



Large Deviations of Return Times and Related Entropy Estimators on Shift Spaces

Noé Cuneo¹ , Renaud Raquépas² 

¹ Université Paris Cité, Sorbonne Université and CNRS, Laboratoire de Probabilités, Statistique et Modélisation, 75013 Paris, France. E-mail: noe.cuneo@u-paris.fr

² Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA

Accepted: 18 March 2024

Published online: 22 May 2024 – © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract: We prove the large deviation principle for several entropy and cross entropy estimators based on return times and waiting times on shift spaces over finite alphabets. We consider shift-invariant probability measures satisfying some decoupling conditions which imply no form of mixing nor ergodicity. We establish precise relations between the rate functions of the different estimators, and between these rate functions and the corresponding pressures, one of which is the Rényi entropy function. For the most commonly used definition of return times, the large-deviation rate function is proved to be nonconvex, except in marginal cases. The results apply in particular to irreducible Markov chains, equilibrium measures for Bowen-regular potentials, g -measures, invariant Gibbs states for absolutely summable interactions in statistical mechanics, and also to probability measures which may be far from Gibbsian, including some hidden Markov models and repeated quantum measurement processes.

Contents

1. Introduction	2
1.1 Setup and hypotheses	4
1.2 Main results	9
1.3 Outline of the proof for a toy model	12
2. Examples	14
2.1 Bernoulli measures	15
2.2 Irreducible Markov measures	17
2.3 Equilibrium measures for Bowen potentials	19
2.4 Beyond Bowen potentials: statistical mechanics and g -measures	22
2.4.1 Absolutely summable interactions in statistical mechanics	22
2.4.2 g -measures	24
2.5 Hidden Markov models and lack of exponential tightness	25
3. Key Estimates	27

3.1	Waiting times	27
3.2	From waiting times to return times	32
4.	Weak LDPs and Ruelle–Lanford Functions	34
4.1	At positive values	36
4.2	At the origin	38
5.	Proof of the Main Results	42
5.1	Proof of Theorem A	44
5.2	Proof of Theorem B	45
5.3	Proof of Theorem C	46
5.4	Nonconvexity in Theorem C	47
6.	Relation to Almost Sure Convergence Results	50
6.1	A brief review of results on almost sure convergence	50
6.2	Zeroes of the rate functions	52
A.	Supports and Subshifts	54
A.1	Specification properties	55
A.2	Measures of maximal entropy	55
B.	Technical Results About Decoupling Assumptions	56
B.1	Alternative definitions	56
B.2	Sufficient conditions	58
B.3	Decoupling properties of g -measures	59
B.4	Relation to mixing	60
C.	From Weak LDPs to Full LDPs	61
C.1	General results for nonnegative random variables	61
C.2	Proof of Theorem 0	63

1. Introduction

Return times play a fundamental role in the theory of dynamical systems. In the specific context of a one-sided shift space over a finite alphabet, there is a vast literature on the connection between return times and entropy. One of the landmark results in this direction is the following theorem, due to Wyner and Ziv [WZ89] and to Ornstein and Weiss [OW93]: if \mathbb{P} is an ergodic probability measure, then for \mathbb{P} -almost every sequence x , the time $R_n(x)$ it takes for the first n letters to reappear down the sequence x (in the same order) grows exponentially with n at a rate equal to the Kolmogorov–Sinai entropy $h(\mathbb{P})$. In other words, return times, once properly rescaled, provide a sequence $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ of universal entropy estimators. Here, “universal” means that the definition of the random variable R_n itself makes no reference to the measure \mathbb{P} ; computing $R_n(x)$ requires no explicit information about the marginals of the measure \mathbb{P} whose entropy is being estimated. Refinements of this behavior in the form of central limit theorems, laws of the iterated logarithm, large deviation principles (LDPs) and multifractal analysis have been an active area of research since the 1990s; see e.g. [Kon98, CGS99, FW01, Ols03, CU05, Joh06, CFM19, AACG23].

In the present paper, we will prove a full LDP for the sequence $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$, and give an expression of the rate function and pressure in terms of those of the LDP accompanying the celebrated Shannon–McMillan–Breiman (SMB) theorem. We will also carry this analysis for a nonoverlapping notion of return times along a sequence (see the definition of V_n below) and for waiting times (see the definition of W_n below) involving a pair of sequences as in [WZ89, Shi93, MS95, Kon98, CDEJR23a]. The latter will be related to recent results on the large deviations of $(-\frac{1}{n} \ln Q_n)_{n \in \mathbb{N}}$ with respect to \mathbb{P} , where

\mathbb{Q}_n is the n -th marginal of a second shift-invariant probability measure \mathbb{Q} . For each of these results, the assumptions on the measures involve the notions of decoupling of [CJPS19, CDEJR23a], in turn inspired by [LPS95, Pfi02]. Roughly speaking, these decoupling assumptions take the form of bounds on how strongly different parts of the sequences may depend on each other. However, unlike mixing conditions, they do not require distant symbols to be asymptotically independent. As we shall see, the decoupling assumptions cover many standard classes of probability measures, including irreducible Markov chains, equilibrium measures for sufficiently regular potentials, g -measures, invariant Gibbs states for summable interactions in statistical mechanics, as well as some less standard measures which may be far from Gibbsian in any sense.

To the best of the authors' knowledge, the large deviations of the sequence $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ were previously only understood locally (i.e. in a bounded interval containing the entropy $h(\mathbb{P})$), and only for measures satisfying the Bowen–Gibbs property [CGS99, AACG23]. The improvements we provide are three-fold. First, we extend the LDP to the whole real line, and describe the possible lack of convexity of the rate function which prevents the Gärtner–Ellis method used in [AACG23] from producing a global result. Second, we go past the Bowen–Gibbs property by focussing on the class of *decoupled measures* for which the LDP accompanying the SMB theorem and its analogue for cross entropy can be derived from results recently established in [CJPS19]. Third, we extend this analysis to the related estimators $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ and $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ to be introduced below.

Our main results also provide a sharp version of the large-deviation upper bounds on $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ that were obtained in [JB13] for mixing measures on shift spaces, and in [CRS18] for more general dynamical systems.

Organization of the paper. In the remainder of Sect. 1, we discuss our setup and main results. Theorem 0, which collects some large-deviation properties of $(-\frac{1}{n} \ln \mathbb{Q}_n)_{n \in \mathbb{N}}$ with respect to \mathbb{P} , is viewed as a starting point. Our main new results, Theorems A, B and C, deal with the large deviations of $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$, $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ and $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ respectively. In Sect. 1.3, we outline the proof on the basis of a toy model consisting of a mixture of geometric random variables, which we connect to the widely studied problem of exponential approximations of hitting and return times.

In Sect. 2, we discuss a selection of applications and provide concrete formulae for the rate functions and pressures whenever possible. It seems that even in the simplest examples (Bernoulli and Markov measures), some of our results are actually new. We also compare our results to those of [CGS99, AACG23] in Sect. 2.3, and explain how Theorems A and C prove a conjecture stated in [AACG23].

In Sect. 3, we first establish some sharp (at the exponential scale) estimates on the waiting times W_n (Sect. 3.1), which we then extend to R_n and V_n in Sect. 3.2.

Section 4 starts with some reminders about (weak) LDPs, and a brief review of the notion of Ruelle–Lanford functions. We then identify the Ruelle–Lanford functions of $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$, $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ and $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ in order to obtain the corresponding weak LDPs.

In Sect. 5, we complete the proofs of Theorems A–C: the weak LDPs of the previous section are promoted to full LDPs, the Legendre–Fenchel duality relations between the rate functions and the associated pressures are established, and the (lack of) convexity of the rate function of $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ is characterized.

In Sect. 6, we gain some insight into the set where the rate functions vanish by combining our LDPs and the corresponding almost sure convergence results (law of large numbers) in the literature.

In order to make the paper accessible to readers who may not be familiar with the tools that we will require from both large-deviation theory and dynamical systems, we include several technical, yet rather standard results in the appendices. We provide in “Appendix A” some definitions regarding subshifts and measures of maximal entropy, which are useful for some of the examples in Sect. 2 as well as for characterizing the convexity of the rate function of $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$. In “Appendix B”, we present technical results about our decoupling assumptions, and in particular some sufficient conditions for them to hold. Finally, in “Appendix C”, we give tools to promote weak LDPs to full ones, and we prove Theorem 0.

Notational conventions. We adopt the convention that $\mathbb{N} = \{1, 2, 3, 4, \dots\}$. Unless otherwise stated, measures are probability measures. We use the conventions $\ln 0 = -\infty$ and $0 \cdot (\pm\infty) = 0$ so that, in particular, $0 \cdot \ln 0 = 0$. Given a function $f : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$, we denote by $f^* : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ the Legendre–Fenchel transform (convex conjugate) of f defined by $f^*(\alpha) = \sup_{s \in \mathbb{R}} (\alpha s - f(s))$. The ball of radius $\epsilon > 0$ around the point s in a metric space is denoted $B(s, \epsilon)$.

1.1. Setup and hypotheses. We consider the one-sided shift space $\Omega := \mathcal{A}^{\mathbb{N}} = \{x = (x_k)_{k \in \mathbb{N}} : x_k \in \mathcal{A} \text{ for all } k \in \mathbb{N}\}$ for some finite alphabet \mathcal{A} . The shift map $T : \Omega \rightarrow \Omega$ is defined by $(Tx)_n := x_{n+1}$. As usual, Ω is equipped with the product topology constructed from the discrete topology on \mathcal{A} , and \mathcal{F} denotes the corresponding Borel σ -algebra.

We use common notations such as: x_k^n for the letters x_k, x_{k+1}, \dots, x_n of the sequence $x \in \Omega$; $[u]$ for the cylinder consisting in the sequences x with $x_1^n = u$; \mathbb{P}_n for the marginal of the probability measure \mathbb{P} on \mathcal{A}^n , i.e. $\mathbb{P}_n(u) = \mathbb{P}([u])$ for all $u \in \mathcal{A}^n$. We let $\Omega_{\text{fin}} := \bigcup_{n \in \mathbb{N}} \mathcal{A}^n$ be the set of words of finite length. The length of a word $u \in \mathcal{A}^n$ is denoted by $|u| = n$. The concatenation of $u, v \in \Omega_{\text{fin}}$ is denoted by uv , and the concatenation of n copies of u is denoted by u^n . We use \mathcal{F}_n for the σ -algebra generated by the cylinders sets $[u]$ with $u \in \mathcal{A}^n$, and $\mathcal{F}_{\text{fin}} := \bigcup_{n \in \mathbb{N}} \mathcal{F}_n \subset \mathcal{F}$ for the set of events involving only finitely many coordinates in Ω .

We denote by $\mathcal{P}_{\text{inv}}(\Omega)$ the set of shift-invariant Borel probability measures on Ω . For $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ and $n \in \mathbb{N}$, the support of \mathbb{P}_n is $\text{supp } \mathbb{P}_n := \{u \in \mathcal{A}^n : \mathbb{P}_n(u) > 0\}$. Moreover, $\text{supp } \mathbb{P} := \{x \in \Omega : \mathbb{P}_n(x) > 0 \text{ for all } n \in \mathbb{N}\}$ is the support of \mathbb{P} , which is a subshift of Ω . We denote by $h_{\text{top}}(\text{supp } \mathbb{P})$ the topological entropy of $\text{supp } \mathbb{P}$, which satisfies $h_{\text{top}}(\text{supp } \mathbb{P}) \leq \ln |\mathcal{A}|$; see “Appendix A”.¹

We now define the three sequences of interest, namely $(R_n)_{n \in \mathbb{N}}, (V_n)_{n \in \mathbb{N}}$ and $(W_n)_{n \in \mathbb{N}}$. The return times $R_n : \Omega \rightarrow \mathbb{N}$ are defined as

$$R_n(x) := \inf\{k \in \mathbb{N} : T^k x \in [x_1^n]\} = \inf\{k \in \mathbb{N} : x_{k+1}^{k+n} = x_1^n\}$$

for all $x \in \Omega$ and $n \in \mathbb{N}$, and their nonoverlapping counterparts $V_n : \Omega \rightarrow \mathbb{N}$ are defined as

$$V_n(x) := \inf\{k \in \mathbb{N} : T^{k+n-1} x \in [x_1^n]\} = \inf\{k \in \mathbb{N} : x_{n+k}^{2n+k-1} = x_1^n\}$$

instead. The waiting times $W_n : \Omega \times \Omega \rightarrow \mathbb{N}$ are defined as

$$W_n(x, y) := \inf\{k \in \mathbb{N} : T^{k-1} y \in [x_1^n]\} = \inf\{k \in \mathbb{N} : y_k^{k+n-1} = x_1^n\}$$

¹ We use “ $|\cdot|$ ” both for the length of words and the cardinality of discrete sets.

for all $(x, y) \in \Omega \times \Omega$ and $n \in \mathbb{N}$. In the literature, the function $W_n(x, \cdot)$ is often called *hitting time* of the set $[x_1^n]$.

We next briefly recall some results about almost sure convergence of the sequences $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$, $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ and $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$, which justify their role of as (cross) entropy estimators. The LDPs we prove in the present paper complement these almost sure convergence results, without relying on them, nor implying them in general; see Sect. 6 for an extended discussion.

First, if $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln R_n(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln V_n(x) = h_{\mathbb{P}}(x), \tag{1.1}$$

for \mathbb{P} -almost every $x \in \Omega$, where $h_{\mathbb{P}}$ is the local entropy function defined by

$$h_{\mathbb{P}}(x) := \lim_{n \rightarrow \infty} -\frac{1}{n} \ln \mathbb{P}_n(x_1^n). \tag{1.2}$$

By virtue of the SMB theorem, the last limit exists and satisfies $h_{\mathbb{P}}(x) = h_{\mathbb{P}}(Tx)$ for \mathbb{P} -almost every x , and its integral with respect to \mathbb{P} is the *Kolmogorov–Sinai entropy* (also commonly called *specific entropy* or simply *entropy*), defined as

$$h(\mathbb{P}) := \lim_{n \rightarrow \infty} -\frac{1}{n} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \ln \mathbb{P}_n(u).$$

When \mathbb{P} is ergodic, we have $h_{\mathbb{P}}(x) = h(\mathbb{P})$ for \mathbb{P} -almost every $x \in \Omega$, and in this case (1.1) can be traced back to [WZ89, OW93]. The relations (1.1) are less standard when \mathbb{P} is merely assumed to be shift invariant; see Sect. 6.1 for references.

Consider now a pair (\mathbb{P}, \mathbb{Q}) of shift-invariant probability measures, possibly with $\mathbb{P} = \mathbb{Q}$. As we will discuss in Sect. 6.1, under some assumptions more general than those of Theorem A below, it has recently been proved in [CDEJR23a] that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln W_n(x, y) = h_{\mathbb{Q}}(x)$$

for $(\mathbb{P} \otimes \mathbb{Q})$ -almost all pairs $(x, y) \in \Omega \times \Omega$, where $h_{\mathbb{Q}}$ is defined as in (1.2) and integrates with respect to \mathbb{P} to the specific *cross entropy* of \mathbb{P} relative to \mathbb{Q} , i.e.

$$h_c(\mathbb{P}|\mathbb{Q}) := \lim_{n \rightarrow \infty} -\frac{1}{n} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \ln \mathbb{Q}_n(u). \tag{1.3}$$

If \mathbb{P} is, in addition, assumed to be ergodic, then $h_{\mathbb{Q}}(x) = h_c(\mathbb{P}|\mathbb{Q})$ for \mathbb{P} -almost every $x \in \Omega$. We stress that the \mathbb{P} -almost sure existence of $h_{\mathbb{Q}}$ and the existence of the limit in (1.3) do not follow from mere shift invariance when $\mathbb{P} \neq \mathbb{Q}$. Note that, with $h_r(\mathbb{P}|\mathbb{Q})$ the specific *relative entropy*, we have $h_c(\mathbb{P}|\mathbb{Q}) = h(\mathbb{P}) + h_r(\mathbb{P}|\mathbb{Q})$ whenever both sides are well defined, and that $h_c(\mathbb{P}|\mathbb{P}) = h(\mathbb{P})$ is always well defined. The numbers $h_r(\mathbb{P}|\mathbb{Q})$ and $h_c(\mathbb{P}|\mathbb{Q})$ are both relevant in information theory (see e.g. [CT06, §5.4]²) and are commonly used in classification tasks (see e.g. [GBC16, §3.13] or [Mur22, §5.1.6]).

² Unfortunately, there is a lack of consensus on terminology. In particular, in [CT06], “cross entropy” is only mentioned as a synonym for relative entropy, and what we call cross entropy here plays a central role in Theorem 5.4.3 of [CT06] but is not given any specific name therein. As far as the authors are aware, the terminology adopted in the present paper reflects the most common usage in modern applications.

The pressure q_R associated with the sequence $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ is the function of $\alpha \in \mathbb{R}$ defined by

$$q_R(\alpha) := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \int e^{\alpha \ln R_n(x)} d\mathbb{P}(x) \tag{1.4}$$

when the limit exists. The function q_R is also referred to as the “rescaled cumulant-generating function” or “ L^α -spectrum” in the literature. Similarly, we define

$$q_V(\alpha) := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \int e^{\alpha \ln V_n(x)} d\mathbb{P}(x) \tag{1.5}$$

and

$$q_W(\alpha) := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \iint e^{\alpha \ln W_n(x,y)} d\mathbb{P}(x)d\mathbb{Q}(y) \tag{1.6}$$

when the limits exist.

The above pressures will be expressed in terms of the pressure of the sequence $(-\frac{1}{n} \ln \mathbb{Q}_n)_{n \in \mathbb{N}}$, which we define as

$$\begin{aligned} q_{\mathbb{Q}}(\alpha) &:= \lim_{n \rightarrow \infty} \frac{1}{n} \ln \int e^{-\alpha \ln \mathbb{Q}_n(x^n)} d\mathbb{P}(x) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \ln \sum_{u \in \text{supp } \mathbb{P}_n} \mathbb{Q}_n(u)^{-\alpha} \mathbb{P}_n(u) \end{aligned} \tag{1.7}$$

when the limit exists. It will be part of our assumptions below that $\mathbb{P}_n \ll \mathbb{Q}_n$ for every $n \in \mathbb{N}$, i.e. that \mathbb{P}_n is absolutely continuous with respect to \mathbb{Q}_n , so the integral and sum in (1.7) are well defined and finite for each n . When $\mathbb{Q} = \mathbb{P}$, the summand in the rightmost expression of (1.7) is $\mathbb{P}_n(u)^{1-\alpha}$; up to some sign and normalization convention, the pressure $q_{\mathbb{P}}$ thus coincides with the Rényi entropy function of \mathbb{P} .

A common route to establishing the LDP, which is followed in particular in [AACG23], is to first study the pressure in detail, and then derive the LDP using an adequate version of the Gärtner–Ellis theorem. Intrinsic to this approach are the differentiability of the pressure and the convexity of the rate function, both of which fail in our setup; see Sect. 2.3 for further discussion of the results in [AACG23]. Our approach goes in the opposite direction: first the LDP is established, and only then is some version of Varadhan’s lemma used to describe the pressure. This method allows to consider significantly more general measures, and to obtain the full LDP with a possibly nonconvex rate function. This path was already followed in [CJPS19] in order to establish, in particular, the LDP for $(-\frac{1}{n} \ln \mathbb{Q}_n)_{n \in \mathbb{N}}$ with respect to \mathbb{P} and the properties of $q_{\mathbb{Q}}$.

Assumptions. In order to state the decoupling assumptions below, we require a sequence $(\tau_n)_{n \in \mathbb{N}}$ in $\mathbb{N} \cup \{0\}$ and a sequence $(C_n)_{n \in \mathbb{N}}$ in $[1, \infty)$, assumed to be fixed and to satisfy

$$\lim_{n \rightarrow \infty} \frac{\tau_n}{n} = \lim_{n \rightarrow \infty} \frac{\ln C_n}{n} = 0. \tag{1.8}$$

We will freely write $\tau_n = o(n)$ and $C_n = e^{o(n)}$ or speak of an “ $o(n)$ -sequence” and an “ $e^{o(n)}$ -sequence” when referring to the conditions (1.8).

Definition 1.1. (UD, SLD, JSLD, admissible pair) Let $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$. We say that \mathbb{P} satisfies the *upper decoupling* assumption (UD) if for all $n, m \in \mathbb{N}$, $u \in \mathcal{A}^n$, $v \in \mathcal{A}^m$ and $\xi \in \Omega^{\tau_n}$,

$$\mathbb{P}_{n+\tau_n+m}(u\xi v) \leq C_n \mathbb{P}_n(u) \mathbb{P}_m(v). \tag{1.9}$$

We say that \mathbb{P} satisfies the *selective lower decoupling* assumption (SLD) if for all $n, m \in \mathbb{N}$, $u \in \mathcal{A}^n$ and $v \in \mathcal{A}^m$, there exist $0 \leq \ell \leq \tau_n$ and $\xi \in \mathcal{A}^\ell$ such that

$$\mathbb{P}_{n+\ell+m}(u\xi v) \geq C_n^{-1} \mathbb{P}_n(u) \mathbb{P}_m(v). \tag{1.10}$$

A pair of measures (\mathbb{P}, \mathbb{Q}) with $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_{\text{inv}}(\Omega)$ is said satisfy the *joint selective lower decoupling* assumption (JSLD) if for all $n, m \in \mathbb{N}$, $u \in \mathcal{A}^n$ and $v \in \mathcal{A}^m$, there exist $0 \leq \ell \leq \tau_n$ and $\xi \in \mathcal{A}^\ell$ such that

$$\mathbb{P}_{n+\ell+m}(u\xi v) \geq C_n^{-1} \mathbb{P}_n(u) \mathbb{P}_m(v) \quad \text{and} \quad \mathbb{Q}_{n+\ell+m}(u\xi v) \geq C_n^{-1} \mathbb{Q}_n(u) \mathbb{Q}_m(v). \tag{1.11}$$

Finally, a pair of measures (\mathbb{P}, \mathbb{Q}) with $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_{\text{inv}}(\Omega)$ is said to be *admissible* if $\mathbb{P}_n \ll \mathbb{Q}_n$ for all $n \in \mathbb{N}$, if \mathbb{P} and \mathbb{Q} satisfy **UD**, and if the pair (\mathbb{P}, \mathbb{Q}) satisfies **JSLD**.

Remark 1.2. If the pair (\mathbb{P}, \mathbb{Q}) satisfies **JSLD**, then obviously both \mathbb{P} and \mathbb{Q} satisfy **SLD**, but the converse need not hold, since in (1.11) both inequalities are required to hold for the *same* ξ . If $\tau_n = 0$ for all n , then **JSLD** becomes equivalent to \mathbb{P} and \mathbb{Q} both satisfying **SLD**.

Remark 1.3. We can always increase the constants C_n and τ_n (see Lemma B.1), so there is no loss of generality in taking the same sequences $(\tau_n)_{n \in \mathbb{N}}$ and $(C_n)_{n \in \mathbb{N}}$ for **JSLD** (resp. **SLD**) and **UD**, as well as for both measures \mathbb{P} and \mathbb{Q} .

