walk in a random environment, Ann. Probab. 22 (1994) 1381–1428.

[D33] G. Grimmett, Large deviations in subadditive processes and first-passage percolation, in: *Particle Systems, Random Media and Large Deviations*, (ed. R. Durrett), Contemporary Mathematics 41, AMS, Providence RI, 1984, pp. 175–194.

[D34] M. Grunwald, Sanov results for Glauber spin-glass dynamics, Probab. Theory Relat. Fields 106 (1996) 187–232.

[D35] A. Guionnet, Averaged and quenched propagation of chaos for spin glass dynamics, Probab. Theory Relat. Fields 109 (1997) 183–215.

[D36] T. Hara and G. Slade, The lace expansion for self-avoiding walk in five or more dimensions, Review in Mathematical Physics 4 (1992) 235–327.

[D37] R. van der Hofstad, The constants in the central limit theorem for the one-dimensional Edwards model, J. Stat. Phys. 90 (1998) 1295–1310.

[D38] R. van der Hofstad and F. den Hollander, Scaling for a random polymer, Commun. Math. Phys. 169 (1995) 397–440.

[D39] R. van der Hofstad, F. den Hollander and W. König, Central limit theorem for a weakly interacting random polymer, Markov Processes Relat. Fields 3 (1997) 1–63.

[D40] I. Iscoe, P. Ney and E. Nummelin, Large deviations of uniformly recurrent Markov additive processes, Adv. Appl. Math. 6 (1985) 373–412.

[D41] H. Kesten, The limit distribution of Sinai's random walk in random environment, Phys. A138 (1986) 299–309.

[D42] F.B. Knight, Random walks and a sojourn density process of Brownian motion, Transactions of the AMS 109 (1963) 56–86.

[D43] W. König, A central limit theorem for a one-dimensional polymer measure, Ann. Probab. 24 (1996) 1-12–1035.

[D44] H. Miller, A convexity property in the theory of random variables defined on a finite Markov chain, Ann. Math. Statist. 32 (1961) 1260–1270.

[D45] G.L. O'Brien, Sequences of capacities, with connections to large-deviation theory, J. Th. Probab. 9 (1996) 19–35.

[D46] G.L. O'Brien and W. Vervaat, Compactness in the theory of large deviations, Stoch. Proc. Appl. 57 (1995) 1–10.

[D47] S. Olla, Large deviations for Gibbs random fields, Probab. Theory Relat. Fields 77 (1988) 343–357.

[D48] S. Orey, Large deviations in ergodic theory, in: *Seminar on Stochastic Processes 12* (eds. E. Cinlar, K.L. Chung and R.K. Getoor), pp. 195–249, Birkhäuser, Boston, 1984.

[D49] S. Orey and S. Pelikan, Large deviation principles for stationary processes, Ann. Probab. 16 (1988) 1481-1495.

[D50] A. Pisztora and T. Povel, Large deviation principle for random walk in a quenched random environment in the low speed regime, to appear in Ann. Probab.

[D51] H. Sakaguchi, Cooperative phenomena in coupled oscillator systems under external fields, Prog. Theor. Phys. 79 (1988) 39–46.

[D52] I.N. Sanov, On the probability of large deviations of random variables, Mat. Sb. 42 (in russian). English translation in: Selected Translations in Mathematical Statistics and Probability I (1961) 213–244.

[D53] Ya.G. Sinai, The limiting behavior of a one-dimensional random walk in a random medium, Theory Probab. Appl. 27 (1982) 256–268.

[D54] F. Solomon, Random walks in a random environment, Ann. Probab. 3 (1975) 1–31.

[D55] S.H. Strogatz, From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators, preprint 1999, to appear in Physica D.

[D56] S.H. Strogatz and R.E. Mirollo, Stability of incoherence in a population of coupled oscillators, J. Stat. Phys. 63 (1991) 613–635.

[D57] S.H. Strogatz, R.E. Mirollo and P.C. Matthews, Coupled nonlinear oscillators below the synchronization threshold: relaxation by generalized Landau damping, Phys. Rev. Lett. 68 (1992) 2730–2733.

[D58] A.-S. Sznitman, Nonlinear reflecting diffusions process, and the propagation of chaos and fluctuations associated, J. Funct. Anal. 56 (1984) 311-336.