Next, the following numbers will play an important role:

$$\gamma_+ := \limsup_{n \rightarrow \infty} \frac{1}{n} \sup_{u \in \mathcal{A}^n} \ln \mathbb{P}_n(u) \quad \text{and} \quad \gamma_- := \liminf_{n \rightarrow \infty} \frac{1}{n} \inf_{u \in \text{supp } \mathbb{P}_n} \ln \mathbb{P}_n(u). \tag{1.12}$$

One easily shows that

$$0 \leq -\gamma_+ \leq h(\mathbb{P}) \leq h_{\text{top}}(\text{supp } \mathbb{P}) \leq -\gamma_- \leq \infty; \tag{1.13}$$

see ‘‘Appendix A’’ for a proof and a definition of the topological entropy h_{top} . For some results, γ_+ will be required to be well approximated by periodic sequences:

Definition 1.4 (PA). A measure $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ satisfies the *periodic approximation* assumption (PA) if for every $\epsilon > 0$, there exists $p \in \mathbb{N}$ and $u \in \mathcal{A}^p$ such that

$$\liminf_{n \rightarrow \infty} \frac{1}{np} \ln \mathbb{P}_{np}(u^n) \geq \gamma_+ - \epsilon.$$

Remark 1.5. If \mathbb{P} satisfies **SLD** with $\tau_n = 0$, then automatically **PA** holds; a more general sufficient condition for **PA** to hold is given in Lemma B.6.

We discuss at the end of Sect. 4.2 possible ways to weaken PA.

Starting point. Our analysis is built on top of the LDP for the sequence $(-\frac{1}{n} \ln \mathbb{Q}_n)_{n \in \mathbb{N}}$ viewed as a family of random variables on (Ω, \mathbb{P}) .³ The following theorem summarizes the large-deviation properties of this sequence and essentially follows from the results of [CJPS19], where the corresponding weak LDP is proved. Taking this weak LDP for granted, the proof of the full theorem only requires some minor adjustments to the arguments in [CJPS19]; for completeness we provide the details in “Appendix C.2”. In some important applications, the conclusions of the theorem are well known and follow from standard methods; see Sects. 2.1–2.3 below. The terminology regarding (weak and full) LDPs, rate functions and exponential tightness is summarized in Sect. 4.

Theorem 0. (LDP for $-\frac{1}{n} \ln \mathbb{Q}_n$) *If (\mathbb{P}, \mathbb{Q}) is an admissible pair, then the following hold:*

- i. *The sequence $(-\frac{1}{n} \ln \mathbb{Q}_n)_{n \in \mathbb{N}}$ on (Ω, \mathbb{P}) satisfies the LDP with a convex rate function $I_{\mathbb{Q}} : \mathbb{R} \rightarrow [0, \infty]$ satisfying $I_{\mathbb{Q}}(s) = \infty$ for all $s < 0$.*
- ii. *The limit defining $q_{\mathbb{Q}}$ in (1.7) exists in $(-\infty, \infty]$ for all $\alpha \in \mathbb{R}$ and defines a non-decreasing, convex, lower semicontinuous function $q_{\mathbb{Q}} : \mathbb{R} \rightarrow (-\infty, \infty]$ satisfying $q_{\mathbb{Q}}(0) = 0$. Moreover, the following Legendre–Fenchel duality relations hold:*

$$q_{\mathbb{Q}} = I_{\mathbb{Q}}^* \quad \text{and} \quad I_{\mathbb{Q}} = q_{\mathbb{Q}}^*. \tag{1.14}$$

- iii. *Assume now that $\mathbb{Q} = \mathbb{P}$. Then, in addition to the above, the following hold:*
 - a. *The sequence $(-\frac{1}{n} \ln \mathbb{P}_n)_{n \in \mathbb{N}}$ on (Ω, \mathbb{P}) is exponentially tight and $I_{\mathbb{P}}$ is a good rate function satisfying*

$$\begin{aligned} I_{\mathbb{P}}(s) &\in [s - h_{\text{top}}(\text{supp } \mathbb{P}), s] && \text{if } s \in [-\gamma_+, -\gamma_-], \\ I_{\mathbb{P}}(s) &= \infty && \text{otherwise,} \end{aligned} \tag{1.15}$$

where $[-\gamma_+, -\gamma_-]$ is understood to be $[-\gamma_+, \infty)$ if $\gamma_- = -\infty$.

- b. *The limit superior (resp. inferior) defining γ_+ (resp. γ_-) is actually a limit.*
- c. *For every $\alpha \leq 1$,*

$$q_{\mathbb{P}}(\alpha) \leq q_{\mathbb{P}}(1) = h_{\text{top}}(\text{supp } \mathbb{P}). \tag{1.16}$$

- d. *Either $\gamma_- > -\infty$ and $q_{\mathbb{P}}(\alpha) < \infty$ for all $\alpha \in \mathbb{R}$, or $\gamma_- = -\infty$ and $q_{\mathbb{P}}(\alpha) = \infty$ for all $\alpha > 1$.*

Remark 1.6. As we shall see in Sect. 6, under the assumptions of Theorem 0, the limit defining $h_c(\mathbb{P}|\mathbb{Q})$ exists in $[0, \infty]$, and we have $I_{\mathbb{Q}}(h_c(\mathbb{P}|\mathbb{Q})) = 0$ when $h_c(\mathbb{P}|\mathbb{Q}) < \infty$ and $\lim_{s \rightarrow \infty} I_{\mathbb{Q}}(s) = 0$ when $h_c(\mathbb{P}|\mathbb{Q}) = \infty$. In particular, $I_{\mathbb{P}}(h(\mathbb{P})) = 0$ when $\mathbb{Q} = \mathbb{P}$. These conclusions will extend to the other rate functions to be introduced below; see again Sect. 6.

Remark 1.7. In the course of the proof, we shall establish and use the relation

$$I_{\mathbb{P}}(s) = s - \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \left| \left\{ u \in \mathcal{A}^n : -\frac{1}{n} \ln \mathbb{P}_n(u) \in B(s, \epsilon) \right\} \right|, \tag{1.17}$$

which can be seen as a classical “energy-entropy competition”. Note that the logarithm is either nonnegative or $-\infty$, which explains why we have either $I_{\mathbb{P}} \leq s$ or $I_{\mathbb{P}} = \infty$ in (1.15).

³ While \mathbb{Q}_n was defined as a measure on \mathcal{A}^n , it is reinterpreted here in the straightforward way as a measurable function on Ω .

Remark 1.8. A standard consequence of (1.14) and (1.15), which can also be derived from the definition of $q_{\mathbb{P}}$ directly, is that if \mathbb{P} satisfies **UD** and **SLD**, then

$$-\gamma_{\mp} = \lim_{\alpha \rightarrow \pm\infty} \frac{q_{\mathbb{P}}(\alpha)}{\alpha}. \tag{1.18}$$

Readers who are unfamiliar with such properties of Legendre–Fenchel transforms may benefit from reading the introductions in Section 2.3 of [DZ09] and Chapter VI of [Eli06], or the in-depth exposition in [Roc70].

Remark 1.9. The assumption that the alphabet \mathcal{A} is finite is important, as many of our estimates rely on the constant $|\mathcal{A}|$ in a way which does not seem easy to circumvent. We do not expect the results to remain true on countably infinite alphabets without further assumptions on the measures at hand. This is a matter we would like to investigate in future research.

1.2. Main results. We shall establish the LDP and express the rate function and pressure for the sequence $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ in terms of $I_{\mathbb{Q}}$ and $q_{\mathbb{Q}}$. Similarly, the rate functions and pressures for $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ and $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ will be expressed in terms of $I_{\mathbb{P}}$ and $q_{\mathbb{P}}$.

Theorem A. (LDP for $\frac{1}{n} \ln W_n$) *If (\mathbb{P}, \mathbb{Q}) is an admissible pair, then the following hold:*

i. *The sequence $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ satisfies the LDP with respect to $\mathbb{P} \otimes \mathbb{Q}$ with the convex rate function I_W given by*

$$I_W(s) := \begin{cases} \infty & \text{if } s < 0, \\ \inf_{r \geq s} (r - s + I_{\mathbb{Q}}(r)) & \text{if } s \geq 0. \end{cases} \tag{1.19}$$

ii. *For all $\alpha \in \mathbb{R}$, the limit in (1.6) exists in $(0, \infty]$,*

$$q_W(\alpha) = \max\{q_{\mathbb{Q}}(\alpha), q_{\mathbb{Q}}(-1)\}, \tag{1.20}$$

and the Legendre–Fenchel duality relations $q_W = I_W^$ and $I_W = q_W^*$ hold.*

iii. *If $\mathbb{Q} = \mathbb{P}$, then I_W is a good rate function and $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ is an exponentially tight family of random variables.*

Remark 1.10. The relations (1.19) and (1.20) can be written as

$$I_W(s) = \begin{cases} s_0 - s + I_{\mathbb{Q}}(s_0) & \text{if } s < s_0, \\ I_{\mathbb{Q}}(s) & \text{if } s \geq s_0, \end{cases} \quad \text{and} \quad q_W(\alpha) = \begin{cases} q_{\mathbb{Q}}(-1) & \text{if } \alpha < -1, \\ q_{\mathbb{Q}}(\alpha) & \text{if } \alpha \geq -1, \end{cases} \tag{1.21}$$

where s_0 is any point such that -1 belongs to the subdifferential of $I_{\mathbb{Q}}$ at s_0 , or equivalently, such that s_0 belongs to the subdifferential of $q_{\mathbb{Q}}$ at -1 . In nontrivial cases, q_W is not differentiable at $\alpha = -1$. The situation is depicted on the basis of an example in Fig. 1.

The next two theorems involve only one measure \mathbb{P} . We remark that the pair (\mathbb{P}, \mathbb{P}) is admissible if and only if \mathbb{P} satisfies both **UD** and **SLD**. As a consequence, the conclusions of Theorem 0 hold under the assumptions of Theorems B and C; in particular $I_{\mathbb{P}}$ and $q_{\mathbb{P}}$ are well defined, and the numbers $\gamma_+ \in (-\infty, 0]$ and $\gamma_- \in [-\infty, 0]$ defined in (1.12) are actual limits.

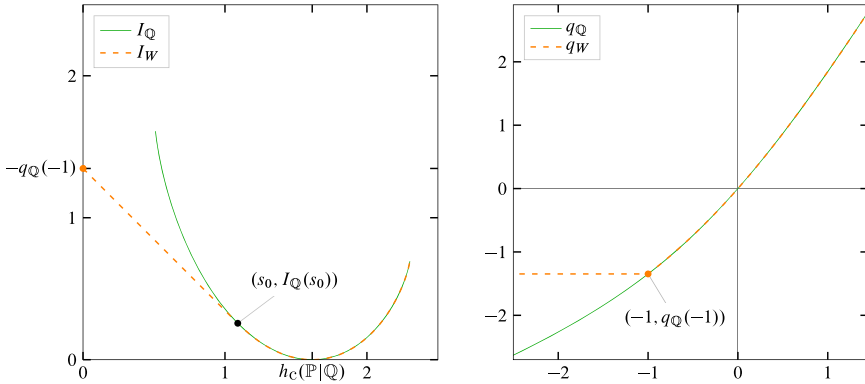


Fig. 1. The rate functions and pressures of Theorems 0 and A for some Bernoulli measures \mathbb{P} and \mathbb{Q} ; see Sect. 2.1. The rate functions are infinite wherever not drawn

Theorem B. (LDP for $\frac{1}{n} \ln V_n$) If $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ satisfies **UD** and **SLD**, then the following hold:

i. The sequence $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ is exponentially tight and satisfies the LDP with respect to \mathbb{P} with the good, convex rate function I_V given by

$$I_V(s) := \begin{cases} \infty & \text{if } s < 0, \\ \inf_{r \geq s} (r - s + I_{\mathbb{P}}(r)) & \text{if } s \geq 0. \end{cases} \tag{1.22}$$

ii. For all $\alpha \in \mathbb{R}$, the limit in (1.5) exists in $(0, \infty]$,

$$q_V(\alpha) = \max\{q_{\mathbb{P}}(\alpha), q_{\mathbb{P}}(-1)\}, \tag{1.23}$$

and the Legendre–Fenchel duality relations $q_V = I_V^*$ and $I_V = q_V^*$ hold. Moreover,

$$I_V(0) = -q_{\mathbb{P}}(-1) \geq -\gamma_+. \tag{1.24}$$

Remark 1.11. The rate function I_V corresponds to I_W in the special case $\mathbb{Q} = \mathbb{P}$. While one can, of course, choose $\mathbb{Q} = \mathbb{P}$ in Theorem A (in this case it suffices that \mathbb{P} satisfies **UD** and **SLD**), this special case is not equivalent to Theorem B. Indeed, W_n and V_n are still distinct in their definition and underlying probability space. It is known that the range of applicability of almost sure entropy estimation via W_n is strictly smaller than that via V_n (or R_n); see [OW93] and [Shi93, §4].

Theorem C. (LDP for $\frac{1}{n} \ln R_n$) If $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ satisfies **UD**, **SLD** and **PA**, then the following hold:

i. The sequence $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ is exponentially tight and satisfies the LDP with respect to \mathbb{P} with the good (possibly nonconvex) rate function I_R given by

$$I_R(s) := \begin{cases} \infty & \text{if } s < 0, \\ -\gamma_+ & \text{if } s = 0, \\ \inf_{r \geq s} (r - s + I_{\mathbb{P}}(r)) & \text{if } s > 0. \end{cases} \tag{1.25}$$

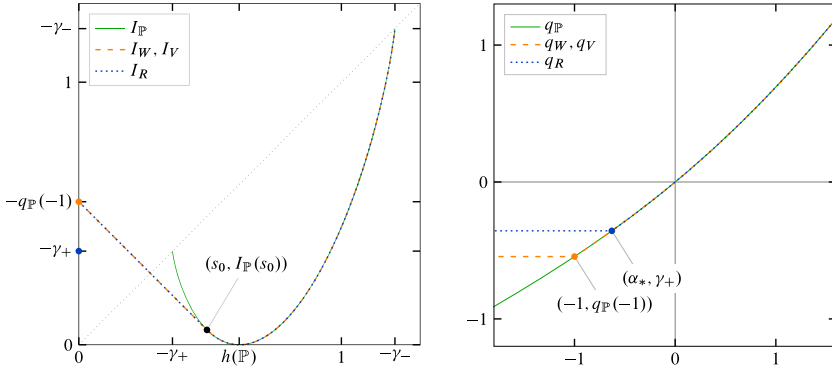


Fig. 2. The rate functions and pressures of Theorems 0 and A–C for some Bernoulli measure \mathbb{P} ; see Sect. 2.1. The rate functions are infinite wherever not drawn

ii. For all $\alpha \in \mathbb{R}$, the limit in (1.4) exists in $(0, \infty]$ and

$$q_R(\alpha) = \max\{q_{\mathbb{P}}(\alpha), \gamma_+\}. \tag{1.26}$$

Moreover, $q_R = I_R^*$.

iii. We have the following relations: I_R is convex $\iff I_R = I_V \iff q_R = q_V \iff \gamma_+ = q_{\mathbb{P}}(-1) \iff I_{\mathbb{P}}(-\gamma_+) = 0 \iff q_{\mathbb{P}}(\alpha) = -\gamma_+\alpha$ for all $\alpha \leq 0 \iff \gamma_+ = \gamma_- \iff I_{\mathbb{P}}(s) = \infty$ for all $s \in \mathbb{R} \setminus \{-\gamma_+\} \iff h(\mathbb{P}) = h_{\text{top}}(\text{supp } \mathbb{P})$. Moreover, if $\tau_n = O(1)$ and $C_n = O(1)$, then all the above properties are actually equivalent.

By their definition, the rate functions I_R and I_V may differ only at 0, and by (1.24) we have $I_R(0) = -\gamma_+ \leq I_V(0)$; see Fig. 2 for an illustration. Part iii gives many technical conditions equivalent to $I_R = I_V$, the most notable of which is the convexity of I_R . The situation where $I_R = I_V$ should be seen as quite degenerate; in the generic case, we have $I_R(0) < I_V(0)$, as strikingly illustrated by the examples in Sects. 2.1 and 2.2. The generic inequality $I_R(0) < I_V(0)$ is due to fact that the definition of R_n allows for overlaps (characterized by $R_n < n$) which are excluded in V_n . These overlaps, which make the LDP for $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ highly interesting, were widely studied in other contexts; see the discussion and references in Sect. 1.3. With help of PA, we will show in Sect. 4.2 that $\mathbb{P}\{x : R_n(x) < n\}$ asymptotically decreases like $e^{n\gamma_+}$, which explains the equality $I_R(0) = -\gamma_+$.

Remark 1.12. If $\gamma_+ = 0$, then $I_{\mathbb{P}}(0) = 0$ by (1.15), and thus $q_{\mathbb{P}}(\alpha) = I_{\mathbb{P}}^*(\alpha) = 0$ for all $\alpha \leq 0$. In this case, we easily see that $I_V = I_R = I_{\mathbb{P}}$ and $q_V = q_R = q_{\mathbb{P}}$.

Remark 1.13. The relation (1.26) can be written as

$$q_R(\alpha) = \begin{cases} \gamma_+ & \text{if } \alpha < \alpha_*, \\ q_{\mathbb{P}}(\alpha) & \text{if } \alpha \geq \alpha_*, \end{cases} \tag{1.27}$$

where $\alpha_* \leq 0$ is such that $q_{\mathbb{P}}(\alpha_*) = \gamma_+$. If $\gamma_+ = 0$, then we can take any $\alpha_* \leq 0$ in view of Remark 1.12. If $\gamma_+ < 0$, then $q_{\mathbb{P}}(-1) \leq \gamma_+ < 0 = q_{\mathbb{P}}(0)$ by (1.24), so the point α_* is unique and satisfies $\alpha_* \in [-1, 0)$, with $\alpha_* = -1$ if and only if $I_V = I_R$.

1.3. Outline of the proof for a toy model. We present a natural toy model, consisting of a mixture of geometric distributions, which not only provides the correct rate functions for W_n and V_n , but also illustrates the method of our proof. We then modify the toy model in order to take overlaps (corresponding to $R_n < n$) into account and guess the correct rate function for R_n .

Approximations of waiting times and return times by geometric random variables (or exponential random variables in the scaling limit) have been widely studied, typically under mixing assumptions; see e.g. [GS97, CGS99, HSV99, AG01, HV10] and [AAG21] for a recent overview and more exhaustive references. As often emphasized in the literature, possible overlaps, which are tightly related to Poincaré recurrence times (see [AV08, AC15, AAG21]), play a crucial role in such approximations. We shall further comment on Poincaré recurrence times at the end of Sect. 4.2.

Our decoupling assumptions do not seem to imply any of the very sharp exponential approximations that are available in the literature. In comparison, the geometric approximation that we prove in Sect. 3.1 are quite loose in the sense that scaling factors and error terms may grow subexponentially. Yet, it suffices to establish the LDPs of interest.

Geometric approximation. We start with the following interpretation of $W_n(x, y)$: first x is drawn at random according to the law \mathbb{P} and y is drawn at random according to the law \mathbb{Q} , independently of x . Then, for each $k \in \mathbb{N}$, we check whether $y_k^{k+n-1} = x_1^n$ or not. Once x_1^n is given, shift invariance implies that $\mathbb{Q}\{y \in \Omega : y_k^{k+n-1} = x_1^n\} = \mathbb{Q}_n(x_1^n)$ for each $k \in \mathbb{N}$, and we can view W_n as the time of the first “success” in a series of attempts (indexed by k). If the attempts were mutually independent, W_n would be a geometric random variable with random parameter $p_n = \mathbb{Q}_n(x_1^n)$. Of course, these attempts are not independent: even if \mathbb{Q} is a Bernoulli measure, the attempts k and k' are only independent for $|k' - k| \geq n$. However, it turns out, due to our decoupling assumptions, that the asymptotic behavior of W_n at the scale that is relevant to our LDP is accurately captured by this simplified geometric model.

We now define the toy model properly. Let \tilde{W}_n the random variable whose law ν_n on \mathbb{N} is given by

$$\nu_n(k) := \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{Q}_n(u) (1 - \mathbb{Q}_n(u))^{k-1}, \tag{1.28}$$

for every $k \in \mathbb{N}$. This is a simple mixture of geometric distributions, motivated by the above discussion. The next proposition describes the large deviations of \tilde{W}_n . Since it is introduced only for illustration purposes, and since the actual proof of the proposition relies on estimates which are similar to—but simpler than—those we provide for W_n in the main body of the paper, we limit ourselves to sketching the proof. The interested reader will easily be able to fill in the details.

Proposition 1.14. *If the pair (\mathbb{P}, \mathbb{Q}) is admissible, then the sequence $(\frac{1}{n} \ln \tilde{W}_n)_{n \in \mathbb{N}}$ satisfies the LDP with the rate function I_W defined in (1.19).*

Sketch of the proof. For each $s \in M_n := \{\frac{1}{n} \ln k : k \in \mathbb{N}\}$, the probability that $\frac{1}{n} \ln \tilde{W}_n$ equals s is given by

$$\nu_n(e^{ns}) = \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{Q}_n(u) (1 - \mathbb{Q}_n(u))^{e^{ns}-1} = \int_{(0, \infty)} e^{-nr} (1 - e^{-nr})^{e^{ns}-1} d\mu_n(r), \tag{1.29}$$

where we denote by μ_n the distribution of $-\frac{1}{n} \ln \mathbb{Q}_n$ with respect to \mathbb{P} . Now using that e^{-nr} is very small for n large, a formal first order Taylor expansion yields

$$(1 - e^{-nr})^{e^{ns}-1} \sim \exp(-(e^{ns} - 1)e^{-nr}) \sim \exp(-e^{n(s-r)}). \tag{1.30}$$

We thus have a cut-off phenomenon: the above vanishes superexponentially when $r < s$, and is very close to 1 when $r > s$. Formal substitution into (1.29) yields

$$v_n(e^{ns}) \sim \int_{(s, \infty)} e^{-nr} d\mu_n(r).$$

We remark that this cut-off argument corresponds to retaining, in the sum in (1.28), only the $u \in \mathcal{A}^n$ such that $\mathbb{Q}_n(u) \lesssim e^{-ns}$.

Now, for $s > 0$ and ϵ small, the set $B(s, \epsilon) \cap M_n$ contains approximately $e^{n(s+\epsilon)} - e^{n(s-\epsilon)} \sim e^{ns}$ points, so

$$v_n \left\{ k \in \mathbb{N} : \frac{1}{n} \ln k \in B(s, \epsilon) \right\} \sim e^{ns} v_n(e^{ns}) \sim \int_{(s, \infty)} e^{n(s-r)} d\mu_n(r). \tag{1.31}$$

Formally, Theorem 0 says that $d\mu_n(r) \sim e^{-nI_{\mathbb{Q}}(r)} dr$, and a saddle-point approximation in (1.31) then yields

$$\frac{1}{n} \ln v_n \left\{ k \in \mathbb{N} : \frac{1}{n} \ln k \in B(s, \epsilon) \right\} \sim \sup_{r \geq s} (s - r - I_{\mathbb{Q}}(r)) = -I_W(s).$$

The same conclusion applies to the case $s = 0$ without the need for any cut-off argument. This local formulation of the LDP (i.e. concerning only small balls) implies the weak LDP (see Sect. 4), which can in turn be promoted to a full LDP using the arguments of Sect. 5. □

The fact that we consider $W_n(x, y)$ for x and y that are mutually independent was an important ingredient of the above argument. When moving on to V_n and R_n (and replacing \mathbb{Q} with \mathbb{P}), we look for occurrences of x_1^n , not in an independent sample y , but in x itself, which introduces more dependence. Conditioned on the event $[u]$ for some $u \in \mathcal{A}^n$, the random variable V_n corresponds to the first success time of a series of attempts, where the k -th attempt is successful if $x_{n+k}^{2n+k-1} = u$. Contrary to the case of W_n , the conditional success probability given $[u]$ of the k -th attempt is not simply given by $\mathbb{P}_n(u)$ since the coordinates x_1^n and x_{n+k}^{2n+k-1} are not independent in general.⁴ However, by our decoupling assumptions, and since the intervals $[1, n]$ and $[n+k, 2k+k-1]$ do not overlap, this added dependence will not actually alter the asymptotics. The arguments of Proposition 1.14 then suggest that V_n obeys the LDP with the rate function I_V of (1.22).

Return time and overlaps. The picture for R_n is more complicated: while the dependence between x_{k+1}^{k+n} and x_1^n will not significantly alter our estimates for k very large, this dependence will play a major role when $k < n$, due to overlap; see the examples in Sect. 2. We shall prove in Sect. 4.2 that, under the assumptions of Theorem C,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : R_n(x) < n\} = \gamma_+. \tag{1.32}$$

⁴ They are, for example, if \mathbb{P} is a Bernoulli measure; see Remark 2.1.

This suggests the following picture: with probability close to $e^{n\gamma_+}$ we have “very quick return” ($R_n < n$), and with probability close to $1 - e^{n\gamma_+}$ we start a series of trials as in the case of V_n . As before, let us pretend these trials are independent, and that their probability of success is exactly $\mathbb{P}_n(x_1^n)$. Noting that the asymptotics of $\frac{1}{n} \ln R_n$ is not affected by adding to R_n quantities of order n , we are led to introduce a toy model \tilde{R}_n for R_n whose law ρ_n on \mathbb{N} is given by

$$\rho_n(k) := e^{n\gamma_+} \delta_{k,1} + (1 - e^{n\gamma_+}) \nu_n(k) \tag{1.33}$$

for every $k \in \mathbb{N}$, where ν_n is as in (1.28) with $\mathbb{Q} = \mathbb{P}$. A straightforward adaptation of Proposition 1.14, taking the first term in (1.33) into account at $s = 0$, and also using that $\gamma_+ \geq \sup_{r \geq 0} (-r - I_{\mathbb{P}}(r))$ by Theorem 0.iii.a., shows that \tilde{R}_n satisfies the LDP with the rate function I_R of (1.25).

Turning the sketch into a proof. We now briefly comment on how the above sketch will be made rigorous in the main body of the paper. In Sect. 3.1, we establish that for fixed x , the random variable $W_n(x, \cdot)$ on (Ω, \mathbb{Q}) is indeed well approximated by a geometric random variable at the exponential scale. The above cut-off argument is made rigorous in Proposition 3.5. The corresponding estimates for R_n and V_n are presented in Sect. 3.2.

The saddle-point approximation mentioned above is made rigorous by the variation of Varadhan’s lemma provided in Lemma 4.1, which then leads to the weak LDP for the sequences of interest. As with the above toy model, we need to treat the cases $s > 0$ (Sect. 4.1) and $s = 0$ (Sect. 4.2) separately—the latter is particularly subtle for R_n and PA will be needed to establish (1.32). With the weak LDPs at hand, the main results are proved in Sect. 5.

2. Examples

Decoupling assumptions are satisfied by many important classes of examples, and have allowed to simplify and unify the proofs of various large deviation principles that existed in many, sometimes rather technical, forms in the literature. The range of applicability of UD and SLD has already been discussed in [CJPS19,BCJP21,CDEJR23a,CDEJR23b].

We describe in this section some classes of examples that serve as illustrations of different features of our main results. Sections 2.1–2.5 each assume some level of familiarity with the specific examples on the reader’s part, and may be skipped entirely without affecting the reader’s ability to understand the proofs of the main theorems. We now briefly summarize the role of each of these examples.

1. In the Bernoulli (IID) case (Sect. 2.1), formulae for the different pressures and rate functions can be quickly derived, and are easy to understand. To the best of our knowledge, the global aspect (i.e. without any restriction to a strict subinterval of \mathbb{R}) of the LDPs in Theorems A–C as well as the ability to consider distinct measures \mathbb{P} and \mathbb{Q} in Theorem A are new even in this most basic class of examples.
2. By going from Bernoulli measures to Markov measures (Sect. 2.2), we start seeing the benefit of inserting the words ξ in the formulation of SLD, as they allow to deal with irreducible Markov chains whose transition probabilities are not all positive. Markov measures also make the role of periodicity in PA very clear.
3. Discussing our assumptions and results in the setup of equilibrium measures for potentials enjoying Bowen’s regularity condition (Sect. 2.3) allows us to compare our results to existing ones, most notably to those of [AACG23] where the pressures q_W and q_R were studied.

4. We then discuss two situations in which Bowen’s regularity condition can be lifted. While carrying distinct history and intuition, they both reveal the same two aspects of our decoupling assumptions. First, they make a case that allowing a certain amount of growth for the sequence $(C_n)_{n \in \mathbb{N}}$ in **UD** and **SLD** is beneficial. Second, they show that our assumptions apply in phase-transition situations, and in particular do not imply ergodicity. These generalizations are:
 - i. equilibrium measures for absolutely summable interactions in statistical mechanics (Sect. 2.4.1);
 - ii. equilibrium measures for g -functions, i.e. g -measures (Sect. 2.4.2).
 While our results do not seem to apply to the class of *weak Gibbs* measures in full generality (see Sect. 2.4), the measures discussed in Sects. 2.4.1 and 2.4.2 are weak Gibbs.
5. Finally, the so-called class of *hidden Markov models* (Sect. 2.5) shows that our assumptions apply to measures which are far from Gibbsian. Hidden Markov models also provide examples of pairs of measures (\mathbb{P}, \mathbb{Q}) where the distributions of $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ lack exponential tightness, showing that I_W (see Theorem A) need not be a good rate function when $\mathbb{Q} \neq \mathbb{P}$. To the authors’ knowledge, even the conclusions of Theorem 0 in this setup are new.

As abundantly discussed in [BJPP18, CJPS19, BCJP21, CDEJR23a, CDEJR23b], repeated quantum measurement processes give rise to a very rich class of measures satisfying our decoupling assumptions, yet displaying remarkable singularities — some being, once again, far from Gibbsian. For reasons of space, we do not repeat such a discussion in the present paper.

2.1. Bernoulli measures. Consider the simple case $\mathbb{P} = \mathbb{P}_1^{\otimes \mathbb{N}}$ and $\mathbb{Q} = \mathbb{Q}_1^{\otimes \mathbb{N}}$, where \mathbb{P}_1 and \mathbb{Q}_1 are measures on \mathcal{A} , with $\mathbb{P}_1 \ll \mathbb{Q}_1$. **SLD**, **UD** and **PA** obviously hold with $C_n = 1$ and $\tau_n = 0$ for all n . Note that, as a random variable on (Ω, \mathbb{P}) , the map $x \mapsto -\frac{1}{n} \ln \mathbb{Q}_n(x^n) = \frac{1}{n} \sum_{i=1}^n (-\ln \mathbb{Q}_1(x_i))$ is simply the average of IID random variables supported on the finite set $\{-\ln \mathbb{Q}_1(a) : a \in \text{supp } \mathbb{P}_1\} \subset \mathbb{R}$. By independence, $q_{\mathbb{Q}}(\alpha)$ is easily seen to coincide with the cumulant-generating function:

$$q_{\mathbb{Q}}(\alpha) = \ln \int_{\mathcal{A}} e^{-\alpha \ln \mathbb{Q}_1(a)} d\mathbb{P}_1(a) = \ln \sum_{a \in \text{supp } \mathbb{P}_1} \mathbb{P}_1(a) \mathbb{Q}_1(a)^{-\alpha}. \tag{2.1}$$

The LDP proved in Theorem 0 then follows from standard results. We mention two methods which lead to different expressions of the rate function $I_{\mathbb{Q}}$.

Method 1. Since $\ln \mathbb{Q}_n(x^n)$ is a sum of IID random variables, Cramér’s theorem [DZ09, §2.2.1] yields the stated LDP with a rate function given by the Legendre–Fenchel transform of the cumulant-generating function (2.1), i.e. with a rate function $I_{\mathbb{Q}} = q_{\mathbb{Q}}^*$. This can be seen as a special case of the Gärtner–Ellis theorem [DZ09, §2.3], which applies here since by (2.1) the function $q_{\mathbb{Q}}$ is differentiable (and actually real-analytic).

Method 2. One can instead appeal to a combination of Sanov’s theorem and the contraction principle [DZ09, §2.1.1–2.1.2] to obtain the LDP with rate function

$$I_{\mathbb{Q}}(s) = \inf_{\mu \in L_s} H_r(\mu | \mathbb{P}_1), \tag{2.2}$$

where H_r denotes the relative entropy and L_s is the set of probability measures $\mu \ll \mathbb{P}_1$ (hence also satisfying $\mu \ll \mathbb{Q}_1$) on \mathcal{A} subject to the constraint

$$-\sum_{a \in \mathcal{A}} \mu(a) \ln \mathbb{Q}_1(a) = s. \tag{2.3}$$

When $s \notin [-\ln \max_{a \in \mathcal{A}} \mathbb{Q}_1(a), -\ln \min_{a \in \text{supp } \mathbb{Q}_1} \mathbb{Q}_1(a)]$, the set L_s is empty and the infimum is set to ∞ by convention. Note that $I_{\mathbb{Q}}$ vanishes at the unique point $s = -\sum_{a \in \mathcal{A}} \mathbb{P}_1(a) \ln \mathbb{Q}_1(a) = h_c(\mathbb{P}|\mathbb{Q})$, where the infimum in (2.2) is attained at $\mu = \mathbb{P}_1$.

We now turn to Theorems B and C, and we discuss in particular the convexity of I_R . The relevant quantities are easily expressed in terms of \mathbb{P}_1 :

$$\begin{aligned} \gamma_+ &= \ln \max_{a \in \text{supp } \mathbb{P}_1} \mathbb{P}_1(a), & h_{\text{top}}(\text{supp } \mathbb{P}) &= \ln |\text{supp } \mathbb{P}_1|, \\ \gamma_- &= \ln \min_{a \in \text{supp } \mathbb{P}_1} \mathbb{P}_1(a), & q_{\mathbb{P}}(-1) &= \ln \sum_{a \in \text{supp } \mathbb{P}_1} \mathbb{P}_1(a)^2. \end{aligned}$$

In view of these expressions, $\gamma_- \leq q_{\mathbb{P}}(-1) \leq \gamma_+$ with strict inequalities *unless* \mathbb{P}_1 is constant on its support. To see this, notice that $\sum_{a \in \mathcal{A}} \mathbb{P}_1(a)^2$ is the expectation of the function $a \mapsto \mathbb{P}_1(a)$ with respect to the measure \mathbb{P}_1 on \mathcal{A} . To discuss the convexity of I_R we distinguish two cases.

Singular case. If \mathbb{P}_1 is constant on its support, then $\gamma_- = q_{\mathbb{P}}(-1) = \gamma_+$, so I_R is convex by Theorem C.iii. Moreover, we readily obtain $h(\mathbb{P}) = h_{\text{top}}(\text{supp } \mathbb{P})$, so \mathbb{P} is indeed the measure of maximal entropy on its support, in accordance with Theorem C.iii. Next, since $-\frac{1}{n} \ln \mathbb{P}_n(x_1^n) = h(\mathbb{P})$ almost surely, the rate function $I_{\mathbb{P}}$ vanishes at $h(\mathbb{P})$ and is infinite everywhere else. Dual to this, $q_{\mathbb{P}}(\alpha) = h(\mathbb{P})\alpha$ for all $\alpha \in \mathbb{R}$. In the quite extreme case where $|\text{supp } \mathbb{P}_1| = 1$, i.e. if \mathbb{P} is a Dirac measure on an orbit of period 1, we find $h(\mathbb{P}) = 0$ and $q_{\mathbb{P}}(\alpha) = 0$ for all $\alpha \in \mathbb{R}$.

Generic case. If \mathbb{P}_1 is not constant on its support, then $\gamma_- < q_{\mathbb{P}}(-1) < \gamma_+$, and thus the rate function I_R is nonconvex by Theorem C.iii. In the present setup, it is easy to understand why $I_R(0) < I_V(0)$, as we now discuss. Let ϵ be small and n be large. On the one hand, if $\hat{a} \in \mathcal{A}$ is such that $\ln \mathbb{P}_1(\hat{a}) = \gamma_+$, we find $R_n(x) = 1$ for all $x \in [\hat{a}^{n+1}]$, so

$$\begin{aligned} \mathbb{P}\{x : R_n(x) \leq e^{\epsilon n}\} &\geq \mathbb{P}\{x : R_n(x) = 1\} \\ &\geq \mathbb{P}_{n+1}(\hat{a}^{n+1}) \\ &= \exp((n+1)\gamma_+). \end{aligned} \tag{2.4}$$

On the other hand, $V_n(x) = k$ implies that $x \in [u] \cap T^{-n-k+1}[u]$ for some $u \in \mathcal{A}^n$, so

$$\begin{aligned} \mathbb{P}\{x : V_n(x) = k\} &\leq \sum_{u \in \mathcal{A}^n} \mathbb{P}([u] \cap T^{-n-k+1}[u]) = \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u)^2 \\ &= \left(\sum_{a \in \mathcal{A}} \mathbb{P}_1(a)^2\right)^n = \exp(nq_{\mathbb{P}}(-1)), \end{aligned}$$

where we have used (2.1) with $\mathbb{Q} = \mathbb{P}$. Thus,

$$\mathbb{P}\{x : V_n(x) \leq e^{\epsilon n}\} \leq \sum_{k=1}^{\lfloor e^{\epsilon n} \rfloor} \mathbb{P}\{x : V_n(x) = k\} \leq \exp(n\epsilon + nq_{\mathbb{P}}(-1)). \tag{2.5}$$

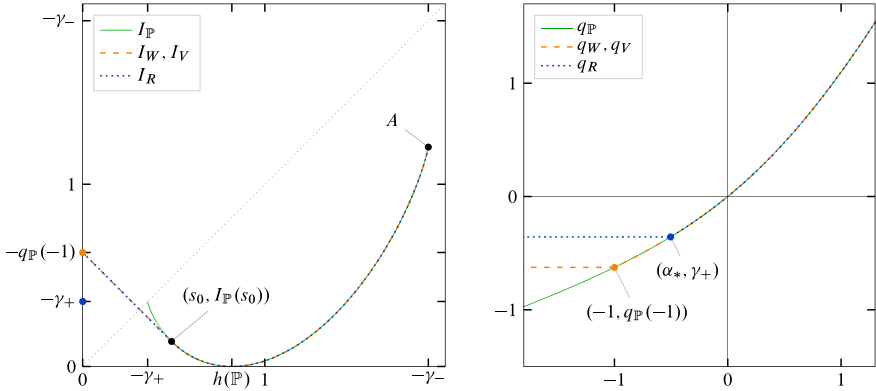


Fig. 3. Rate functions and pressures for a Bernoulli measure with degenerate least probable letter. The coordinates of the point A are $(-\gamma_-, -\gamma_- - \ln 2)$. Rate functions are infinite wherever not drawn

Therefore, for ϵ small enough, the right-hand side of (2.4) decays exponentially faster than the right-hand side of (2.5) since $q_{\mathbb{P}}(-1) < \gamma_+$. The estimates (2.4) and (2.5), despite the rather crude inequalities in (2.4), turn out to be sharp at the exponential scale. Indeed, we find $I_V(0) = -q_{\mathbb{P}}(-1)$ and $I_R(0) = -\gamma_+$ in Theorems B and C.

Remark 2.1. For Bernoulli measures, in the case $\mathbb{Q} = \mathbb{P}$, W_n and V_n have the same law.

We provide three figures corresponding to Bernoulli measures, the first two of which were displayed in Sect. 1.2:

- Fig. 1: $\mathcal{A} = \{0, 1, 2\}$ and $\mathbb{P}_1 = 0.2 \delta_0 + 0.3 \delta_1 + 0.5 \delta_2$ and $\mathbb{Q}_1 = 0.6 \delta_0 + 0.3 \delta_1 + 0.1 \delta_2$.
- Fig. 2: $\mathcal{A} = \{0, 1\}$ and $\mathbb{P}_1 = \mathbb{Q}_1 = 0.3 \delta_0 + 0.7 \delta_1$.
- Fig. 3: $\mathcal{A} = \{0, 1, 2\}$ and $\mathbb{P}_1 = \mathbb{Q}_1 = 0.15 \delta_0 + 0.15 \delta_1 + 0.7 \delta_2$. Notice that $\ln \mathbb{P}_1(0) = \ln \mathbb{P}_1(1) = -\gamma_-$. Such degeneracy implies that the second term in (1.17) takes the value $\ln 2$ for $s = -\gamma_-$. Indeed, when ϵ is very small, the cardinality in (1.17) is essentially 2^n . As a consequence, $I_{\mathbb{P}}(-\gamma_-) = -\gamma_- - \ln 2$. A similar phenomenon occurs at $-\gamma_+$ if several letters in \mathcal{A} have maximum probability.

2.2. Irreducible Markov measures. Let \mathbb{P} and \mathbb{Q} be two stationary Markov measures on \mathcal{A} , with transition matrices $P = [P_{i,j}]_{i,j \in \mathcal{A}}$ and $Q = [Q_{i,j}]_{i,j \in \mathcal{A}}$ respectively. We assume that the matrices P and Q are *irreducible* in the sense that there exists $M \in \mathbb{N}$ such that all entries in the matrices $\sum_{i=1}^M P^i$ and $\sum_{i=1}^M Q^i$ are positive. Then,

$$\mathbb{P}_n(x_1^n) = \mathbb{P}_1(x_1) \prod_{k=1}^{n-1} P_{x_k, x_{k+1}} \quad \text{and} \quad \mathbb{Q}_n(x_1^n) = \mathbb{Q}_1(x_1) \prod_{k=1}^{n-1} Q_{x_k, x_{k+1}}, \quad (2.6)$$

where \mathbb{P}_1 and \mathbb{Q}_1 are the (unique and fully supported) invariant probability vectors for the matrices P and Q respectively. We assume, furthermore, that $Q_{i,j} = 0$ implies $P_{i,j} = 0$, which ensures that $\mathbb{P}_n \ll \mathbb{Q}_n$ for all $n \in \mathbb{N}$. It follows from Lemma A.3 in [CJPS19] that **JSLD** and **UD** hold with $C_n = O(1)$ and $\tau_n = M - 1$, so the pair (\mathbb{P}, \mathbb{Q}) is admissible. Similar arguments also yield **PA**. It is worth noting that if P is further

assumed to be *aperiodic*, i.e. if there exists $M \in \mathbb{N}$ such that all entries in the matrix P^M are positive, then we can fix $\ell = \tau_n = M - 1$ in (1.10), whereas if P is merely irreducible, then ℓ must be allowed depend on u and v . This illustrates the importance of the condition $\ell \leq \tau_n$ in SLD instead of $\ell = \tau_n$; see the discussion in [CJPS19, §§2.5;A.1].

Our results thus apply to the setup described here. As in the previous example, the conclusions of Theorem 0 can be derived using classical methods, which yield natural expressions for $I_{\mathbb{Q}}$ and $q_{\mathbb{Q}}$, as we now show.

Method 1. In view of (2.6), the pressure is easily seen to be given by

$$q_{\mathbb{Q}}(\alpha) = \ln \operatorname{spr}(M(\alpha)),$$

where the spectral radius is computed for the deformation of the stochastic matrix P defined by $M_{i,j}(\alpha) := P_{i,j} Q_{i,j}^{-\alpha}$. By the Perron–Frobenius theorem and analytic perturbation theory, $\alpha \mapsto q_{\mathbb{Q}}(\alpha)$ is real-analytic, so the LDP for $(-\frac{1}{n} \ln \mathbb{Q}_n)_{n \in \mathbb{N}}$ with the rate function $I_{\mathbb{Q}} = q_{\mathbb{Q}}^*$ follows from the Gärtner–Ellis theorem [DZ09, §2.3].

Method 2. From (2.6), we see that

$$\ln \mathbb{Q}_n(x_1^n) = \sum_{k=1}^{n-1} \ln Q_{x_k, x_{k+1}} + O(1)$$

for all $x \in \operatorname{supp} \mathbb{P} \subseteq \operatorname{supp} \mathbb{Q}$, so the LDP for $(-\frac{1}{n} \ln \mathbb{Q}_n)_{n \in \mathbb{N}}$ reduces to that of a sequence of Birkhoff sums. By Sanov’s theorem applied to the pair empirical measures $\mathcal{E}_n(x) := \frac{1}{n} \sum_{k=1}^n \delta_{(x_k, x_{k+1})} \in \mathcal{P}(\mathcal{A}^2)$ and the contraction principle, one derives the expression

$$I_{\mathbb{Q}}(s) = \inf_{\mu \in L_s^{(2)}} H_{\mathbb{T}}(\mu | \mu_1 \otimes P), \tag{2.7}$$

where $L_s^{(2)}$ is the set of probability measures μ on $\mathcal{A} \times \mathcal{A}$ such that $\mu \ll \mu_1 \otimes P$, $\mu_1 = \mu_2$, and

$$\sum_{i,j \in \mathcal{A}} \mu(i, j) \ln Q_{i,j} = -s;$$

see [DZ09, §3.1.3]⁵ or Lemma 4.49 in [DS89]. Here, μ_1 (resp. μ_2) denotes the first (resp. second) marginal of μ , and $\mu_1 \otimes P$ denotes the measure on $\mathcal{A} \times \mathcal{A}$ defined by $(\mu_1 \otimes P)(i, j) = \mu_1(i) P_{i,j}$. Note that $I_{\mathbb{Q}}$ vanishes at the single point $s = -\sum_{i,j \in \mathcal{A}} \mathbb{P}_2(i, j) \ln Q_{i,j} = h_{\mathbb{C}}(\mathbb{P} | \mathbb{Q})$, where the infimum in (2.7) is attained at $\mu = \mathbb{P}_2$.

We now discuss I_R and its convexity, with $\mathbb{P} = \mathbb{Q}$ a Markov measure as above. It is well known that

$$\gamma_+ = \max_{p \leq |\mathcal{A}|} \max_{u \in \mathcal{A}^p} \frac{1}{p} \ln \prod_{k=1}^p P_{u_k, u_{k+1}}, \tag{2.8}$$

with cyclic identification $u_{p+1} = u_1$; see e.g. Remark 1.ii in [Szp93] or [AACG23, §3.4]. The heuristic interpretation is the following: while in the generic IID case discussed in Sect. 2.1 the probability $\mathbb{P}\{x : R_n(x) \leq e^{\epsilon n}\}$ was, at exponential scale, captured by the

⁵ There it is assumed that $P_{i,j} > 0$ for all $i, j \in \mathcal{A}$.

Table 1. A summary of the notational changes between our results and those of [AACG23]

	Pressures		Variable		Important points		
Present paper	$q_{\mathbb{P}}$	q_W	q_R	α	α_*	γ_+	γ_-
[AACG23]	\mathcal{M}_φ	\mathcal{W}_φ	\mathcal{R}_φ	q	q_φ^*	γ_φ^+	$-v_\varphi^+$

subset $\{x : x_1^{n+1} = \hat{a}\hat{a}\hat{a} \cdots \hat{a}\hat{a}\}$ of the event $\{x : R_n(x) = 1\}$ for n large and ϵ small, now $\mathbb{P}\{x : R_n(x) \leq e^{\epsilon n}\}$ is essentially accounted for by the subset $\{x : x_1^{n+p} = uuu \cdots uu^r\}$ of the event $\{x : R_n(x) = p\}$, for any p and u that saturate (2.8), and $r = n - \lfloor \frac{n}{p} \rfloor p$. In other words, the key scenario for small return times consists of a periodic orbit repeating some optimal cycle.

By the last part of Theorem C, the function I_R is convex if and only if \mathbb{P} is the measure of maximal entropy on $\text{supp } \mathbb{P}$, that is if and only if \mathbb{P} is the Parry measure of $\text{supp } \mathbb{P}$, which is characterized by the Perron–Frobenius data of the adjacency matrix of the chain [Par64]; see also [CGS99, §2]. In particular, if $P_{i,j} > 0$ for all $i, j \in \mathcal{A}$, then I_R is convex if and only if $\mathbb{P} = \pi^{\otimes \mathbb{N}}$ with π uniform on \mathcal{A} .