[D59] A.-S. Sznitman and M.P.W. Zerner, A law of large numbers for random walks in random environment, preprint 1999.

[D60] B. Tóth, The "true" self-avoiding walk with bond repulsion on $\mathbb{Z}$: limit theorems, Ann. Probab. 23 (1995) 1523–1556.

[D61] S.R.S. Varadhan, Asymptotic probabilities and differential equations, Comm. Pure Appl. Math. 19 (1966) 261-286.

[D62] M.P.W. Zerner, Lyapunov exponents and quenched large deviations for multi-dimensional random walk in random environment, Ann. Probab. 26 (1998) 1446–1476.

Index

Glossary of Symbols

α_n, VI.1
α, VII.1
$A_{r,\beta}(i,j)$, IX.7
$A^{\circlearrowleft}_{r,\beta}(i,j)$, IX.9
A_θ, IX.6
$A^{\circlearrowleft}_\theta$, IX.9
$A^{\omega,r}(x)$, X.6

$\underline{b}(t)$, X.1
$\underline{b}[0,T]$, X.1
β_n, VI.1
$\beta^{\omega,q}(x)$, X.4
$B_a(\rho)$, II.1
$B_a(\rho \times \rho)$, II.2
$B_a(\rho^N)$, II.5
$B_a(\rho^{\mathbb{N}})$, II.5
$B_\epsilon(x)$, III.1
$B^{\omega,r}(x)$, X.6

γ_0, γ_1, VI.1
$C_b(\mathcal{X})$, III.2
$c_\rho(x,y)$, VIII.4
Γ, II.1

$d(\mu,\nu)$, II.1, II.2, II.5, II.7
$d_N(\pi_N\mu, \pi_N\nu)$, II.5
Δ_n, VI.1
D_I, I.3
D_Λ, V.1
D_{Λ^*}, V.2
D_ϕ, I.4

η, VII.7
E, V.2
$\mathbb{E}$, I.1
$\mathbb{E}_\omega$, VII.1
$\mathbb{E}_x$, VIII.1

$f(x;\omega,\pi)$, X.1
$\varphi(t)$, I.3
$\varphi_0(t)$, VI.2
$\varphi_n(t)$, V.1
$\varphi(r,\omega)$, VII.3
$\widetilde{\varphi}(r,\omega)$, VII.3
F, I.3, III.2, III.3, IV.1
$\widehat{F}$, I.3
$F_d(\nu)$, VIII.6
$F_\beta(\nu)$, IX.6
$F^{\circlearrowleft}_\beta(\nu)$, IX.9
$F(Q)$, X.2
$\Phi_\mu(r)$, X.6

$g(x;\omega)$, X.1
G, IV.4
$\chi(\rho)$, VIII.3
$\chi^N(\rho)$, $\widehat{\chi}^N(\rho)$, VIII.5

$\widehat{h}(\nu)$, II.6
$H(\nu|\rho)$, II.1
$H(\nu|\bar{\nu} \times \rho)$, II.2, II.5
$H(\pi_N\nu|\pi_{N-1}\nu \times \rho)$, II.5
$H(Q|W \times \mu)$, X.3
$H(t)$, VIII.1
$H_N(\underline{x}, \underline{\omega})$, X.1

$I(z)$, I.3
$I_\rho(\nu)$, II.1, II.7
$I^2_\rho(\nu)$, II.2
$I^N_\rho(\nu)$, II.5
$I^\infty_\rho(\nu)$, II.5
$I_P(\nu)$, IV.2
$I^2_P(\nu)$, IV.2
$I^N_P(\nu)$, IV.3
$I^\infty_P(\nu)$, IV.3
$\widetilde{I}_G(\nu)$, IV.4
$I(x)$, III.1
$\widehat{I}(A)$, III.7
$I(S)$, III.1
$I_0(z)$, VI.2
$I(\theta)$, VII.2
$I(\nu)$, VIII.2
$I_\beta(\theta)$, IX.9
I_n, IX.1
$\widehat{I}_n$, IX.2
$I(Q)$, X.3