Remark 2.2. One can show that irreducible multi-step Markov measures also satisfy the assumptions in the present paper. In fact, Markov measures and multi-step Markov measures are merely special cases of the equilibrium measures discussed in Sect. 2.3. For example, in the notation of Sect. 2.3, one takes $\varphi(x) = \ln P_{x_1, x_2}$ in the case of Markov measures. An explicit computation of $q_{\mathbb{P}}$ for a specific Markov chain is provided in [AACG23].

2.3. Equilibrium measures for Bowen potentials. As mentioned above, our assumptions cover the setups of [CGS99] and [AACG23], where the large-deviation results are local in the sense that there is some strict subinterval $A \subset \mathbb{R}$ such that the large-deviation lower bound and upper bound (see (4.1) and (4.2) below) are only shown to be valid respectively for open sets O contained in A and closed sets Γ contained in A . The former work considers a single measure \mathbb{P} that is the equilibrium measure for a Hölder-continuous potential on a topologically mixing Markov subshift; the latter work, for a potential of summable variations on the full shift.

In particular, Theorem A.i and Theorem C.i prove the following conjecture stated in [AACG23, §3.3]:

We believe that there exists a non-trivial rate function describing the large deviation asymptotic [on the whole real line] for both return and waiting times, but this has to be proven using another method.

Moreover, the results of [AACG23, §3.2] are recovered as special cases of Theorem A.ii and Theorem C.ii. In order to facilitate the translation, we compare notations in Table 1. Note also that what we refer to as pressure is called “ L^q -spectrum” in [AACG23].

We first consider the class of *Bowen-regular* potentials on the full shift, and then discuss the extension to some subshifts (including those of [CGS99]) in Remark 2.3. We recall that a potential φ , i.e. a continuous function $\varphi : \Omega \rightarrow \mathbb{R}$, is called Bowen regular if

$$\sup \left\{ \left| \sum_{i=0}^{n-1} \varphi(T^i x) - \sum_{i=0}^{n-1} \varphi(T^i y) \right| : x \in \Omega, n \in \mathbb{N}, y \in [x_1^n] \right\} < \infty, \quad (2.9)$$

and that this class strictly contains the class of potentials with summable variations, which in turn strictly contains the class of Hölder-continuous potentials; we refer the reader to [Bow74] and [Wal01, §4] for a thorough discussion. We also recall that the topological pressure of any potential φ is given by

$$p_{\text{top}}(\varphi) := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \sum_{u \in \mathcal{A}^n} e^{\sup_{x \in [u]} \sum_{i=0}^{n-1} \varphi(T^i x)}. \tag{2.10}$$

Suppose that \mathbb{P} and \mathbb{Q} are the (necessarily unique [Bow74]) equilibrium measures for the Bowen-regular potentials φ and ψ on Ω in the sense that they belong to $\mathcal{P}_{\text{inv}}(\Omega)$ and satisfy

$$h(\mathbb{P}) + \int \varphi d\mathbb{P} = p_{\text{top}}(\varphi) \quad \text{and} \quad h(\mathbb{Q}) + \int \psi d\mathbb{Q} = p_{\text{top}}(\psi), \tag{2.11}$$

respectively. The measure \mathbb{P} then satisfies the *Bowen–Gibbs property* with respect to φ , i.e. there exists a constant $K \geq 1$ such that

$$K^{-1} e^{\sum_{i=0}^{n-1} \varphi(T^i x) - np_{\text{top}}(\varphi)} \leq \mathbb{P}([x_1^n]) \leq K e^{\sum_{i=0}^{n-1} \varphi(T^i x) - np_{\text{top}}(\varphi)}, \tag{2.12}$$

for every $x \in \Omega$ [Wal01, §4], and one then deduces that it satisfies **UD** and **SLD** with $\tau_n = 0$ and $C_n = K^3$. The same is true for \mathbb{Q} with ψ , so **JSLD** and **PA** follow from Remarks 1.2 and 1.5 respectively, and thus the pair (\mathbb{P}, \mathbb{Q}) is admissible. We note that (2.12) is central in the analysis in [AACG23, §2.1]. In order to simplify some formulae, we assume for the remainder of this subsection that

$$p_{\text{top}}(\varphi) = p_{\text{top}}(\psi) = 0, \tag{2.13}$$

which results in no loss of generality since adding constants to φ and ψ does not alter the set of equilibrium measures.

In the setup of the present subsection, the conclusions of Theorem 0 are well known and can be obtained more directly as follows.

Method 1. The bounds (2.12) imply that

$$q_{\mathbb{Q}}(\alpha) = p_{\text{top}}(\varphi - \alpha\psi). \tag{2.14}$$

In particular, we remark that $q_{\mathbb{P}}(-1) = p_{\text{top}}(2\varphi)$; this quantity plays an important role in the formula for q_W , and is equal to $-I_W(0)$. In this setup, $q_{\mathbb{Q}}$ is differentiable and, for all $\alpha \in \mathbb{R}$,

$$q'_{\mathbb{Q}}(\alpha) = - \int \psi d\mu_{\alpha},$$

where μ_{α} is the equilibrium measure for the Bowen-regular potential $\varphi - \alpha\psi$; see e.g. Theorems 4.3.3 and 4.3.5 in [Kel98]. The LDP of Theorem 0 then follows from the Gärtner–Ellis theorem.

Method 2. The same LDP can be obtained by noticing that, in view of (2.12), the large deviations of $(-\frac{1}{n} \ln \mathbb{Q}_n)_{n \in \mathbb{N}}$ are the same as those of the ergodic averages of $-\psi$ with respect to \mathbb{P} , which is a well-studied problem; see e.g. [You90, Kif90, Com09, PS17]. The rate function is then given, for all $s \in \mathbb{R}$, by

$$I_{\mathbb{Q}}(s) = - \sup \left\{ \int \varphi d\eta + h(\eta) : \eta \in \mathcal{P}_{\text{inv}}, \int \psi d\eta = -s \right\}. \tag{2.15}$$

The rate function $I_{\mathbb{Q}}$ vanishes at a single point $s = -\int \psi d\mathbb{P} = h_c(\mathbb{P}|\mathbb{Q})$; the supremum in (2.15) is then reached at $\eta = \mathbb{P}$.

For all $s \geq 0$, the relations (1.19) and (1.22) give

$$I_W(s) = -s - \sup \left\{ \int (\varphi + \psi) d\eta + h(\eta) : \eta \in \mathcal{P}_{\text{inv}}, \int \psi d\eta \leq -s \right\},$$

$$I_V(s) = -s - \sup \left\{ 2 \int \varphi d\eta + h(\eta) : \eta \in \mathcal{P}_{\text{inv}}, \int \varphi d\eta \leq -s \right\}.$$

When $\mathbb{Q} = \mathbb{P}$, we obtain $I_W = I_V$ and, for all $s \in \mathbb{R}$,

$$I_{\mathbb{P}}(s) = s - \sup \left\{ h(\eta) : \eta \in \mathcal{P}_{\text{inv}}, \int \varphi d\eta = -s \right\}.$$

Next, by (1.25), I_R differs from I_V only at 0, with

$$I_R(0) = -\gamma_+ = - \sup_{\eta \in \mathcal{P}_{\text{inv}}} \int \varphi d\eta;$$

the second identity is proved [AACG23, §4.2].

The very last assertion of Theorem C applies, and I_R is convex if and only if \mathbb{P} is the measure of maximal entropy on Ω (note that $\text{supp } \mathbb{P} = \Omega$ by (2.12)), i.e. if \mathbb{P} is the uniform measure. Equivalently, in terms of potentials, I_R is convex if and only if φ is cohomologous to a constant [AACG23, §3.2].

The authors of [AACG23] first derive the expression (1.20) for q_W and the expression (1.23) for q_R . Then, since under the assumptions at hand, the pressure q_W is differentiable on the domain where it is equal to $q_{\mathbb{P}}$, a version of the Gärtner–Ellis theorem implies the LDP for $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ restricted to the interval $[q'_{\mathbb{P}}(-1), -\gamma_-]$, i.e. where $I_{\mathbb{P}}$ and I_W coincide and are finite. In the same way, $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ is shown to obey the LDP restricted to the interval $[q'_{\mathbb{P}}(\alpha_*), -\gamma_-]$, i.e. where $I_{\mathbb{P}}$ and the convex envelope of I_R coincide and are finite; see e.g. Fig. 2. By construction, the approach of [AACG23] cannot describe the LDP on the interval $(0, q'_{\mathbb{P}}(-1))$ (resp. $(0, q'_{\mathbb{P}}(\alpha_*))$), and in particular cannot capture the nonconvexity of I_R , since I_R is not the Legendre–Fenchel transform of q_R in general.

As mentioned in the introduction, our method is very different in spirit. In addition to the explicit singularities in (1.21) and (1.27), the pressures suffer from the fact that even $q_{\mathbb{Q}}$ may fail to be differentiable under our decoupling assumptions. As a consequence, the Gärtner–Ellis theorem cannot be used to obtain the LDP, even limited to the above-mentioned intervals. We are able to circumvent these limitations by going in the opposite direction: we first establish the LDPs directly, using the Ruelle–Lanford method, and then we obtain the properties of the pressures as corollaries.

While [CGS99] does not rely on any version of the Gärtner–Ellis theorem, the LDP there is still restricted to a nonexplicit interval, which is contained in that of [AACG23]. More precisely, the LDP is restricted to an interval of values of s on which $I_{\mathbb{P}}(s)$ is small enough so that some exponentially decaying error terms that arise in the proof decay faster than $e^{-nI_{\mathbb{P}}(s)}$.

Remark 2.3. The above discussion and the results in [AACG23] are limited to the full shift Ω , while [CGS99] discusses Markov subshifts Ω' that are topologically mixing. However, the above conclusions extend in a straightforward way to any subshift $\Omega' \subseteq \Omega$

satisfying the *flexible specification property* of Definition A.2 with $\tau_n = O(1)$, and in particular to any transitive subshift of finite type; see Remark A.3. Indeed, in this context, one can show that if \mathbb{P} and \mathbb{Q} are equilibrium measures for two potentials φ and ψ on Ω' satisfying Bowen’s condition, i.e. (2.9) with x and y restricted to Ω' in the supremum, then the Bowen–Gibbs property (2.12) holds for both measures and for all $x \in \Omega'$; see Remark 2.2 in [CT13] and [CT16, §6.5] (with $\mathcal{G} = \mathcal{L}$ in the notation therein). This in turn implies JSLD and PA through Lemmas A.4, B.7 and B.8, once we have extended \mathbb{P} and \mathbb{Q} to measures on Ω by setting $\mathbb{P}(\Omega \setminus \Omega') = \mathbb{Q}(\Omega \setminus \Omega') = 0$. Thus, our results apply, and the above expressions for the rate functions and pressures remain valid. In particular, I_R is convex if and only if \mathbb{P} is the measure of maximal entropy on Ω' .

We emphasize that the restriction to $\tau_n = O(1)$ is important in two ways in the above argument: first, in order to apply Lemma A.4, and second, in order to derive (2.12), on which the hypotheses (A.3) and (A.4) of Lemmas B.7 and B.8 rely in the present setup.

2.4. Beyond Bowen potentials: statistical mechanics and g -measures. The analysis in Sect. 2.3 relies heavily on the Bowen–Gibbs property (2.12), which is obtained as a consequence of the (quite restrictive) Bowen condition (2.9) imposed on the potentials φ and ψ .

For the coming discussion, we introduce a weaker version of (2.12). We say that \mathbb{P} is *weak Gibbs* for the potential φ if there exists an $e^{o(n)}$ -sequence $(K_n)_{n \in \mathbb{N}}$ such that for all $x \in \Omega$,

$$K_n^{-1} e^{\sum_{i=0}^{n-1} \varphi(T^i x) - n p_{\text{top}}(\varphi)} \leq \mathbb{P}_n(x_1^n) \leq K_n e^{\sum_{i=0}^{n-1} \varphi(T^i x) - n p_{\text{top}}(\varphi)}. \tag{2.16}$$

The notion of weak Gibbs measure was introduced in [Yur02], and in many interesting situations, equilibrium measures for non-Bowen potentials can still be shown to be weak Gibbs, see for example [PS20].

The conclusions of Theorem 0 remain valid if \mathbb{P} and \mathbb{Q} are assumed to be weak Gibbs for φ and ψ respectively; indeed, in view of (2.16), the LDP of Theorem 0 again boils down to the LDP for the ergodic averages of ψ with respect to \mathbb{P} , which is a well-studied problem; see e.g. [Com09, §5], [Var12], [PS17], and [CJPS19, §A.3], as well as [EKW94] in the specific setup of Sect. 2.4.1 and [CO00] in the specific setup of Sect. 2.4.2 below.

However, the weak Gibbs condition does not seem to imply UD and SLD; see Remark B.10. To the best of the authors’ knowledge, the LDP for $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$, $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ and $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ as well as the validity of (1.20), (1.23) and (1.26) are open problems for weak Gibbs measures.

We discuss in Sects. 2.4.1 and 2.4.2 below two important classes of measures enjoying the weak Gibbs property, which are shown to satisfy our decoupling assumptions, using specific arguments distinct from those in Sect. 2.3. Although they have a large intersection, these two classes are distinct; see [FGM11, BEvELN18].

Contrary to the regularity conditions in Sect. 2.3, the setups of Sects. 2.4.1 and 2.4.2 allow for phase transitions: the equilibrium measures may fail to be unique and ergodic. Also, the pressure $q_{\mathbb{Q}}$ may fail to be differentiable, so the conclusions of Theorem 0 cannot be obtained using the Gärtner–Ellis theorem anymore.

2.4.1. Absolutely summable interactions in statistical mechanics An important situation where less regular potentials arise is the statistical mechanics of one-dimensional, translation-invariant systems; see e.g. [Rue04, Ch. 3–5] or [Sim93, Ch. II–III]. Indeed, if

$\Phi := \{\Phi_X\}_{X \in \mathbb{Z}}$ is a translation-invariant collection of functions $\Phi_X : \mathcal{A}^X \rightarrow \mathbb{R}$ (interchangeably considered as functions on $\mathcal{A}^{\mathbb{Z}}$), called *interactions*, satisfying the absolute summability condition

$$\sum_{\substack{X \in \mathbb{Z} \\ 0 \in X}} \|\Phi_X\|_\infty < \infty, \tag{2.17}$$

then there is a well-known correspondence between translation-invariant Gibbs states — either defined using the *Dobrushin–Lanford–Ruelle equations*, or using convex combinations of weak limits of finite-volume Gibbs measures — and equilibrium measure (on $\mathcal{A}^{\mathbb{Z}}$) for the “energy per site” potential

$$\varphi = \sum_{\substack{X \in \mathbb{Z} \\ \min X = 1}} \Phi_X,$$

which we see both as a function on $\mathcal{A}^{\mathbb{Z}}$ and on $\Omega = \mathcal{A}^{\mathbb{N}}$.⁶

Let now \mathbb{P} (resp. \mathbb{Q}) be the marginal on Ω of some translation-invariant Gibbs state on $\mathcal{A}^{\mathbb{Z}}$ for the absolutely summable interaction Φ (resp. Ψ). Equivalently, \mathbb{P} (resp. \mathbb{Q}) is an equilibrium measure for the potential φ (resp. ψ) on Ω .

Since φ and ψ may fail to satisfy Bowen’s regularity condition, we cannot use the Bowen–Gibbs property (2.12). However, the Dobrushin–Lanford–Ruelle equations allow to prove that \mathbb{P} and \mathbb{Q} satisfy **UD** and **SLD** with $\tau_n = 0$, but with a possibly unbounded⁷ sequence $(C_n)_{n \in \mathbb{N}}$; see Lemma 9.2 in [LPS95]. **JSLD** and **PA** then follow from Remarks 1.2 and 1.5. Our main results thus apply, and we now identify some of the quantities at play in physical terms.

The associated topological pressure $p_{\text{top}}(\varphi)$ can be thought of as a free energy density: with

$$U_n := \sum_{X \subseteq [1, n]} \Phi_X$$

the Hamiltonian (up to a factor of minus the inverse temperature) corresponding to Φ in the finite volume $[1, n]$ with free boundary conditions, we have

$$U_n = \sum_{i=0}^{n-1} \varphi \circ T^i + o(n), \tag{2.18}$$

so

$$F(\Phi) := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \sum_{u \in \mathcal{A}^n} e^{U_n(u)} = p_{\text{top}}(\varphi).$$

Using the Dobrushin–Lanford–Ruelle equations, one can show that

$$\mathbb{P}_n(u) = e^{U_n(u) - nF(\Phi) + o(n)}, \tag{2.19}$$

⁶ By our choice of summation along $X \in \mathbb{Z}$ such that $\min X = 1$, the function φ only depends on the positive coordinates of $x \in \mathcal{A}^{\mathbb{Z}}$. This convention is equivalent to the more common choice of summing over $X \in \mathbb{Z}$ such that $0 \in X$, taking care to divide each term by the cardinality of X ; see e.g. [Rue04, §3.3].

⁷ A notable case where $C_n = O(1)$ is when the interactions have finite range; in this case also Bowen’s regularity condition and (2.12) hold. We refer the reader to [Rue04, Ch. 5].

which, combined with (2.18), shows that \mathbb{P} is weak Gibbs in the sense of (2.16). The same is true of \mathbb{Q} . To simplify the discussion, we shall assume going forward that

$$F(\Phi) = F(\Psi) = 0, \tag{2.20}$$

which can be achieved by adding suitable constants to $\Phi_{\{i\}}$ and $\Psi_{\{i\}}$ for every $i \in \mathbb{Z}$.

The expressions for $q_{\mathbb{Q}}$ and the rate functions obtained in Sect. 2.3 remain valid, as replacing (2.12) with (2.16) does not affect these computations (note that (2.20) implies (2.13)). There, $\int \varphi d\eta$ and $\int \psi d\eta$ are simply understood as the specific energies of the state η .

We then obtain, either by (2.14) or by direct computation using (2.19), the relations

$$q_{\mathbb{Q}}(\alpha) = F(\Phi - \alpha\Psi) \quad \text{and} \quad q_{\mathbb{P}}(\alpha) = F((1 - \alpha)\Phi),$$

valid for all $\alpha \in \mathbb{R}$. We remark that $q_{\mathbb{P}}(\alpha)$ is related to the free energy density at (minus the) inverse temperature $1 - \alpha$. Moreover, in this setup, γ_+ takes the form of (minus) the asymptotic ground-state energy per unit volume:

$$\gamma_+ = \lim_{n \rightarrow \infty} \frac{1}{n} \sup_{u \in \mathcal{A}^n} U_n(u). \tag{2.21}$$

The summability condition (2.17) allows for phase transitions, i.e. the coexistence of several equilibrium measures, and thus the measures \mathbb{P} and \mathbb{Q} may fail to be ergodic, while still satisfying our decoupling assumptions. As a consequence, $q_{\mathbb{Q}}$ is not differentiable in general.

Remark 2.4. Again, we have limited the above discussion to the full shift Ω , but the conclusions remain true on many subshifts. In particular, they hold if the subshift satisfies the flexible specification property of Definition A.2 with $\tau_n = O(1)$. Indeed, although the technical details are slightly involved, one can then adapt the arguments of Lemma 9.2 in [LPS95, §9] in order to obtain (A.3) and (A.4), so that Lemmas B.7 and B.8, together with Lemma A.4, establish UD, SLD, JSLD and PA. In particular, this applies to any transitive subshift of finite type.

2.4.2. *g-measures* A continuous function $g : \Omega \rightarrow (0, 1]$ is called a *g-function* on Ω if

$$\sum_{y \in T^{-1}\{x\}} g(y) = 1$$

for all $x \in \Omega$. In this case, the potential $\varphi = \ln g$ has vanishing topological pressure ($p_{\text{top}}(\varphi) = 0$), and any equilibrium measure (recall (2.11)) for the potential φ is called a *g-measure* on Ω ; see e.g. [Wal75, PPW78, Wal05]. Let \mathbb{P} be such an equilibrium measure. It is then well known that $\text{supp } \mathbb{P} = \Omega$ and that

$$g_n(x) := \frac{\mathbb{P}_n(x_1^n)}{\mathbb{P}_{n-1}(x_2^n)} \tag{2.22}$$

defines a sequence of continuous functions that converges uniformly to g on Ω ; see e.g. [PPW78, §4]. As shown in Lemma B.11, this uniform convergence has two important consequences. First, it implies that \mathbb{P} satisfies the weak Gibbs condition. Second, it implies that there is an $e^{o(n)}$ -sequence $(D_n)_{n \in \mathbb{N}}$ such that

$$D_n^{-1} \mathbb{P}_n(x_1^n) \mathbb{P}_m(x_{n+1}^{n+m}) \leq \mathbb{P}_{n+m}(x_1^{n+m}) \leq D_n \mathbb{P}_n(x_1^n) \mathbb{P}_m(x_{n+1}^{n+m})$$

for all $x \in \Omega$, and all $n, m \in \mathbb{N}$, which implies **UD** and **SLD** with $\tau_n = 0$, but with a possibly unbounded sequence $(C_n)_{n \in \mathbb{N}}$. By Remark 1.5, **PA** holds as well. If \mathbb{Q} is a second g -measure on Ω (for a possibly different g -function g'), then **JSLD** holds as well by Remark 1.2. Our results thus apply, and once again, the expressions obtained in Sect. 2.3 remain valid thanks to (2.12), with $\varphi = \ln g$ and $\psi = \ln g'$.

Remark 2.5. The above discussion can easily be generalized to transitive Markov subshifts $\Omega' \subseteq \Omega$. However, developing a theory of g -measures on more general subshifts seems to be more delicate, and goes beyond the scope of the present paper. Still, without any reference to the general theory of g -measures, one obtains by combining Lemmas B.11, B.7 and B.8 that our decoupling assumptions hold if Ω' satisfies the *flexible specification property* and the *abundant periodic orbit property* of Definition A.2, and if one assumes that the sequence $(g_n)_{n \in \mathbb{N}}$ defined by (2.22) converges uniformly to some continuous function $g : \Omega' \rightarrow (0, 1]$; see Lemma B.11 for a precise statement.

We conclude this subsection by noting two interesting references. First, Hulse constructed in [Hul06] an example of a g -measure that is not ergodic, showing once again that **UD**, **SLD** and **PA** do not imply ergodicity.⁸ Second, in the framework of g -measures, the large deviations of empirical entropies (which are also entropy estimators) were studied in [CG05].

2.5. Hidden Markov models and lack of exponential tightness. A hidden Markov measure \mathbb{P} on Ω with finite hidden alphabet \mathcal{A}^H is obtained from a shift-invariant Markov measure \mathbb{P}^H on $\Omega^H := (\mathcal{A}^H)^\mathbb{N}$ and a surjective map $f : \mathcal{A}^H \rightarrow \mathcal{A}$ by prescribing the marginals $\mathbb{P}_n := \mathbb{P}_n^H \circ (f^{\otimes n})^{-1}$ for all $n \in \mathbb{N}$. The name “hidden Markov model” refers to the pair $(\mathbb{P}^H, \mathbb{P})$. There exist several different characterizations of hidden Markov measures; one which is particularly useful from the point of view of decoupling properties is the representation in terms of products of matrices discussed in Proposition 2.25 in [BCJP21]. The reader is also encouraged to consult Example 2.25 in [CJPS19]. Using that $|\mathcal{A}^H| < \infty$, it is straightforward to show that, if \mathbb{P}^H and \mathbb{Q}^H are irreducible, stationary Markov measures on Ω^H that satisfy $\mathbb{P}_n^H \ll \mathbb{Q}_n^H$ for all $n \in \mathbb{N}$, then \mathbb{P} and \mathbb{Q} defined using the same function f satisfy **UD**, **JSLD** and **PA**, and Theorems 0 and A–C apply.

Obviously, this is a generalization of the setup of Sect. 2.2, but in a completely different spirit from that of Sect. 2.3: hidden Markov measures can be far from Gibbsian; see e.g. Theorem 2.10 in [BCJP21].⁹ The assumption $|\mathcal{A}^H| < \infty$ ensures the existence of a constant $c \geq 0$ such that

$$\inf_{x \in \text{supp } \mathbb{P}} \mathbb{P}_n(x_1^n) \geq e^{-cn} \quad \text{and} \quad \inf_{x \in \text{supp } \mathbb{Q}} \mathbb{Q}_n(x_1^n) \geq e^{-cn}, \tag{2.23}$$

so $0 \leq -\frac{1}{n} \ln \mathbb{Q}_1(x_1^n) \leq c$ for \mathbb{P} -almost every $x \in \Omega$. These bounds imply exponential tightness and thus guarantee goodness of the rate function $I_{\mathbb{Q}}$. In fact, $I_{\mathbb{Q}}$ is infinite on (c, ∞) , and the same is then automatically true of I_W, I_V and I_R . By the same token, the bounds (2.23) imply that $q_{\mathbb{Q}} \leq c\alpha$ for all $\alpha \geq 0$, so $q_{\mathbb{Q}}, q_W, q_R$ and q_V are finite everywhere. The bounds (2.23) and these consequences are a common feature of all the examples in Sects. 2.1–2.4.

⁸ Ergodicity follows from **SLD** if we further assume that $\tau_n = O(1)$ and $C_n = O(1)$; see Lemma A.2 in [CJPS19] for a slightly more general sufficient condition.

⁹ Historical details and references are given in the discussion of Blackwell–Furstenberg–Walters–van den Berg measures in [BCJP21, §2.2]. Important references on the topic of Gibbsianity of hidden Markov measures (or lack thereof) include [LMV98, §3], [CU11] and [Ver11].

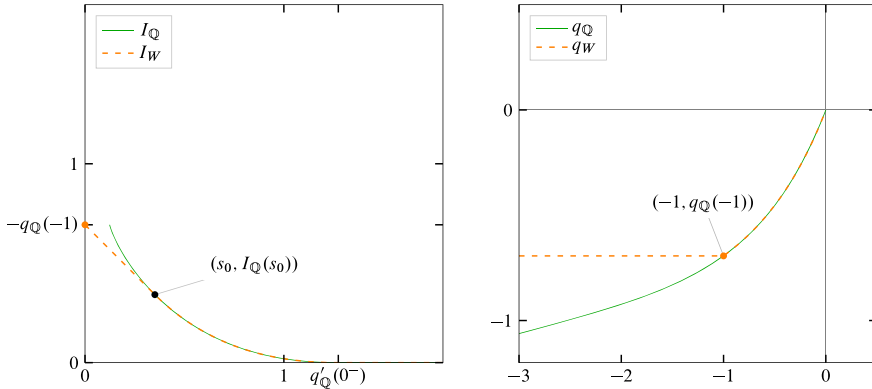


Fig. 4. Illustration of Example 2.6. In the right picture, $q_Q(\alpha) = q_W(\alpha) = \infty$ for all $\alpha > 0$

Dropping the assumption that \mathcal{A}^H is finite, the class of hidden Markov measures allows for examples where (2.23) fails, but one then needs to verify on a case-by-case basis whether UD, JSLD and PA are satisfied in order for our results to apply. Using $\mathcal{A}^H = \{0\} \cup \mathbb{N}$ as in [CJPS19, §A.2], one easily constructs fully supported, admissible pairs (\mathbb{P}, \mathbb{Q}) on $\mathcal{A} = \{\mathbf{a}, \mathbf{b}\}$, satisfying also PA, but violating either or both inequalities in (2.23). We limit ourselves to providing two sets of parameters following the notation in [CJPS19, §A.2]: Examples 2.6 and 2.7 are illustrated in Figs. 4 and 5, respectively.

Example 2.6. Define \mathbb{P} using $\gamma(n) = n \ln 2$, and \mathbb{Q} using $\hat{\gamma}(n) = 2n + 0.15n^2$. Then, the measure \mathbb{P} is uniform and $\mathbb{Q}(\mathbf{b}^n) \sim e^{-\hat{\gamma}(n)}$. Note that the quadratic term in the definition of $\hat{\gamma}(n)$ makes the second bound in (2.23) fail. Here,

$$\begin{aligned} q_Q(\alpha) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \ln(\mathbb{P}_n(\mathbf{b}^n) \mathbb{Q}_n(\mathbf{b}^n)^{-\alpha}) \\ &= -\ln 2 + \alpha \liminf_{n \rightarrow \infty} \frac{\hat{\gamma}(n)}{n}, \end{aligned}$$

so $q_Q(\alpha) = \infty$ for all $\alpha > 0$. One can show that $q'_Q(0^-) < \infty$ and that $I_Q(s) = I_W(s) = 0$ for all $s \geq q'_Q(0^-)$. Thus, the sequence $(-\frac{1}{n} \ln \mathbb{Q}_n)_{n \in \mathbb{N}}$ is not exponentially tight with respect to \mathbb{P} , and the rate functions I_Q and I_W are not good; see Fig. 4.

Example 2.7. Define $\mathbb{P} = \mathbb{Q}$ using $\gamma(0) = \hat{\gamma}(0) = 0$, $\gamma(n) = \hat{\gamma}(n) = 0.6 e^n$ for $n \in \mathbb{N}$. Then, we have $\mathbb{P}(\mathbf{b}^n) = \mathbb{Q}(\mathbf{b}^n) \sim e^{-\gamma(n)}$ and both bounds in (2.23) fail. Here, $q_{\mathbb{P}}$ is infinite for all $\alpha > 1$, and $q'_{\mathbb{P}}(1^-) = \infty$. Each sequence of interest is exponentially tight and has a rate function that is good, but remains finite on an unbounded interval; in other words $\gamma_- = -\infty$.

Remark 2.8. The possible lack of exponential tightness is specific to the case where $\mathbb{P} \neq \mathbb{Q}$; when $\mathbb{P} = \mathbb{Q}$, our assumptions imply that all the random variables discussed in the paper are exponentially tight and that the corresponding rate functions are good. At the root of this fact is the exponential tightness of $(-\frac{1}{n} \ln \mathbb{P}_n)_{n \in \mathbb{N}}$ with respect to \mathbb{P} (Theorem 0.iii). This exponential tightness will be derived as a consequence of (1.16), but it can also be seen directly: for any $M \geq 0$,

$$\mathbb{P} \left\{ x : -\frac{1}{n} \ln \mathbb{P}_n(x_1^n) \geq M \right\} = \mathbb{P} \{ x : \mathbb{P}_n(x_1^n) \leq e^{-nM} \} \leq |\mathcal{A}|^n e^{-nM}. \quad (2.24)$$

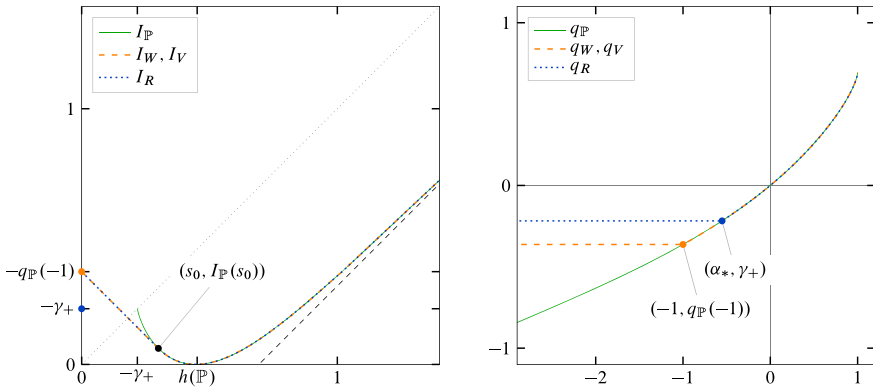


Fig. 5. Illustration of Example 2.7. In the left picture, the dashed black asymptote is given by the equation $y = x - q_{\mathbb{P}}(1)$, and in the right picture, $q_{\mathbb{P}}(\alpha) = q_W(\alpha) = q_V(\alpha) = q_R(\alpha) = \infty$ for all $\alpha > 1$

We will come back to this point in Remark 5.3.

Remark 2.9. This class of examples also allows for cases where the pressure $q_{\mathbb{Q}}$ is not differentiable—and hence where the Gärtner–Ellis route cannot provide the conclusions of Theorem 0. Indeed, using the parameters $\gamma(n) = \hat{\gamma}(n) = \frac{n}{10} + 4 \ln(1 + \frac{n}{4})$ yields a situation where $q_{\mathbb{Q}}$ is not differentiable at $\alpha_0 \approx -0.1769$ and the rate functions are affine on an interval corresponding to the subdifferential of $q_{\mathbb{Q}}$ at α_0 .

3. Key Estimates

This section is devoted to technical estimates on the distribution of W_n, R_n and V_n at large but finite $n \in \mathbb{N}$. We start with estimates for W_n and then use those for our analysis of R_n and V_n . In order to do so, we first need a convenient reformulation of our decoupling assumptions.

We show in Lemma B.2 that, at the cost of replacing C_n with $C_n|A|^{\tau_n}$ (which also satisfies (1.8)), UD implies that for every $n \in \mathbb{N}, A \in \mathcal{F}_n$ and $B \in \mathcal{F}$,

$$\mathbb{P}(A \cap T^{-n-\tau_n} B) \leq C_n \mathbb{P}(A) \mathbb{P}(B). \tag{3.1}$$

In the same way, we show in Lemma B.3 that at the cost of replacing C_n with $(\tau_n + 1)C_n$, SLD implies that for every $n \in \mathbb{N}, A \in \mathcal{F}_n$ and $B \in \mathcal{F}$,

$$\max_{0 \leq \ell \leq \tau_n} \mathbb{P}(A \cap T^{-n-\ell} B) \geq C_n^{-1} \mathbb{P}(A) \mathbb{P}(B). \tag{3.2}$$

We shall freely use the form (3.1) of UD and (3.2) of SLD throughout the paper.

3.1. Waiting times. Because $W_n : \Omega \times \Omega \rightarrow \mathbb{N}$ is $(\mathcal{F}_n \otimes \mathcal{F})$ -measurable, we will sometimes identify it to a function $\mathcal{A}^n \times \Omega \rightarrow \mathbb{N}$ denoted by the same symbol.

Lemma 3.1. Assume \mathbb{Q} satisfies SLD. Then, for all $m, n \in \mathbb{N}$ and $u \in \mathcal{A}^n$,

$$\mathbb{Q}\{y : W_n(u, y) = m\} \leq \mathbb{Q}_n(u) (1 - C_n^{-1} \mathbb{Q}_n(u))^{\lfloor \frac{m-1}{n+\tau_n} \rfloor} \tag{3.3}$$

and

$$\mathbb{Q}\{y : W_n(u, y) \geq m\} \leq (1 - C_n^{-1}\mathbb{Q}_n(u)) \lfloor \frac{m-1}{n+\tau_n} \rfloor. \tag{3.4}$$

Proof. Let us fix n, m and u as in the statement, and let $k := \lfloor \frac{m-1}{n+\tau_n} \rfloor$. First, if $k = 0$, then (3.4) is trivial and (3.3) holds since $\{y : W_n(u, y) = m\} \subseteq T^{1-m}[u]$ and \mathbb{Q} is shift invariant. Consider now the case $k \in \mathbb{N}$. In view of SLD (recall (3.2)) and shift invariance, we may inductively pick k integers $0 \leq \ell_1, \ell_2, \dots, \ell_k \leq \tau_n$ such that the intersections inductively defined by

$$A_0 := [u] \quad \text{and} \quad A_j := [u]^c \cap T^{-n-\ell_j} A_{j-1}$$

for $j = 1, \dots, k$ satisfy $\mathbb{Q}([u] \cap T^{-n-\ell_j} A_{j-1}) \geq C_n^{-1}\mathbb{Q}_n(u)\mathbb{Q}(A_{j-1})$, and thus also

$$\mathbb{Q}(A_j) = \mathbb{Q}(A_{j-1}) - \mathbb{Q}([u] \cap T^{-n-\ell_j} A_{j-1}) \leq \mathbb{Q}(A_{j-1})(1 - C_n^{-1}\mathbb{Q}_n(u)). \tag{3.5}$$

Iterating (3.5) starting from A_0 yields

$$\mathbb{Q}(A_k) \leq \mathbb{Q}_n(u)(1 - C_n^{-1}\mathbb{Q}_n(u))^k. \tag{3.6}$$

Let $M := \{m - jn - \sum_{i=1}^j \ell_i\}_{j=1}^k$. By construction, $kn + \sum_{i=1}^k \ell_j \leq k(n + \tau_n) < m$, so $M \subseteq [1, m - n] \subseteq [1, m - 1]$. As a consequence,

$$\begin{aligned} \{y : W_n(u, y) = m\} &= T^{1-m}[u] \cap \bigcap_{m'=1}^{m-1} T^{1-m'}[u]^c \\ &\subseteq T^{1-m}[u] \cap \bigcap_{m' \in M} T^{1-m'}[u]^c \\ &= T^{1-m+nk+\sum_{j=1}^k \ell_j} A_k. \end{aligned}$$

Thus, by shift invariance and (3.6), we readily obtain (3.3). The proof of (3.4) is exactly the same with A_0 replaced by Ω . □

Remark 3.2. The bounds (3.5) and (3.6) are inspired by [Kon98, §2], and were already adapted to selective decoupling conditions in [CDEJR23a]. It might be surprising that the upper bound (3.3) relies on the lower decoupling assumption SLD. In fact, even if $\tau_n = 0$ for all n , using UD would yield $\mathbb{Q}(A_k) \leq C_n^k \mathbb{Q}_n(u)(1 - \mathbb{Q}_n(u))^k$ instead of (3.6); the extra factor of C_n^k is too crude since we will be interested in the case where $k \gg n$. The opposite will happen in Lemma 3.4, where a lower bound will be proved using UD; see (3.13).

If $\mathbb{Q}_n(u) = 0$, then $\mathbb{Q}\{y : W_n(u, y) = m\} \leq \mathbb{Q}(T^{1-m}[u]) = \mathbb{Q}_n(u) = 0$ for all $m \in \mathbb{N}$, and thus $W_n(u, \cdot)$ is almost surely infinite. Conversely, if \mathbb{Q} satisfies SLD, the bound (3.4) ensures that $W_n(u, \cdot)$ is almost surely finite whenever $\mathbb{Q}_n(u) > 0$. This last observation is at the heart of the next lemma.

Lemma 3.3. *Let $n \in \mathbb{N}$, let \mathbb{Q} satisfy SLD, and assume that $\mathbb{P}_n \ll \mathbb{Q}_n$. Then, the random variable W_n is $(\mathbb{P} \otimes \mathbb{Q})$ -almost surely finite.*

Proof. This is a consequence of the decomposition

$$W_n(x, y) = \sum_{u \in \mathcal{A}^n} 1_{[u]}(x) W_n(u, y),$$

the bound (3.4), and the absolute continuity assumption. \square

We now turn to lower bounds on the distribution of $W_n(u, \cdot)$. The following lemma will be useful when $(1 - C_n \mathbb{Q}_n(u))^{\lceil \frac{m-1}{n+\tau_n} \rceil}$ is close to 1.

Lemma 3.4. *Assume \mathbb{Q} satisfies UD. Then, for all $n, m \in \mathbb{N}$ and $u \in \mathcal{A}^n$ such that $C_n \mathbb{Q}_n(u) \leq 1$,*

$$\begin{aligned} \mathbb{Q}\{y : W_n(u, y) = m\} &\geq \frac{1}{n + \tau_n} \mathbb{Q}_n(u) \left(1 - (n + \tau_n) \left(1 - (1 - C_n \mathbb{Q}_n(u))^{\lceil \frac{m-1}{n+\tau_n} \rceil} \right) \right). \end{aligned} \tag{3.7}$$

Proof. Let us fix n, m and u as in the statement. We first prove that it suffices to establish the bound

$$\begin{aligned} \mathbb{Q}\{y : k(n + \tau_n) < W_n(u, y) \leq (k + 1)(n + \tau_n)\} &\geq \mathbb{Q}_n(u) \left(1 - (n + \tau_n) \left(1 - (1 - C_n \mathbb{Q}_n(u))^k \right) \right) \end{aligned} \tag{3.8}$$

for all $k \in \mathbb{N} \cup \{0\}$. Since

$$\{y : W_n(u, y) = t + 1\} \subseteq T^{-1}\{y : W_n(u, y) = t\} \tag{3.9}$$

for all $t \in \mathbb{N}$, the probability $\mathbb{Q}\{y : W_n(u, y) = t\}$ is nonincreasing in t . As a consequence, the left-hand side of (3.8) is bounded above by $(n + \tau_n) \mathbb{Q}\{y : W_n(u, y) = k(n + \tau_n) + 1\}$, so (3.8) implies that

$$\begin{aligned} \mathbb{Q}\{y : W_n(u, y) = 1 + k(n + \tau_n)\} &\geq \frac{1}{n + \tau_n} \mathbb{Q}_n(u) \left(1 - (n + \tau_n) \left(1 - (1 - C_n \mathbb{Q}_n(u))^k \right) \right). \end{aligned}$$

By nonincreasingness, the same lower bound applies to $\mathbb{Q}\{y : W_n(u, y) = m\}$ for all k such that $1 + k(n + \tau_n) \geq m$, and thus (3.8) indeed implies (3.7).

We now establish (3.8). For every $y \in T^{1-(k+1)(n+\tau_n)}[u] =: A$ we have $W_n(u, y) \leq (k + 1)(n + \tau_n)$, and for every $y \in \bigcap_{r=1}^{k(n+\tau_n)} T^{1-r}[u]^c$ we have $W_n(u, y) > k(n + \tau_n)$. As a consequence,

$$\mathbb{Q}\{y : k(n + \tau_n) < W_n(u, y) \leq (k + 1)(n + \tau_n)\} \geq \mathbb{Q} \left(A \cap \bigcap_{r=1}^{k(n+\tau_n)} T^{1-r}[u]^c \right). \tag{3.10}$$

We now write $\bigcap_{r=1}^{k(n+\tau_n)} T^{1-r}[u]^c = \bigcap_{j=1}^{n+\tau_n} B_j$, where for $1 \leq j \leq n + \tau_n$,

$$B_j := \bigcap_{d=0}^{k-1} T^{1-d(n+\tau_n)-j}[u]^c.$$

Notice that for each j , the events whose intersection defines B_j are separated by “gaps” of size τ_n , which will allow to use **UD** below. By a union bound, (3.10) implies that

$$\begin{aligned} \mathbb{Q}\{y : k(n + \tau_n) < W_n(u, y) \leq (k + 1)(n + \tau_n)\} &\geq \mathbb{Q}\left(A \cap \bigcap_{j=1}^{n+\tau_n} B_j\right) \\ &= \mathbb{Q}(A) - \mathbb{Q}\left(\bigcup_{j=1}^{n+\tau_n} (A \cap B_j^c)\right) \\ &\geq \mathbb{Q}(A) - \sum_{j=1}^{n+\tau_n} \mathbb{Q}(A \cap B_j^c) \\ &= \mathbb{Q}(A) - \sum_{j=1}^{n+\tau_n} (\mathbb{Q}(A) - \mathbb{Q}(A \cap B_j)). \end{aligned}$$

By shift invariance, $\mathbb{Q}(A) = \mathbb{Q}_n(u)$, and thus the proof of (3.8) will be complete once we have shown that

$$\mathbb{Q}(A \cap B_j) \geq \mathbb{Q}_n(u)(1 - C_n \mathbb{Q}_n(u))^k. \tag{3.11}$$

Fix $1 \leq j \leq n + \tau_n$. We have

$$A \cap B_j = T^{1-j} \left(T^{j-(k+1)(n+\tau_n)}[u] \cap B_1 \right) = T^{1-j} F_k, \tag{3.12}$$

with F_0, \dots, F_k inductively defined by

$$F_0 := T^{j-n-\tau_n}[u] \quad \text{and} \quad F_i := [u]^c \cap T^{-n-\tau_n} F_{i-1}$$

for $i = 1, \dots, k$. By **UD** (recall (3.1)),

$$\mathbb{Q}(F_i) = \mathbb{Q}(F_{i-1}) - \mathbb{Q}([u] \cap T^{-n-\tau_n} F_{i-1}) \geq \mathbb{Q}(F_{i-1})(1 - C_n \mathbb{Q}_n(u)).$$

Iterating this bound starting with F_0 yields

$$\mathbb{Q}(F_k) \geq \mathbb{Q}_n(u)(1 - C_n \mathbb{Q}_n(u))^k. \tag{3.13}$$

Combining this with (3.12) and using shift invariance establishes (3.11), as claimed. \square

The next proposition makes precise the “cut-off” phenomenon sketched in (1.30) and uses the notation

$$U_n(s) := \{u \in \mathcal{A}^n : \mathbb{Q}_n(u) \leq e^{-ns}\} \tag{3.14}$$

for $s > 0$.

Proposition 3.5. *Suppose that \mathbb{Q} satisfies **UD** and **SLD**, and let $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ satisfy $\mathbb{P}_n \ll \mathbb{Q}_n$ for all $n \in \mathbb{N}$. Then, for all $s > 0$ and all $0 < \delta \leq \epsilon < \frac{s}{2}$, we have, for all large enough n ,*

$$\begin{aligned} \mathbb{P} \otimes \mathbb{Q}\{(x, y) : \frac{1}{n} \ln W_n(x, y) \in B(s, \epsilon)\} \\ \leq e^{n(s+\epsilon)} \sum_{u \in U_n(s-\epsilon-\delta)} \mathbb{Q}_n(u) \mathbb{P}_n(u) + \exp(-e^{\frac{n\delta}{2}}) \end{aligned} \tag{3.15}$$

and

$$\mathbb{P} \otimes \mathbb{Q}\{(x, y) : \frac{1}{n} \ln W_n(x, y) \in B(s, \epsilon)\} \geq e^{n(s-\delta)} \sum_{u \in U_n(s+\delta)} \mathbb{Q}_n(u) \mathbb{P}_n(u). \quad (3.16)$$

Proof. The key idea in this proof is to show that, when $\frac{1}{n} \ln m \in B(s, \epsilon)$, the $\lfloor \frac{m-1}{n+\tau_n} \rfloor$ -th power of $(1 - C_n^{-1} \mathbb{Q}_n(u))$ in (3.3) vanishes superexponentially for all $u \notin U_n(s - \epsilon - \delta)$ as $n \rightarrow \infty$, and that the $\lceil \frac{m-1}{n+\tau_n} \rceil$ -th power of $(1 - C_n \mathbb{Q}_n(u))$ in (3.7) is very close to 1 for all $u \in U_n(s + \delta)$ as $n \rightarrow \infty$. We fix s, ϵ and δ as in the statement and first note that

$$\begin{aligned} \mathbb{P} \otimes \mathbb{Q}\{(x, y) : \frac{1}{n} \ln W_n(x, y) \in B(s, \epsilon)\} \\ = \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{Q}\{y : \frac{1}{n} \ln W_n(u, y) \in B(s, \epsilon)\}. \end{aligned} \quad (3.17)$$

Proof of (3.15). Notice that $|\{m \in \mathbb{N} : \frac{1}{n} \ln m \in B(s, \epsilon)\}| \leq e^{n(s+\epsilon)}$. Thus, using (3.17), the nonincreasingness of $t \mapsto \mathbb{Q}\{y : W_n(u, y) = t\}$ (recall (3.9)), and then Lemma 3.1, we find

$$\begin{aligned} \mathbb{P} \otimes \mathbb{Q}\{(x, y) : \frac{1}{n} \ln W_n(x, y) \in B(s, \epsilon)\} \\ \leq e^{n(s+\epsilon)} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{Q}\{y : W_n(u, y) = m_n\} \\ \leq e^{n(s+\epsilon)} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{Q}_n(u) (1 - C_n^{-1} \mathbb{Q}_n(u))^{k_n}, \end{aligned} \quad (3.18)$$

where $m_n := \lceil e^{n(s-\epsilon)} \rceil$ and $k_n := \lfloor \frac{m_n-1}{n+\tau_n} \rfloor$. We now split the above sum into a sum over $U_n(s - \epsilon - \delta)$ and a sum over $U_n^c(s - \epsilon - \delta)$. Clearly,

$$\sum_{u \in U_n(s-\epsilon-\delta)} \mathbb{P}_n(u) \mathbb{Q}_n(u) (1 - C_n^{-1} \mathbb{Q}_n(u))^{k_n} \leq \sum_{u \in U_n(s-\epsilon-\delta)} \mathbb{P}_n(u) \mathbb{Q}_n(u), \quad (3.19)$$

while on the other hand,

$$\sum_{u \in U_n^c(s-\epsilon-\delta)} \mathbb{P}_n(u) \mathbb{Q}_n(u) (1 - C_n^{-1} \mathbb{Q}_n(u))^{k_n} \leq (1 - C_n^{-1} e^{-n(s-\epsilon-\delta)})^{k_n}. \quad (3.20)$$

But using the inequality $(1 - x)^{k_n} \leq e^{-x k_n}$, we find

$$\begin{aligned} e^{n(s+\epsilon)} (1 - C_n^{-1} e^{-n(s-\epsilon-\delta)})^{k_n} &\leq \exp\left(n(s+\epsilon) - C_n^{-1} e^{-n(s-\epsilon-\delta)} k_n\right) \\ &\leq \exp(-e^{\frac{n\delta}{2}}) \end{aligned}$$

for n large enough, thanks to the fact that $k_n \geq e^{n(s-\epsilon)-o(n)}$ and $C_n = e^{o(n)}$ as $n \rightarrow \infty$. Therefore, using the estimates (3.19) and (3.20) in (3.18) indeed yields (3.15).