$J(y)$, III.5
$J(u)$, VII.3
$J(\nu)$, VIII.2
$J_\beta(\theta)$, $\widehat{J}_\beta(\theta)$, IX.3
$J^{\circlearrowleft}_\beta(\theta)$, IX.9

$\widetilde{J}_{\beta}^{\circlearrowleft}(\theta)$, IX.9

$\widehat{K}_n(\theta)$, IX.3
$\widetilde{K}_n(\theta)$, IX.4

$\ell_n(x)$, IX.2
$\ell(x)$, IX.4
$\ell_t(z)$, VIII.5
$\lambda(r)$, VII.3
$\lambda_{r,\beta}$, IX.7
$\lambda_\eta(r)$, VII.7
λ, X.1
L_n, I.1, II.1
L_n^2, II.2
L_n^N, II.5
L_t, IV.4
$L_t(z)$, VIII.5
$L_N(\underline{x}[0,T], \underline{\omega})$, X.2
$\Lambda(t)$, V.1
$\Lambda^*(x)$, V.1

$m(x)$, IX.5
μ, I.1, I.3
μ, II.1, II.2, II.5, II.7
μ, X.1
$\mathfrak{M}_1(\Gamma)$, II.1
$\widetilde{\mathfrak{M}}_1(\Gamma \times \Gamma)$, II.2
$\widetilde{\mathfrak{M}}_1(\Gamma^N)$, II.5
$\widehat{\mathfrak{M}}_1(\Gamma^{\mathbb{N}})$, II.5
$\mathfrak{M}_1(\mathbb{N})$, II.7

ν, II.1, II.2, II.5, II.7
$\bar{\nu}$, II.2, II.5
ν^Q, X.5

ω, VII.1
$\underline{\omega}$, X.1

π_s, IV.1
π_N, II.5, II.7
π_t, X.4
(P_n), III.1
P_{st}, IV.1
$\mathbb{P}$, I.1, III.7, V.1, VI.1, VII.1
$\mathbb{P}_x$, VIII.1
$\mathbb{P}_\omega$, VII.1
$\mathbb{P}_n$, IX.1
$\mathbb{P}^X, \mathbb{P}^Y$, IV.1
$P_\theta(i,j)$, IX.5
$\mathbb{P}_N^{\underline{\omega}}$, X.1
$\mathbb{P}_N$, X.3
$P^{\omega,Q}$, X.4
P^Q, X.4

q, X.4
q_l, X.5
q_l^ω, X.5
(Q_n), III.5
$\mathbb{Q}_n^\beta$, IX.1

Q, X.2
Q^ω, X.5
Q_*, X.5
Q_*^ω, X.5

r, VII.3, IX.7, X.6
$r_\beta(\theta)$, IX.7
$r_\eta(\theta)$, VII.7
$r^*(\beta)$, IX.8
$r^{**}(\beta)$, IX.9
ρ, II.1, VIII.1
ρ_x, VII.1
R_n, II.5

σ, I.1, I.3
σ, II.5
$\sigma^*(\beta)$, IX.10
S_n, I.1, I.3, IX.1
$\mathbb{S}_N$, X.3

$\theta^*(\beta)$, IX.2
$\theta^{**}(\beta)$, IX.9
$\tau_{r,\beta}$, IX.7
T, III.5
$T_n(X_1, \ldots, X_n)$, VI.1
T_k, VII.3
T_N, VIII.4

$u(x,t)$, VIII.1

v_α, VII.1
$v_\rho(x)$, VIII.3

$w_\rho(x)$, VIII.4
W, X.1
$\mathbb{W}_N$, X.1

$\underline{x}(t)$, X.1
$\underline{x}[0,T]$, X.1
$\underline{x}$, X.1
$\xi(x)$, VIII.1
(X_i), I.1, I.3, II.1, IV.1, VI.1, VII.1
$\mathcal{X}$, III.1

(Y_i), III.5, IV.1
$\mathcal{Y}$, III.5

Z_n, V.1
$Z(t)$, VIII.1

$\simeq$, I.1
int, cl, III.1
rint, V.2
$\langle \cdot \rangle$, VII.1, VIII.1
$\langle \cdot, \cdot \rangle$, IV.4, V.1
$(*)_n$, IX.3
$(\circ)_n$, IX.4