Proof of (3.16). This time, note that $|\{m \in \mathbb{N} : \frac{1}{n} \ln m \in (s - \epsilon, s)\}| \geq \frac{1}{2} e^{ns}$ for n large enough, and set $m_n := \lfloor e^{ns} \rfloor$ and $k_n := \lceil \frac{m_n-1}{n+\tau_n} \rceil$. Then, by (3.17) and the

nonincreasingness of $t \mapsto \mathbb{Q}\{y : W_n(u, y) = t\}$, we obtain

$$\begin{aligned} & \mathbb{P} \otimes \mathbb{Q}\{(x, y) : \frac{1}{n} \ln W_n(x, y) \in B(s, \epsilon)\} \\ & \geq \mathbb{P} \otimes \mathbb{Q}\{(x, y) : \frac{1}{n} \ln W_n(x, y) \in (s - \epsilon, s]\} \\ & \geq \frac{e^{ns}}{2} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{Q}\{y : W_n(u, y) = m_n\} \\ & \geq \frac{e^{ns}}{2} \sum_{u \in U_n(s+\delta)} \mathbb{P}_n(u) \mathbb{Q}\{y : W_n(u, y) = m_n\}. \end{aligned}$$

Then, for n large enough so that $C_n e^{-n(s+\delta)} \leq 1$, we can apply Lemma 3.4 to every $u \in U_n(s + \delta)$, and we obtain that

$$\begin{aligned} & \mathbb{P} \otimes \mathbb{Q}\{(x, y) : \frac{1}{n} \ln W_n(x, y) \in B(s, \epsilon)\} \\ & \geq \frac{e^{ns}}{2} \sum_{u \in U_n(s+\delta)} \mathbb{P}_n(u) \mathbb{Q}_n(u) \left((n + \tau_n)^{-1} - \left(1 - (1 - C_n \mathbb{Q}_n(u))^{k_n} \right) \right) \tag{3.21} \\ & \geq \frac{e^{ns}}{2} \sum_{u \in U_n(s+\delta)} \mathbb{P}_n(u) \mathbb{Q}_n(u) \left((n + \tau_n)^{-1} - \left(1 - (1 - C_n e^{-n(s+\delta)})^{k_n} \right) \right). \end{aligned}$$

Notice that by Bernoulli’s inequality, $(1-x)^{k_n} \geq 1-xk_n$ for all $x \leq 1$. As a consequence, for all n large enough,

$$(1 - C_n e^{-n(s+\delta)})^{k_n} \geq 1 - C_n e^{-n(s+\delta)} k_n \geq 1 - e^{-\frac{n\delta}{2}},$$

where we have used that $k_n \leq e^{ns+o(n)}$ and $C_n = e^{o(n)}$. Substitution into (3.21) yields

$$\begin{aligned} & \mathbb{P} \otimes \mathbb{Q}\{(x, y) : \frac{1}{n} \ln W_n(x, y) \in B(s, \epsilon)\} \\ & \geq \frac{e^{ns}}{2} \sum_{u \in U_n(s+\delta)} \mathbb{P}_n(u) \mathbb{Q}_n(u) \left((n + \tau_n)^{-1} - e^{-\frac{n\delta}{2}} \right), \end{aligned}$$

from which the lower bound (3.16) readily follows when n is large enough. □

3.2. From waiting times to return times.

Lemma 3.6. *Assume \mathbb{P} satisfies UD. Then, for all $s > 0$ and all $0 < \epsilon < s$, we have, for all large enough n ,*

$$\mathbb{P}\{x : \frac{1}{n} \ln R_n(x) \in B(s, \epsilon)\} \leq C_n \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{P}\{x : \frac{1}{n} \ln W_n(u, x) \in B(s, 2\epsilon)\}$$

and

$$\mathbb{P}\{x : \frac{1}{n} \ln V_n(x) \in B(s, \epsilon)\} \leq C_n \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{P}\{x : \frac{1}{n} \ln W_n(u, x) \in B(s, 2\epsilon)\}.$$

Proof. Fix $s > \epsilon > 0$, fix $n \in \mathbb{N}$ large enough so that

$$e^{n(s-\epsilon)} > n + \tau_n - 1 \quad \text{and} \quad e^{n(s-2\epsilon)} \leq e^{n(s-\epsilon)} + 1 - n - \tau_n, \quad (3.22)$$

and fix $u \in \mathcal{A}^n$. Then, for all $x \in [u]$ such that $\frac{1}{n} \ln R_n(x) \in B(s, \epsilon)$, we have $R_n(x) = W_n(u, Tx) \in (e^{n(s-\epsilon)}, e^{n(s+\epsilon)})$. The first condition in (3.22) implies that $W_n(u, T^{n+\tau_n}x) = W_n(u, Tx) + 1 - n - \tau_n$ for all such x , so that, also using the second condition in (3.22),

$$W_n(u, T^{n+\tau_n}x) \in (e^{n(s-\epsilon)} + 1 - n - \tau_n, e^{n(s+\epsilon)} + 1 - n - \tau_n) \subset (e^{n(s-2\epsilon)}, e^{n(s+2\epsilon)}).$$

As a consequence, by UD,

$$\begin{aligned} \mathbb{P}([u] \cap \{x : \frac{1}{n} \ln R_n(x) \in B(s, \epsilon)\}) &\leq \mathbb{P}([u] \cap \{x : \frac{1}{n} \ln W_n(u, T^{n+\tau_n}x) \in B(s, 2\epsilon)\}) \\ &\leq C_n \mathbb{P}_n(u) \mathbb{P}\{x : \frac{1}{n} \ln W_n(u, x) \in B(s, 2\epsilon)\}. \end{aligned}$$

Taking the sum over $u \in \mathcal{A}^n$ proves the desired bound for R_n . The proof of the bound for V_n is almost identical: for all $x \in [u]$ such that $\frac{1}{n} \ln V_n(x) \in B(s, \epsilon)$, we have this time $W_n(u, T^n x) \in (e^{n(s-\epsilon)}, e^{n(s+\epsilon)})$, so that $W_n(u, T^{n+\tau_n}x) \in (e^{n(s-\epsilon)} - \tau_n, e^{n(s+\epsilon)} - \tau_n) \subset (e^{n(s-2\epsilon)}, e^{n(s+2\epsilon)})$. The remainder of the argument is unchanged. \square

Lemma 3.7. *Assume \mathbb{P} satisfies SLD. Then, for all $s > 0$ and all $0 < \epsilon < s$, we have, for all large enough n ,*

$$\mathbb{P}\{x : \frac{1}{n} \ln R_n(x) \in B(s, \epsilon)\} \geq \frac{C_n^{-1}}{n + \tau_n} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{P}\{x : \frac{1}{n} \ln W_n(u, x) \in B(s, \frac{1}{2}\epsilon)\}$$

and

$$\mathbb{P}\{x : \frac{1}{n} \ln V_n(x) \in B(s, \epsilon)\} \geq \frac{C_n^{-1}}{1 + \tau_n} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{P}\{x : \frac{1}{n} \ln W_n(u, x) \in B(s, \frac{1}{2}\epsilon)\}.$$

Proof. We first prove the statement concerning R_n . Fix $s > \epsilon > 0$, fix n large enough so that

$$e^{n(s+\frac{1}{2}\epsilon)} + n + \tau_n - 1 < e^{n(s+\epsilon)}, \quad (3.23)$$

and fix $u \in \mathcal{A}^n$. By SLD, there exists $0 \leq \ell \leq \tau_n$ such that the set

$$A_u := [u] \cap T^{-n-\ell}\{x : \frac{1}{n} \ln W_n(u, x) \in B(s, \frac{1}{2}\epsilon)\}$$

satisfies

$$\mathbb{P}(A_u) \geq C_n^{-1} \mathbb{P}_n(u) \mathbb{P}\{x : \frac{1}{n} \ln W_n(u, x) \in B(s, \frac{1}{2}\epsilon)\}. \quad (3.24)$$

We now claim that

$$A_u \subseteq \bigcup_{j=1}^{n+\tau_n} T^{1-j}\{x : \frac{1}{n} \ln R_n(x) \in B(s, \epsilon)\}. \quad (3.25)$$

To prove (3.25), take an arbitrary $x \in A_u$, and let $m := W_n(u, T^{n+\ell}x)$, which, by the definition of A_u , satisfies $\frac{1}{n} \ln m \in B(s, \frac{1}{2}\epsilon)$. Then, the set $I := \{i \in \mathbb{N} : x_i^{i+n-1} = u\}$

contains 1 and $n + \ell + m$, but excludes $\{n + \ell + 1, n + \ell + 2, \dots, n + \ell + m - 1\}$. Hence, with $j := \max\{i \in I : i \leq n + \ell\}$ we find $R_n(T^{j-1}x) = n + \ell + m - j \in [m, m + n + \tau_n - 1]$, so by (3.23) we obtain (3.25).

By taking the union over $u \in \mathcal{A}^n$ in the left-hand side of (3.25), and using shift invariance to bound the probability of the right-hand side, we obtain

$$\sum_{u \in \mathcal{A}^n} \mathbb{P}(A_u) = \mathbb{P}\left(\bigcup_{u \in \mathcal{A}^n} A_u\right) \leq (n + \tau_n) \mathbb{P}\{x : \frac{1}{n} \ln R_n(x) \in B(s, \epsilon)\}. \tag{3.26}$$

In view of (3.24), we have completed the proof of the statement about R_n .

To adapt the proof for V_n , it suffices to replace the definition of j with $j := \max\{i \in I : i \leq 1 + \ell\}$; then $V_n(T^{j-1}x) = 1 + \ell + m - j \in [m, m + \tau_n]$ and the same arguments apply, with the factor $(n + \tau_n)$ replaced by $(1 + \tau_n)$ in (3.26). \square

Lemma 3.8. *Assume \mathbb{P} satisfies SLD. Then, for every $n \in \mathbb{N}$, the random variables R_n and V_n are \mathbb{P} -almost surely finite.*

Proof. In view of (3.4), the random variable $W_n(u, \cdot)$ is \mathbb{P} -almost surely finite for every fixed $u \in \mathcal{A}^n$ with $\mathbb{P}_n(u) > 0$. By shift invariance, so is $W_n(u, \cdot) \circ T^k$ for each k . In view of the expressions

$$R_n(x) = \sum_{u \in \mathcal{A}^n} 1_{[u]}(x) W_n(u, Tx),$$

$$V_n(x) = \sum_{u \in \mathcal{A}^n} 1_{[u]}(x) W_n(u, T^n x),$$

the conclusion is immediate. \square

Remark 3.9. An alternative way of showing that R_n and V_n are almost surely finite is to use the Poincaré recurrence theorem, see e.g. Theorem 1.4 in [Wal82].

4. Weak LDPs and Ruelle–Lanford Functions

We now briefly recall some terminology from the theory of large deviations, limiting ourselves to sequences of real-valued random variables. See for example [DS89, Eli06, DZ09] for proper introductions to the field. Let $(Z_n)_{n \in \mathbb{N}}$ be a sequence of (almost surely finite) real-valued random variables on a probability space (Ω_*, \mathbb{P}_*) . The cases of interest will be

- $\Omega_* = \Omega \times \Omega$, $\mathbb{P}_* = \mathbb{P} \otimes \mathbb{Q}$ and $Z_n = \frac{1}{n} \ln W_n$ for Theorem A;
- $\Omega_* = \Omega$, $\mathbb{P}_* = \mathbb{P}$, with $Z_n = -\frac{1}{n} \ln Q_n$, $Z_n = \frac{1}{n} \ln V_n$ and $Z_n = \frac{1}{n} \ln R_n$ for, respectively, Theorems 0, B and C.

The sequence $(Z_n)_{n \in \mathbb{N}}$ is said to satisfy the *large deviation principle* (LDP) if there exists a lower semicontinuous function $I : \mathbb{R} \rightarrow [0, \infty]$ such that

$$-\inf_{s \in O} I(s) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}_*\{x : Z_n(x) \in O\} \tag{4.1}$$

for every open set $O \subseteq \mathbb{R}$ and

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}_*\{x : Z_n(x) \in \Gamma\} \leq -\inf_{s \in \Gamma} I(s) \tag{4.2}$$

for every closed set $\Gamma \subseteq \mathbb{R}$. The bounds (4.1) and (4.2) are respectively called the *large-deviation lower bound* and the *large-deviation upper bound*, and the function I — which is unique when it exists — is called the *rate function*. Following standard terminology, we say that I is a *good* rate function if it properly diverges as $s \rightarrow \pm\infty$. We also recall that the large-deviation upper bound applied to the set \mathbb{R} implies that $\inf_{s \in \mathbb{R}} I(s) = 0$. Ubiquitous in the theory of large deviations is the question of whether the rate function is convex and can be expressed as the Legendre–Fenchel transform of the corresponding pressure. A detailed analysis of these considerations for the random variables of interest is postponed to Sect. 5.

Our analysis will require additional vocabulary which is discussed e.g. in [DZ09, §1.2]. The sequence $(Z_n)_{n \in \mathbb{N}}$ is said to satisfy the *weak large deviation principle* if (4.1) holds for all open sets $O \subseteq \mathbb{R}$, and (4.2) holds for all *compact* sets $\Gamma \subseteq \mathbb{R}$. We shall sometimes refer to the standard LDP as the *full* LDP when we need to emphasize the contrast to the weak LDP used as a stepping stone towards the full LDP. The following notion will play a role in doing so: the sequence $(Z_n)_{n \in \mathbb{N}}$ is said to be *exponentially tight* if, for every $\beta \in \mathbb{R}$, there exists $M > 0$ such that $\mathbb{P}_* \{x : Z_n(x) \notin [-M, M]\} \leq e^{-\beta n}$ for all n large enough. To be more precise, we will appeal to the two following facts for real-valued sequences. First, if the weak LDP holds and the sequence is exponentially tight, then the full LDP holds with a good rate function. Second, if the full LDP holds with a good rate function, then the sequence is exponentially tight. While all our LDPs are full, and while exponential tightness does play a role in our analysis, we emphasize that the sequence $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ need not be exponentially tight, as illustrated in Sect. 2.5.

As mentioned, we will first prove the weak LDP. We will do so using *Ruelle–Lanford (RL) functions*. We introduce the *lower RL function* $\underline{I} : \mathbb{R} \rightarrow [0, \infty]$ defined by¹⁰

$$\underline{I}(s) := - \lim_{\epsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}_* \{x : Z_n(x) \in B(s, \epsilon)\},$$

and the *upper RL function* $\bar{I} : \mathbb{R} \rightarrow [0, \infty]$ defined by

$$\bar{I}(s) := - \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}_* \{x : Z_n(x) \in B(s, \epsilon)\}.$$

It follows from their definition that \bar{I} and \underline{I} are lower semicontinuous, and that $\bar{I} \leq \underline{I}$. Moreover, the weak LDP holds if and only if we have the equality

$$\underline{I}(s) = \bar{I}(s) \tag{4.3}$$

for every $s \in \mathbb{R}$; see e.g. [DZ09, §4.1.2] or [CJPS19, §3.2]. The common value in (4.3) must then coincide with $I(s)$.

The core of this section is devoted to proving the weak LDP for the sequences of interest via the validity of (4.3). To be more precise, for each sequence, both RL functions are shown to be equal to the proposed rate function, as detailed in Table 2. We consider separately the case $s > 0$ in Sect. 4.1 and the case $s = 0$ in Sect. 4.2, noting that the case $s < 0$ is trivial by mere nonnegativity of the random variables under study. We denote by $\underline{I}_W, \underline{I}_V$ and \underline{I}_R (resp. \bar{I}_W, \bar{I}_V and \bar{I}_R) the lower (resp. upper) RL functions associated with our sequences.

¹⁰ The name “RL function” was first used in [LPS94, LPS95]. The method of RL functions is often used in conjunction with subadditive arguments; this is in particular the case of the derivation of Theorem 0 in [CJPS19]. In the present paper, once Theorem 0 is taken for granted, the proof of our results is not of the subadditive kind.

Table 2. Assumptions, proposed rate function and references to where the key equality (4.3) is established for the three weak LDPs to be proved

Random variable	Probability space	Assumptions	Proposed rate function	Eq. (4.3) for $s > 0$	Eq. (4.3) for $s = 0$
$\frac{1}{n} \ln W_n$	$(\Omega^2, \mathbb{P} \otimes \mathbb{Q})$	admissibility	I_W (1.19)	Prop. 4.2	Prop. 4.4
$\frac{1}{n} \ln V_n$	(Ω, \mathbb{P})	SLD, UD	I_V (1.22)	Prop. 4.3	Prop. 4.5
$\frac{1}{n} \ln R_n$	(Ω, \mathbb{P})	SLD, UD, PA	I_R (1.25)	Prop. 4.3	Prop. 4.7

4.1. *At positive values.* Our first goal is to prove equality of the upper and lower Ruelle–Lanford functions at positive values of s . We start with the RL functions \underline{I}_W and \bar{I}_W of the sequence $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$. Most of the work to show that $\underline{I}_W(s) = \bar{I}_W(s)$ when $s > 0$ was done in Sect. 3.1: in view of Proposition 3.5, it only remains to estimate the quantity

$$J_n(s) := \sum_{u \in U_n(s)} \mathbb{Q}_n(u) \mathbb{P}_n(u) = \int_{[s, \infty)} e^{-rn} d\mu_n(r), \tag{4.4}$$

where μ_n is the distribution of $-\frac{1}{n} \ln \mathbb{Q}_n$ with respect to \mathbb{P} , and where $U_n(s)$ was defined in (3.14). The integral in (4.4) allows to express the limiting behavior of J_n in terms of the rate function $I_{\mathbb{Q}}$ of Theorem 0, as shown by the following straightforward variation of Varadhan’s lemma.

Lemma 4.1. *Assume (\mathbb{P}, \mathbb{Q}) is admissible. Then, for all $s > 0$,*

$$\sup_{r > s} (-r - I_{\mathbb{Q}}(r)) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \ln J_n(s) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \ln J_n(s) \leq \sup_{r \geq s} (-r - I_{\mathbb{Q}}(r)).$$

Proof. Let us fix $s > 0$. For the lower bound, note that for every choice of $r > s$ and $0 < \epsilon < r - s$,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \ln J_n(s) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \int_{B(r, \epsilon)} e^{-r'n} d\mu_n(r') \\ &\geq -r - \epsilon + \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mu_n(B(r, \epsilon)) \\ &\geq -r - \epsilon - I_{\mathbb{Q}}(r), \end{aligned}$$

where we have used the large-deviation lower bound of Theorem 0. For the upper bound, let $\beta := \inf_{r \geq s} (r + I_{\mathbb{Q}}(r)) \in [s, \infty)$. Then,

$$J_n(s) = \int_{[s, \beta]} e^{-rn} d\mu_n(r) + \int_{(\beta, \infty)} e^{-rn} d\mu_n(r) \leq \int_{[s, \beta]} e^{-rn} d\mu_n(r) + e^{-\beta n}.$$

As a consequence, it suffices to show that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \int_{[s, \beta]} e^{-rn} d\mu_n(r) \leq -\beta, \tag{4.5}$$

which, since $[s, \beta]$ is compact, follows from a standard covering argument, see e.g. Lemma 4.3.6 in [DZ09] or the proof of Proposition C.1.iv below. \square

Proposition 4.2. *Assume the pair (\mathbb{P}, \mathbb{Q}) is admissible. Then, for all $s > 0$, the RL functions for $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ with respect to $\mathbb{P} \otimes \mathbb{Q}$ satisfy*

$$\underline{I}_W(s) = \bar{I}_W(s) = -s + \inf_{r \geq s} (r + I_{\mathbb{Q}}(r)). \tag{4.6}$$

Proof. Let $s > 0$. By Proposition 3.5 and Lemma 4.1, we find that for all $0 < \delta \leq \epsilon < \frac{1}{2}s$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P} \otimes \mathbb{Q} \{ (x, y) : \frac{1}{n} \ln W_n(x, y) \in B(s, \epsilon) \} \leq s + \epsilon - \inf_{r \geq s - \epsilon - \delta} (r + I_{\mathbb{Q}}(r)) \tag{4.7}$$

and

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P} \otimes \mathbb{Q} \{ (x, y) : \frac{1}{n} \ln W_n(x, y) \in B(s, \epsilon) \} \geq s - \delta - \inf_{r > s + \delta} (r + I_{\mathbb{Q}}(r)). \tag{4.8}$$

By taking the limit as $\delta \rightarrow 0$ first and then $\epsilon \rightarrow 0$ in (4.7), and since $I_{\mathbb{Q}}$ is lower semicontinuous, we obtain that $-\bar{I}_W(s) \leq s - \inf_{r \geq s} (r + I_{\mathbb{Q}}(r))$. Taking the same limits in (4.8) yields $-\underline{I}_W(s) \geq s - \inf_{r > s} (r + I_{\mathbb{Q}}(r))$. We then have

$$-s + \inf_{r \geq s} (r + I_{\mathbb{Q}}(r)) \leq \bar{I}_W(s) \leq \underline{I}_W(s) \leq -s + \inf_{r > s} (r + I_{\mathbb{Q}}(r)). \tag{4.9}$$

For all $s' \in (0, s)$, the last inequality in (4.9) applied to s' yields

$$\underline{I}_W(s') \leq -s' + \inf_{r > s'} (r + I_{\mathbb{Q}}(r)) \leq -s' + \inf_{r \geq s} (r + I_{\mathbb{Q}}(r)).$$

Since \underline{I}_W is lower semicontinuous (as a RL function), this in turn implies that

$$\underline{I}_W(s) \leq \liminf_{s' \uparrow s} \underline{I}_W(s') \leq -s + \inf_{r \geq s} (r + I_{\mathbb{Q}}(r)),$$

so the first three quantities in (4.9) are actually equal, as desired. □

Proposition 4.3. *Assume \mathbb{P} satisfies UD and SLD. Then, for all $s > 0$, the RL functions for $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ and $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ with respect to \mathbb{P} satisfy*

$$\underline{I}_R(s) = \bar{I}_R(s) = -s + \inf_{r \geq s} (r + I_{\mathbb{P}}(r))$$

and

$$\underline{I}_V(s) = \bar{I}_V(s) = -s + \inf_{r \geq s} (r + I_{\mathbb{P}}(r)).$$

Proof. Let $s > 0$. First, we remark that Proposition 4.2 applies to the admissible pair (\mathbb{P}, \mathbb{P}) , so that $\bar{I}_W(s) = \underline{I}_W(s) = -s + \inf_{r \geq s} (r + I_{\mathbb{P}}(r))$. Then, Lemma 3.6 implies that $\bar{I}_R(s) \geq \bar{I}_W(s)$ and $\bar{I}_V(s) \geq \bar{I}_W(s)$; recall also (3.17). In the same way, Lemma 3.7 implies that $\underline{I}_R(s) \leq \underline{I}_W(s)$ and $\underline{I}_V(s) \leq \underline{I}_W(s)$. Since also $\bar{I}_R \leq \underline{I}_R$ and $\bar{I}_V \leq \underline{I}_V$, the proof is complete. □

4.2. *At the origin.* In this subsection, we prove that, for each of the sequences $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$, $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ and $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$, the upper and lower RL functions match at $s = 0$. We recall that the limit

$$q_{\mathbb{Q}}(-1) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{Q}_n(u) \tag{4.10}$$

exists for any admissible pair (\mathbb{P}, \mathbb{Q}) by Theorem 0.ii.

Proposition 4.4. *If the pair (\mathbb{P}, \mathbb{Q}) is admissible, then*

$$\underline{I}_W(0) = \bar{I}_W(0) = -q_{\mathbb{Q}}(-1) = \inf_{r \geq 0} (r + I_{\mathbb{Q}}(r)). \tag{4.11}$$

Proof. For all $\epsilon > 0$ and $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P} \otimes \mathbb{Q}\{(x, y) : \frac{1}{n} \ln W_n(x, y) \in B(0, \epsilon)\} &\geq \mathbb{P} \otimes \mathbb{Q}\{(x, y) : W_n(x, y) = 1\} \\ &= \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{Q}_n(u). \end{aligned}$$

In view of (4.10) and the definition of \underline{I}_W , we have $\underline{I}_W(0) \leq -q_{\mathbb{Q}}(-1)$. To obtain the opposite inequality for $\bar{I}_W(0)$, observe that

$$\mathbb{P} \otimes \mathbb{Q}\{(x, y) : W_n(x, y) = k\} \leq \mathbb{P} \otimes \mathbb{Q}\{(x, y) : x_1^n = y_{1+k}^{n+k}\} = \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{Q}_n(u)$$

for every $u \in \mathcal{A}^n$, $n \in \mathbb{N}$ and $k \in \mathbb{N}$. Therefore, a union bound gives, for every $\epsilon > 0$,

$$\mathbb{P} \otimes \mathbb{Q}\{(x, y) : W_n(x, y) \leq e^{\epsilon n}\} \leq e^{\epsilon n} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \mathbb{Q}_n(u).$$

By (4.10) and the definition of \bar{I}_W , we conclude that $\bar{I}_W(0) \geq -q_{\mathbb{Q}}(-1)$. Since also $\underline{I}_W(0) \geq \bar{I}_W(0)$ we have thus established the first two equalities in (4.11).

To complete the proof, it remains to observe that, since $I_{\mathbb{Q}}(r) = \infty$ for all $r < 0$, and since $q_{\mathbb{Q}} = I_{\mathbb{Q}}^*$ by Theorem 0.ii,

$$\inf_{r \geq 0} (r + I_{\mathbb{Q}}(r)) = \inf_{r \in \mathbb{R}} (r + I_{\mathbb{Q}}(r)) = -\sup_{r \in \mathbb{R}} (-r - I_{\mathbb{Q}}(r)) = -q_{\mathbb{Q}}(-1), \tag{4.12}$$

which establishes the last identity in (4.11). □

Proposition 4.5. *If \mathbb{P} satisfies UD and SLD, then*

$$\underline{I}_V(0) = \bar{I}_V(0) = -q_{\mathbb{P}}(-1) = \inf_{r \geq 0} (r + I_{\mathbb{P}}(r)). \tag{4.13}$$

Proof. The stated assumptions allow to apply Theorem 0 to the pair (\mathbb{P}, \mathbb{P}) , and in particular its consequences (4.10) and (4.12) with $\mathbb{Q} = \mathbb{P}$. Since the third equality in (4.13) is a special case of (4.12), and since $\bar{I}_V(0) \leq \underline{I}_V(0)$ by definition, it suffices to establish the inequalities $\underline{I}_V(0) \leq -q_{\mathbb{P}}(-1)$ and $\bar{I}_V(0) \geq -q_{\mathbb{P}}(-1)$ in order to complete the proof. To this end, we let $\epsilon > 0$ be arbitrary and restrict our attention to n large enough so that $\min\{n, e^{\epsilon n}\} > \tau_n + 1$.

For each $u \in \mathcal{A}^n$, **SLD** implies that there is $0 \leq \ell \leq \tau_n$ such that $\mathbb{P}([u] \cap T^{-n-\ell}[u]) \geq C_n^{-1} \mathbb{P}_n(u)^2$. Since $[u] \cap \{x : V_n(x) \leq \tau_n + 1\} \supseteq [u] \cap T^{-n-\ell}[u]$, this in turn implies that

$$\mathbb{P}\{x : V_n(x) < e^{\epsilon n}\} \geq \mathbb{P}\{x : V_n(x) \leq \tau_n + 1\} \geq C_n^{-1} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u)^2.$$

Combining this with (4.10) for $\mathbb{Q} = \mathbb{P}$ establishes the inequality $\underline{I}_V(0) \leq -q_{\mathbb{P}}(-1)$. On the other hand, for each $k \in \mathbb{N}$,

$$\{x : V_n(x) = k\} \subseteq \bigcup_{u \in \mathcal{A}^n} [u] \cap T^{1-n-k}[u] \subseteq \bigcup_{v \in \mathcal{A}^{n-\tau_n}} [v] \cap T^{1-n-k}[v].$$

Assuming without loss of generality that the sequence $(\tau_n)_{n \in \mathbb{N}}$ is nondecreasing (so in particular $\tau_{n-\tau_n} \leq \tau_n$), we obtain from **UD** that

$$\mathbb{P}\{x : V_n(x) = k\} \leq C_{n-\tau_n} \sum_{v \in \mathcal{A}^{n-\tau_n}} \mathbb{P}_{n-\tau_n}(v)^2.$$

Considering the union over $k = 1, 2, \dots, \lceil e^{\epsilon n} \rceil - 1$, we further obtain

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : V_n(x) < e^{\epsilon n}\} \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \left(e^{\epsilon n} C_{n-\tau_n} \sum_{v \in \mathcal{A}^{n-\tau_n}} \mathbb{P}_{n-\tau_n}(v)^2 \right) \\ & = \epsilon + \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u)^2, \end{aligned} \tag{4.14}$$

where we have used that $\lim_{n \rightarrow \infty} \frac{n-\tau_n}{n} = 1$. Combining this with (4.10) for $\mathbb{Q} = \mathbb{P}$ establishes the inequality $\bar{I}_V(0) \geq -q_{\mathbb{P}}(-1)$, so the proof is complete. \square

Let us now turn to $\underline{I}_R(0)$ and $\bar{I}_R(0)$, whose comparison is significantly more involved. We start with a technical lemma.

Lemma 4.6. *If \mathbb{P} satisfies **SLD** and **UD**, then*

$$\gamma_+ = \lim_{n \rightarrow \infty} \frac{1}{n} \sup_{u \in \mathcal{A}^n} \ln \mathbb{P}_n(u) \geq q_{\mathbb{P}}(-1). \tag{4.15}$$

Proof. From Theorem 0.iii.b, we know that the limit superior in the definition (1.12) of γ_+ is actually a limit, which is the first equality in (4.15). We now prove the inequality, noting that, by Theorem 0.ii, the limit defining $q_{\mathbb{P}}(-1)$ exists. For each $n \in \mathbb{N}$, we find $\sup_{v \in \mathcal{A}^n} \mathbb{P}(v) = \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \sup_{v \in \mathcal{A}^n} \mathbb{P}_n(v) \geq \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u)^2$, and the claim follows from (4.10) with $\mathbb{Q} = \mathbb{P}$. \square

The following proposition shows that $\underline{I}_R(0) = \bar{I}_R(0) = -\gamma_+$ under the assumptions of Theorem C. We give several additional inequalities in order to underline the role of **PA**. The proposition also shows that, while $\bar{I}_R(0)$ and $\underline{I}_R(0)$ are defined in terms of $\{x : R_n(x) < e^{\epsilon n}\}$, the subset $\{x : R_n(x) < n\}$ accounts for the full behavior of the probability at the exponential scale.

Proposition 4.7. *The following hold:*

i. *If \mathbb{P} satisfies UD, then*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : R_n(x) < n\} \leq -\bar{I}_R(0) \leq \gamma_+. \tag{4.16}$$

ii. *If \mathbb{P} satisfies PA, then*

$$-\underline{I}_R(0) \geq \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : R_n(x) < n\} \geq \gamma_+. \tag{4.17}$$

iii. *If \mathbb{P} satisfies UD and PA, then*

$$-\underline{I}_R(0) = -\bar{I}_R(0) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : R_n(x) < n\} = \gamma_+.$$

Proof. Given $u \in \mathcal{A}^n$, we denote by $\text{per}(u)$ the period of u , i.e. the smallest $p \in \mathbb{N}$ such that $u_i = u_{i+p}$ for all $1 \leq i \leq n - p$. Since the condition is vacuously true for $p = n$, we have the bound $\text{per}(u) \leq n$. The key observation is that, for every $1 \leq k < n$ and $x \in \Omega$,

$$R_n(x) = k \iff \text{per}(x_1^{k+n}) = k. \tag{4.18}$$

i. Since $\{x : R_n(x) < n\} \subseteq \{x : R_n(x) < e^{\epsilon n}\}$ for all large enough n , the first inequality in (4.16) readily follows from the definition of \bar{I}_R . We now prove the second inequality. Let n be large enough so that $n > \max\{1, \tau_n\}$ and let $1 \leq k < n$. The map $\{u \in \mathcal{A}^n : \text{per}(u) = k\} \rightarrow \mathcal{A}^k$ given by $u \mapsto u_{n-k+1}^n$ is injective. By (4.18), we thus have

$$\mathbb{P}\{x : R_n(x) = k\} = \sum_{u \in \mathcal{A}^n : \text{per}(u) = k} \mathbb{P}_{n+k}(uu_{n-k+1}^n) \leq \sum_{v \in \mathcal{A}^k} \sup_{u \in \mathcal{A}^n} \mathbb{P}_{n+k}(uv).$$

As in the proof of Proposition 4.5, we assume without loss of generality that $(\tau_n)_{n \in \mathbb{N}}$ is nondecreasing. Then, UD yields

$$\mathbb{P}_{n+k}(uv) \leq \mathbb{P}([u_1^{n-\tau_n}] \cap T^{-n}[v]) \leq C_{n-\tau_n} \mathbb{P}_{n-\tau_n}(u_1^{n-\tau_n}) \mathbb{P}_k(v),$$

and so

$$\begin{aligned} \mathbb{P}\{x : R_n(x) = k\} &\leq C_{n-\tau_n} \sum_{v \in \mathcal{A}^k} \sup_{u \in \mathcal{A}^n} \mathbb{P}_{n-\tau_n}(u_1^{n-\tau_n}) \mathbb{P}_k(v) \\ &= C_{n-\tau_n} \sup_{u \in \mathcal{A}^n} \mathbb{P}_{n-\tau_n}(u_1^{n-\tau_n}) \\ &= C_{n-\tau_n} \sup_{u \in \mathcal{A}^{n-\tau_n}} \mathbb{P}_{n-\tau_n}(u). \end{aligned}$$

Taking a union over $1 \leq k < n$ gives

$$\mathbb{P}\{x : R_n(x) < n\} \leq (n - 1) C_{n-\tau_n} \sup_{u \in \mathcal{A}^{n-\tau_n}} \mathbb{P}_{n-\tau_n}(u).$$

Comparing with the definition (1.12) of γ_+ , and using that $\lim_{n \rightarrow \infty} \frac{n-\tau_n}{n} = 1$, we deduce that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : R_n(x) < n\} \leq \gamma_+. \tag{4.19}$$

Now, observe that

$$\mathbb{P}\{x : R_n(x) \in [n, e^{n\epsilon}]\} \leq \mathbb{P}\{x : V_n(x) \in [1, e^{n\epsilon} + 1 - n]\} \leq \mathbb{P}\{x : V_n(x) < e^{n\epsilon}\}.$$

By (4.14), which only relies on UD, this implies

$$\lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : R_n(x) \in [n, e^{n\epsilon}]\} \leq \gamma_+. \tag{4.20}$$

By definition of \bar{I}_R , the inequalities (4.19) and (4.20) imply $-\bar{I}_R(0) \leq \gamma_+$, as desired.

ii. As above, the first inequality in (4.17) is immediate by the definition of \underline{I}_R . We now establish the second one. Let $p \in \mathbb{N}$ and $u \in \mathcal{A}^p$ be arbitrary. By (4.18), for $n > p$ and $r_n = \lceil \frac{n}{p} \rceil + 1$, we have $[u^{r_n}] \subseteq \{x : R_n(x) \leq p\}$. Therefore,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : R_n(x) < n\} &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : R_n(x) \leq p\} \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}([u^{r_n}]) \\ &= \liminf_{r \rightarrow \infty} \frac{1}{pr} \ln \mathbb{P}([u^r]). \end{aligned} \tag{4.21}$$

Since the expression in the last line of (4.21) can be made arbitrarily close to γ_+ by PA, the proof of Part ii is complete.

iii. The conclusion is just the combination of Parts i and ii. □

Proposition 4.7 is the only place where PA is ever used in our proofs. We do not know if PA can be lifted, nor are we aware of any example of a measure satisfying UD and SLD but not PA. We take the remainder of this subsection to briefly discuss what remains true if PA is dropped or weakened. In this discussion, we always assume that \mathbb{P} satisfies UD and SLD.

First, if PA is dropped, we remark that Propositions 4.7.i and 4.3 still ensure that $\bar{I}_R(s) \geq I_R(s)$ for all $s \in \mathbb{R}$, with I_R defined in (1.25). Thus, the weak large-deviation upper bound for $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ holds. By retracing the proofs, one easily concludes that also the full large-deviation upper bound holds with the rate function I_R , and that $\bar{q}_R = \bar{I}_R^* \leq I_R^*$, where \bar{q}_R is defined as in (1.4) with a limit superior.

In Proposition 4.7, PA is only used to obtain the large-deviation lower bound at $s = 0$. More specifically, all we actually derive from PA is that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : R_n(x) < n\} \geq \gamma_+. \tag{4.22}$$

So instead of PA, one could have taken (4.22) as an assumption, or any other condition implying it.

In fact, one could even obtain a full LDP for $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ without (4.22). Indeed, if one can show by some means that the limit $D := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{x : R_n(x) < n\}$ exists, then necessarily $D \leq \gamma_+$ by Proposition 4.7.i, and the proofs can easily be adapted to obtain the full LDP, by merely replacing $-\gamma_+$ with $\min\{-D, I_V(0)\}$ in the definition (1.25) of I_R . We do not know under what conditions the limit defining D exists, and we have been unable to produce any counter example.

We now return to means of establishing (4.22). Under PA, the proof of Proposition 4.7.ii actually shows that the probability of $\{x : R_n(x) < n\}$ is asymptotically

captured by periodic orbits of period much smaller than n . PA can be slightly relaxed so as to take into account words u that can be repeated many times, but not necessarily infinitely many times as PA requires:

Definition 4.8 (WPA). A measure $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ satisfies the assumption of *weak periodic approximation* (WPA) if for every $\epsilon > 0$ there exists $M \in \mathbb{N}$ such that for all $m \geq M$ there is a word $u \in \Omega_{\text{fin}}$ with $|u| \leq \epsilon m$, and such that

$$\frac{1}{m} \ln \mathbb{P}_{|u|^{\lfloor \frac{m}{|u|} \rfloor}} \left(u^{\lfloor \frac{m}{|u|} \rfloor} \right) \geq \gamma_+ - \epsilon.$$

Then, an easy modification of the proof of Proposition 4.7 shows that (4.22) still holds assuming WPA instead of PA, and thus so do the conclusions of Theorem C.

In order to conclude the discussion, we briefly comment on results from the literature about Poincaré recurrence times (see [AV08, AC15, AAG21] and references therein) which can be used to establish (4.22). The Poincaré recurrence times are defined by $T_n(x) = \inf\{k \in \mathbb{N} : \mathbb{P}([x_1^n] \cap T^{-k}[x_1^n]) > 0\}$.¹¹ The asymptotic behavior of T_n is overall very different from that of R_n ; in particular $T_n \leq n + \tau_n$ almost surely by SLD. However, the following two relations hold \mathbb{P} -almost surely: first, $T_n \leq R_n$, and second, $T_n < n$ implies that $R_{n-T_n} \leq T_n$. It follows that if one can show that $\lim_{\epsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \ln \frac{1}{n} \mathbb{P}\{T_n < \epsilon n\} \geq \gamma_+$, then also $\lim_{\epsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \ln \frac{1}{n} \mathbb{P}\{x : R_n(x) < \epsilon n\} \geq \gamma_+$, and in particular (4.22) holds. Such a result is proved in [AV08] under an assumption called ‘‘Hypothesis 1’’, which is very similar to WPA. The same bound is obtained in [AC15] under an assumption called ‘‘Assumption 1’’ which, in spirit, also plays role similar to that of WPA.

5. Proof of the Main Results

At this stage, we have proved that if the pair (\mathbb{Q}, \mathbb{P}) is admissible, then $\bar{I}_W = \underline{I}_W = I_W$, with I_W defined by (1.19); see Propositions 4.2 and 4.4, and notice that all three functions are infinite on the negative real axis. This implies that the sequence $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ satisfies the weak LDP with respect to $\mathbb{P} \otimes \mathbb{Q}$, with the rate function I_W ; see the beginning of Sect. 4. In the same way, by Propositions 4.3 and 4.5, we have proved that, if \mathbb{P} satisfies UD and SLD, then the sequence $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ satisfies the weak LDP with respect to \mathbb{P} , with the rate function I_V given in (1.22). Finally, combining Propositions 4.3 and 4.7, we have shown that for \mathbb{P} satisfying UD, SLD and PA, the sequence $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ satisfies the weak LDP with respect to \mathbb{P} , with the rate function I_R given in (1.25). See Table 2 for a summary.

Since upper and lower Ruelle–Lanford functions are always lower semicontinuous, we conclude that I_W, I_V and I_R are lower semicontinuous. Alternatively, lower semicontinuity can be checked explicitly using the expressions (1.19), (1.22) and (1.25), together with the fact that the rate function $I_{\mathbb{Q}}$ in Theorem 0 is lower semicontinuous.

In this section, we promote the weak LDPs to full ones and establish the claimed relations about the rate functions and accompanying pressures. This will conclude the proofs of Theorems A, B and C. The proofs of these theorems, and actually also that of Theorem 0, have many (rather standard) arguments in common, which we have extracted as Proposition C.1.

¹¹ We use the condition $\mathbb{P}([x_1^n] \cap T^{-k}[x_1^n]) > 0$ instead of $[x_1^n] \cap T^{-k}[x_1^n] \neq \emptyset$ because the discussion is not limited to the subshift $\text{supp } \mathbb{P}$.

Lemma 5.1. *Under the assumptions of Theorem A, the rate function I_W is convex. In particular, if \mathbb{P} satisfies **UD** and **SLD**, then I_V is convex.*

Proof. We claim that for every $\lambda \in [0, 1]$ and $s_1, s_2 \in \mathbb{R}$,

$$\lambda I_W(s_1) + (1 - \lambda)I_W(s_2) \geq I_W(s)$$

where $s := \lambda s_1 + (1 - \lambda)s_2$. The stated inequality is obvious if $s_1 < 0$ or $s_2 < 0$. For $s_1, s_2 \geq 0$,

$$\begin{aligned} & \lambda I_W(s_1) + (1 - \lambda)I_W(s_2) \\ &= -s + \inf_{r_1 \geq s_1, r_2 \geq s_2} (\lambda r_1 + (1 - \lambda)r_2 + \lambda I_{\mathbb{Q}}(r_1) + (1 - \lambda)I_{\mathbb{Q}}(r_2)) \\ &\geq -s + \inf_{r_1 \geq s_1, r_2 \geq s_2} (\lambda r_1 + (1 - \lambda)r_2 + I_{\mathbb{Q}}(\lambda r_1 + (1 - \lambda)r_2)) \\ &= -s + \inf_{r \geq s} (r + I_{\mathbb{Q}}(r)) = I_W(s), \end{aligned}$$

where the inequality in the second line relies on the convexity of $I_{\mathbb{Q}}$, by Theorem 0.i. The second part of the lemma follows from the identity $I_V = I_W$ when $\mathbb{P} = \mathbb{Q}$; recall Remark 1.11. \square

Lemma 5.2. *Suppose that \mathbb{Q} satisfies **SLD**. Then, for every $\alpha > 0$, there exists an $e^{o(n)}$ -sequence $(\kappa_{\alpha, n})_{n \in \mathbb{N}}$ such that*

$$\int W_n(u, y)^\alpha d\mathbb{Q}(y) \leq \kappa_{\alpha, n} \mathbb{Q}_n(u)^{-\alpha}$$

for each $u \in \text{supp } \mathbb{Q}_n$.

Proof. Fix $n \in \mathbb{N}$, $u \in \text{supp } \mathbb{Q}_n$, and let $q := 1 - C_n^{-1} \mathbb{Q}_n(u)$. Without loss of generality, we assume that $C_n \geq C_1 > 1$, so that $0 < 1 - C_1^{-1} \leq q < 1$. For all $t \geq 0$, the bound (3.4) of Lemma 3.1 yields

$$\begin{aligned} \mathbb{Q}\{y : W_n(u, y) > t\} &= \mathbb{Q}\{y : W_n(u, y) \geq \lfloor t \rfloor + 1\} \\ &\leq q^{\lfloor \frac{t}{n+\tau_n} \rfloor} \\ &\leq q^{\frac{t}{n+\tau_n} - 2} \\ &\leq (1 - C_1^{-1})^{-2} q^{\frac{t}{n+\tau_n}}. \end{aligned}$$

Then, by a standard consequence of Fubini’s theorem (see e.g. Theorem 8.16 in [Rud87]),

$$\begin{aligned} \int W_n(u, y)^\alpha d\mathbb{Q}(y) &= \alpha \int_0^\infty t^{\alpha-1} \mathbb{Q}\{y : W_n(u, y) > t\} dt \\ &\leq \frac{\alpha}{(1 - C_1^{-1})^2} \int_0^\infty t^{\alpha-1} q^{\frac{t}{n+\tau_n}} dt \\ &= \frac{\alpha}{(1 - C_1^{-1})^2} \Gamma(\alpha) \left(\frac{1}{n + \tau_n} \right)^{-\alpha} (-\ln q)^{-\alpha}. \end{aligned}$$

Further using that $-\ln(q) = -\ln(1 - C_n^{-1} \mathbb{Q}_n(u)) \geq C_n^{-1} \mathbb{Q}_n(u)$ completes the proof. \square

We are now in a position to prove the main results. With Proposition C.1 from “Appendix C” and the weak LDPs at hand, the proofs of Theorems A–C only consist in providing a few remaining estimates specific to each case.

5.1. *Proof of Theorem A.* Since $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ satisfies the weak LDP, the assumptions of Proposition C.1 are satisfied with $Z_n := \frac{1}{n} \ln W_n$ on $(\Omega_*, \mathbb{P}_*) := (\Omega \times \Omega, \mathbb{P} \otimes \mathbb{Q})$, with the convex (recall Lemma 5.1) rate function $I := I_W$. We now show that the conclusions of Theorem A follow from Proposition C.1.

We first prove that q_W exists and that $q_W = I_W^*$. For this we define \underline{q}_W and \bar{q}_W as the limit inferior and limit superior corresponding to the definition (1.6) of q_W . The bound $\underline{q}_W \geq I_W^*$ is provided by Proposition C.1.iii, and by Proposition C.1.v we have $q_W(\alpha) = I_W^*(\alpha)$ for all $\alpha < 0$. Since $I_W \leq I_Q$ by definition, we have $I_W^*(0) \geq I_Q^*(0) = q_Q(0) = 0 = q_W(0)$, where we have used Theorem 0.ii. Finally, for $\alpha > 0$, Lemma 5.2 gives

$$\begin{aligned} \iint W_n(x, y)^\alpha d\mathbb{P}(x)d\mathbb{Q}(y) &= \sum_{u \in \text{supp } \mathbb{P}_n} \mathbb{P}_n(u) \int W_n(u, y)^\alpha d\mathbb{Q}(y) \\ &\leq \kappa_{n,\alpha} \sum_{u \in \text{supp } \mathbb{P}_n} \mathbb{P}_n(u) \mathbb{Q}_n(u)^{-\alpha}. \end{aligned} \tag{5.1}$$

Note that we have used the absolute continuity granted by admissibility of the pair (\mathbb{P}, \mathbb{Q}) . We conclude that

$$\bar{q}_W(\alpha) \leq q_Q(\alpha) = I_Q^*(\alpha) \leq I_W^*(\alpha),$$

where from left to right, we have used (5.1), Theorem 0.ii, and the fact that $I_W \leq I_Q$ by definition. We have thus proved that $q_W = I_W^*$. By Proposition C.1.vii, the weak LDP then extends to a full one. This completes the proof of Theorem A.i.

For Part ii, we have already established that $q_W = I_W^*$ (in particular q_W exists as a limit), and since I_W is convex and lower semicontinuous, this implies that also $I_W = q_W^*$. To establish (1.20), note that

$$q_W(\alpha) = I_W^*(\alpha) = \sup_{s \geq 0} (\alpha s - \inf_{r \geq s} (r - s + I_Q(r))) = \sup_{r \geq s \geq 0} ((\alpha + 1)s - r - I_Q(r)).$$

Since the quantity to optimize is linear in s , it suffices to consider the extremal points $s \in \{0, r\}$, which yields, using again Theorem 0.ii,

$$\begin{aligned} q_W(\alpha) &= \sup_{r \geq 0} (\max\{0, (\alpha + 1)r\} - r - I_Q(r)) \\ &= \max \left\{ \sup_{r \geq 0} (-r - I_Q(r)), \sup_{r \geq 0} (\alpha r - I_Q(r)) \right\} \\ &= \max\{q_Q(\alpha), q_Q(-1)\}, \end{aligned}$$

so the proof of Theorem A.ii is complete.

In order to prove Theorem A.iii, we simply note that when $\mathbb{Q} = \mathbb{P}$, then by (1.20) and (1.16), we have $q_W(1) = q_{\mathbb{P}}(1) = h_{\text{top}}(\text{supp } \mathbb{P}) < \infty$, so Proposition C.1.vi applies. \square

5.2. *Proof of Theorem B.* The proof is almost the same as that of Theorem A, applying this time Proposition C.1 to $Z_n := \frac{1}{n} \ln V_n$ on $(\Omega_*, \mathbb{P}_*) := (\Omega, \mathbb{P})$, with the convex rate function $I := I_V$. We will only require a slightly more involved argument to derive the inequality $\bar{q}_V(\alpha) \leq I_V^*(\alpha)$ when $\alpha > 0$, where \bar{q}_V is defined by taking the limit superior in the definition (1.5) of q_V . For $\alpha > 0$, recalling that $W_n(u, x) \leq W_n(u, T^k x) + k$, we obtain

$$\begin{aligned} \int V_n^\alpha d\mathbb{P} &= \sum_{u \in \mathcal{A}^n} \int 1_{[u]}(x) W_n(u, T^n x)^\alpha d\mathbb{P}(x) \\ &\leq \sum_{u \in \mathcal{A}^n} \int 1_{[u]}(x) (W_n(u, T^{n+\tau_n} x) + \tau_n)^\alpha d\mathbb{P}(x). \end{aligned}$$

Then, by UD,

$$\begin{aligned} &\int 1_{[u]}(x) (W_n(u, T^{n+\tau_n} x) + \tau_n)^\alpha d\mathbb{P}(x) \\ &= \sum_{j=1}^\infty \mathbb{P}([u] \cap T^{-n-\tau_n} \{x : W_n(u, x) = j\}) (j + \tau_n)^\alpha \\ &\leq C_n \sum_{j=1}^\infty \mathbb{P}_n(u) \mathbb{P}\{x : W_n(u, x) = j\} (j + \tau_n)^\alpha \\ &= C_n \mathbb{P}_n(u) \int (W_n(u, x) + \tau_n)^\alpha d\mathbb{P}(x), \end{aligned}$$

so

$$\begin{aligned} \int V_n^\alpha d\mathbb{P} &\leq C_n \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \int (W_n(u, x) + \tau_n)^\alpha d\mathbb{P}(x) \\ &\leq C_n \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \int (W_n(u, x) (1 + \tau_n))^\alpha d\mathbb{P}(x) \\ &\leq C_n (1 + \tau_n)^\alpha \kappa_{\alpha,n} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u)^{1-\alpha}, \end{aligned}$$

where the last inequality was obtained using Lemma 5.2 with $\mathbb{Q} = \mathbb{P}$. By this and Theorem 0.ii, and since $I_V \leq I_{\mathbb{P}}$ by definition, we conclude that $\bar{q}_V(\alpha) \leq q_{\mathbb{P}}(\alpha) = I_{\mathbb{P}}^*(\alpha) \leq I_V^*(\alpha)$, as claimed.

The same arguments as in the proof of Theorem A then provide the full LDP together with the Legendre–Fenchel duality relations and (1.23), which is merely a special case of (1.20).

Next, the exponential tightness and goodness assertions in Theorem B.i follow, as in the proof of Theorem A.iii, from Proposition C.1.vi and the bound $q_V(1) = q_{\mathbb{P}}(1) = h_{\text{top}}(\text{supp } \mathbb{P}) < \infty$.

Finally, (1.24) is proved as follows: by Lemma 4.6 we have $q_{\mathbb{P}}(-1) \leq \gamma_+$ and by (1.23) and Legendre–Fenchel duality we have $I_V(0) = \sup_{\alpha \in \mathbb{R}} (-q_V(\alpha)) = -q_{\mathbb{P}}(-1)$; this can also be obtained using (4.13). \square

5.3. *Proof of Theorem C.* Once more, we plan to apply Proposition C.1, this time with $Z_n := \frac{1}{n} \ln R_n$ on $(\Omega_*, \mathbb{P}_*) := (\Omega, \mathbb{P})$ and the rate function $I := I_R$, keeping in mind that the latter is not convex in general.

The proof that $q_R = I_R^*$ follows the same argument as in Theorems A and B, and once again, only the proof of the inequality $\bar{q}_R(\alpha) \leq I_R^*(\alpha)$ when $\alpha > 0$ needs to be adapted, where \bar{q}_R is again defined by taking the limit superior in the definition (1.4) of q_R . We have, for $\alpha > 0$,

$$\begin{aligned} \int R_n^\alpha d\mathbb{P} &= \sum_{u \in \mathcal{A}^n} \int 1_{[u]}(x) W_n(u, Tx)^\alpha d\mathbb{P}(x) \\ &\leq \sum_{u \in \mathcal{A}^n} \int 1_{[u]}(x) (W_n(u, T^{n+\tau_n}x) + n + \tau_n - 1)^\alpha d\mathbb{P}(x) \\ &\leq C_n (n + \tau_n)^\alpha \kappa_{\alpha,n} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u)^{1-\alpha}, \end{aligned}$$

where the steps giving the last inequality are exactly as in the proof of Theorem B with τ_n replaced by $\tau_n + n - 1$. As previously, we conclude that $\bar{q}_R(\alpha) \leq q_{\mathbb{P}}(\alpha) = I_{\mathbb{P}}^*(\alpha) \leq I_R^*(\alpha)$, since $I_R \leq I_{\mathbb{P}}$. We thus have $q_R = I_R^*$.

Unlike in the previous theorems, the rate function is not convex in general and we cannot invert the Legendre–Fenchel transform. For the same reason, we cannot apply Proposition C.1.vii to obtain the full LDP. Nevertheless, since $q_R(1) = q_{\mathbb{P}}(1) = h_{\text{top}}(\text{supp } \mathbb{P}) < \infty$, the full LDP, exponential tightness and goodness of I_R follow at once from Proposition C.1.vi. This completes the proof of Part i.

We now turn to the proof of Part ii. We have already seen that $q_R = I_R^*$, so it remains to prove (1.26), which we do by singling out the pathological case where $\gamma_- = 0$.

Case 1: $\gamma_- = 0$. In this case, $\gamma_+ = 0$ as well by (1.13), so Theorem 0.iii yields that $I_{\mathbb{P}}(0) = 0$ and $I_{\mathbb{P}}(s) = \infty$ for $s \neq 0$. It follows from its definition that $I_R = I_{\mathbb{P}}$, and taking the Legendre–Fenchel transform gives $q_R = q_{\mathbb{P}} = 0$. In particular, both sides of (1.26) vanish identically.

Case 2: $\gamma_- < 0$. Since $I_R(s) = I_V(s)$ for all $s > 0$ by definition, and since $\lim_{s \downarrow 0} I_V(s) = I_V(0)$ when $\gamma_- < 0$, we find

$$\begin{aligned} q_R(\alpha) &= \sup_{s \in \mathbb{R}} (\alpha s - I_R(s)) \\ &= \max \left\{ -I_R(0), \sup_{s > 0} (\alpha s - I_V(s)) \right\} \\ &= \max \{ -I_R(0), I_V^*(\alpha) \}. \end{aligned}$$

Further using that $-I_R(0) = \gamma_+$ by definition and that

$$I_V^*(\alpha) = q_V(\alpha) = \max \{ q_{\mathbb{P}}(\alpha), q_{\mathbb{P}}(-1) \}$$

by (1.23), we conclude that $q_R(\alpha) = \max \{ \gamma_+, q_{\mathbb{P}}(\alpha), q_{\mathbb{P}}(-1) \}$. In view of (1.24), the quantity $q_{\mathbb{P}}(-1)$ can be omitted from the maximum, and (1.26) follows.

Part iii of Theorem C, about (non)convexity of I_R , is proved in Sect. 5.4. □

Remark 5.3. In order to follow a common route for the proof of our three main theorems, our arguments to promote the weak LDP to a full one did not rely on exponential tightness, since the latter does not hold in general in the setup Theorem A.i–ii. We have instead

made extensive use of the properties of the pressures. We now briefly argue that in the context of Theorems A.iii, B and C, exponential tightness could have been proved more directly, and the full LDP could have thus been obtained without any reference to pressures.

In the setup of Theorem A.iii, that is in the case of $(\frac{1}{n} \ln W_n)_{n \in \mathbb{N}}$ with $\mathbb{Q} = \mathbb{P}$, the idea is that multiplying (3.4) in the case $m \sim e^{nM}$ by $\mathbb{P}_n(u)$ and summing over $u \in \mathcal{A}^n$ yields terms that fall in either of two categories: those with $\mathbb{P}_n(u) \gg e^{-nM}$ decay super-exponentially, while those with $\mathbb{P}_n(u) \lesssim e^{-nM}$ are controlled in view of the exponential tightness estimate (2.24). The argument relies on SLD only, and the details are easily written by following the proof of the estimate (3.15). Further assuming UD, exponential tightness can be extended to $(\frac{1}{n} \ln R_n)_{n \in \mathbb{N}}$ and $(\frac{1}{n} \ln V_n)_{n \in \mathbb{N}}$ by introducing a minor variation of Lemma 3.6 where balls are replaced by intervals of the form (s, ∞) and $(s - \epsilon, \infty)$.

5.4. Nonconvexity in Theorem C. In this subsection, we prove the numerous relations stated in Theorem C.iii. The topological entropy $h_{\text{top}}(\text{supp } \mathbb{P})$ as well as the notion of measure of maximal entropy (MME) on $\text{supp } \mathbb{P}$ are recalled in ‘‘Appendix A’’.

Proposition 5.4. *Let \mathbb{P} satisfy the assumptions of Theorem C, and consider the following properties:*

- (a) I_R is convex;
- (b) $I_R = I_V$;
- (c) $q_R = q_V$;
- (d) $\gamma_+ = q_{\mathbb{P}}(-1)$;
- (e) $I_{\mathbb{P}}(-\gamma_+) = 0$;
- (f) $q_{\mathbb{P}}(\alpha) = -\gamma_+ \alpha$ for all $\alpha \leq 0$;
- (g) $\gamma_+ = \gamma_-$;
- (h) $I_{\mathbb{P}}(s) = \infty$ for all $s \in \mathbb{R} \setminus \{-\gamma_+\}$;
- (i) $h(\mathbb{P}) = h_{\text{top}}(\text{supp } \mathbb{P})$, i.e. \mathbb{P} is a MME on $\text{supp } \mathbb{P}$.

Then, Properties (a)–(f) are equivalent, Properties (g)–(i) are equivalent, and Properties (g)–(i) imply Properties (a)–(f). If, in addition, $\tau_n = O(1)$ and $C_n = O(1)$, then Properties (a)–(i) are equivalent.

Proof. Let us summarize the information we have gathered on the rate functions. First, $I_V(s)$ and $I_R(s)$ coincide for $s \neq 0$ by comparison of (1.22) and (1.25). By the same two equations, and since we have already established that $-\gamma_+ \leq -q_{\mathbb{P}}(-1)$ in Lemma 4.6 and that $-q_{\mathbb{P}}(-1) = I_V(0)$ in Theorem B.ii, we obtain

$$0 \leq I_R(0) = -\gamma_+ \leq -q_{\mathbb{P}}(-1) = I_V(0) \leq I_{\mathbb{P}}(0). \tag{5.2}$$

We now prove that (a)–(f) are equivalent.

- (a) \Rightarrow (b) If $\gamma_+ = 0$, Theorem 0.iii.a implies that $I_{\mathbb{P}}(0) = 0$, so by (5.2) we obtain $I_V(0) = I_R(0)$ and thus $I_V = I_R$.¹² Assume now $\gamma_+ < 0$. By Theorem 0.iii.a and the formulae (1.22) and (1.25), the functions I_V and I_R are finite on $[0, -\gamma_+]$, and equal on $(0, -\gamma_+]$. Now, if (a) holds, then both I_V and I_R are convex, and since they are also lower semicontinuous, their restriction to $[0, \gamma_+]$ is continuous, so $I_R(0) = I_V(0)$, and thus $I_R = I_V$.

¹² Note that the argument when $\gamma_+ = 0$ does not actually use (a); in fact properties (a)–(f) always hold true if $\gamma_+ = 0$.

- (a) \Leftarrow (b) The function I_V is convex by Lemma 5.1.
- (b) \Rightarrow (c) By Theorems B.ii and C.ii, assuming (b) yields $q_V = I_V^* = I_R^* = q_R$.
- (c) \Rightarrow (d) As mentioned in Remarks 1.10 and 1.13, for all $\alpha \leq -1$ we have $q_V(\alpha) = q_{\mathbb{P}}(-1)$ and $q_R(\alpha) = \gamma_+$, so by (c) we find $q_{\mathbb{P}}(-1) = \gamma_+$.
- (d) \Rightarrow (b) Combining (d) and (5.2) gives again $I_V(0) = I_R(0)$.
- (b) \Rightarrow (e) Assuming (b), the identity $I_R(0) = I_V(0)$ reads $-\gamma_+ = \inf_{s \geq 0}(s + I_{\mathbb{P}}(s))$ according to (1.22) and (1.25), so that for all $\epsilon > 0$, there exists $s_\epsilon \geq 0$ such that $I_{\mathbb{P}}(s_\epsilon) < -\gamma_+ - s_\epsilon + \epsilon$. Since also $s_\epsilon \geq -\gamma_+$ by Theorem 0.iii.a, we have

$$0 \leq I_{\mathbb{P}}(s_\epsilon) < -\gamma_+ - s_\epsilon + \epsilon \leq \epsilon.$$

It follows that $s_\epsilon \rightarrow -\gamma_+$ and $I_{\mathbb{P}}(s_\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$. By nonnegativity and lower semicontinuity of $I_{\mathbb{P}}$, this allows to conclude that $I_{\mathbb{P}}(-\gamma_+) = 0$.

- (e) \Rightarrow (f) By Theorem 0.iii.a and (e), we have $I_{\mathbb{P}}(-\gamma_+) = 0$ and $I_{\mathbb{P}}(s) = \infty$ for $s < -\gamma_+$. Since $I_{\mathbb{P}}$ is nonnegative, for all $\alpha \leq 0$, the duality relations (1.14) imply that $q_{\mathbb{P}}(\alpha) = \sup_{s \geq -\gamma_+}(\alpha s - I_{\mathbb{P}}(s)) = -\alpha\gamma_+$, where the supremum is reached at $s = -\gamma_+$.
- (f) \Rightarrow (d) The latter is a special case of the former with $\alpha = -1$.

We now prove the equivalence of (g)–(i).

- (g) \Leftrightarrow (h) This is a consequence of Theorem 0.iii.a.
- (g) \Leftrightarrow (i) This is a consequence of Proposition A.5.i.

To show that (g)–(i) imply (a)–(f) we remark that if (h) holds, then we have (e) because $\inf_{s \in \mathbb{R}} I_{\mathbb{P}}(s) = 0$. Finally, to obtain the last statement, we prove in Proposition 5.5 below that if $\tau_n = O(1)$ and $C_n = O(1)$, then (d) implies (h). \square

Proposition 5.5. *Suppose that \mathbb{P} satisfies UD and SLD with $\tau_n = O(1)$ and $C_n = O(1)$. If $q_{\mathbb{P}}(-1) = \gamma_+$, then $I_{\mathbb{P}}(s) = \infty$ for all $s \in \mathbb{R} \setminus \{-\gamma_+\}$.*

Proof. Seeking a contradiction, suppose that there is $s > -\gamma_+$ such that $I_{\mathbb{P}}(s) < \infty$ (recall that $I_{\mathbb{P}}(s) = \infty$ for all $s < -\gamma_+$ by Theorem 0.iii.a). The idea of the proof is that, if such s exists, then there are words whose probability decreases like $e^{-sn} \ll e^{\gamma_+ n}$, and that these words appear often enough to make $q_{\mathbb{P}}(-1)$ strictly smaller than γ_+ , which contradicts the assumption. Let $\tau := \sup_{n \in \mathbb{N}} \tau_n < \infty$, $C := \sup_{n \in \mathbb{N}} C_n < \infty$, and $\epsilon := s + \gamma_+ > 0$. For $\ell, m \in \mathbb{N}$ and $w \in \mathcal{A}^m$, the computations below will involve

$$R_1 := \frac{\ln(1 - C\mathbb{P}_m(w))}{2(m + \tau)} + \frac{\ln 2}{\ell} + \frac{\ln C}{2\ell} - \frac{\ln(1 - C\mathbb{P}_m(w))}{\ell}$$

and

$$R_2 := \frac{m}{2\ell} \left(-\gamma_+ + \frac{\ln \mathbb{P}_m(w)}{m} - \frac{2\tau\gamma_+}{m} + \frac{3 \ln C}{m} \right).$$

Let us show that m, ℓ , and w can be chosen so that R_1 and R_2 are negative, which will later lead to the contradiction we are seeking. Since $I_{\mathbb{P}}(s) < \infty$, the LDP implies that for all m large enough we can choose a word $w \in \mathcal{A}^m$ such that $\mathbb{P}_m(w) \leq e^{-m(s - \frac{1}{2}\epsilon)}$, and thus $\gamma_+ - \frac{\ln \mathbb{P}_m(w)}{m} \geq \frac{\epsilon}{2}$. We fix m large enough and $w \in \mathcal{A}^m$ so that not only the above holds, but also $-\frac{2\tau\gamma_+}{m} + \frac{3 \ln C}{m} \leq \frac{\epsilon}{4}$ and $e^{-m(s - \frac{1}{2}\epsilon)} < C^{-1}$ (so in particular $1 - C\mathbb{P}_m(w) > 0$). Thus, we have

$$R_2 \leq -\frac{m\epsilon}{8\ell} < 0,$$

and, since the first term in R_1 is negative, taking ℓ large enough yields

$$R_1 \leq \frac{\ln(1 - C\mathbb{P}_m(w))}{4(m + \tau)} < 0.$$

We remark that the above condition on ℓ implies in particular that $\ell > 2(m + \tau)$, which will be useful to have in mind below.

For the remainder of the proof, m , w and ℓ are fixed as above, so that R_1 and R_2 are negative. Let $k \in \mathbb{N}$ and consider the intervals $L_j := [1 + j\ell, (j + 1)\ell - \tau]$ for $j = 0, 1, \dots, 2k - 1$. These $2k$ intervals are separated by gaps of length τ . Let A_k be the set of points $x \in \Omega$ for which at least k out of the $2k$ intervals contain *no* occurrence of w (an interval $[i_1, i_2]$ contains no occurrence of w if there is no $i_1 \leq i \leq i_2 - m + 1$ such that $x_i^{i+m-1} = w$). Clearly,

$$\begin{aligned} \sum_{u \in \mathcal{A}^{2k\ell}} \mathbb{P}_{2k\ell}^2(u) &= \int_{A_k} \mathbb{P}_{2k\ell}([x_1^{2k\ell}]) d\mathbb{P}(x) + \int_{A_k^c} \mathbb{P}_{2k\ell}([x_1^{2k\ell}]) d\mathbb{P}(x) \\ &\leq \mathbb{P}(A_k) \sup_{x \in \Omega} \mathbb{P}_{2k\ell}(x_1^{2k\ell}) + \sup_{x \in A_k^c} \mathbb{P}_{2k\ell}(x_1^{2k\ell}). \end{aligned} \tag{5.3}$$

We will show below that

$$\limsup_{k \rightarrow \infty} \frac{1}{2k\ell} \ln(\mathbb{P}(A_k) \sup_{x \in \Omega} \mathbb{P}_{2k\ell}(x_1^{2k\ell})) \leq \gamma_+ + R_1, \tag{5.4}$$

$$\limsup_{k \rightarrow \infty} \frac{1}{2k\ell} \ln \sup_{x \in A_k^c} \mathbb{P}_{2k\ell}(x_1^{2k\ell}) \leq \gamma_+ + R_2. \tag{5.5}$$

Combining this with (5.3) yields, since the limit defining $q_{\mathbb{P}}(-1)$ exists,

$$\begin{aligned} q_{\mathbb{P}}(-1) &= \lim_{n \rightarrow \infty} \frac{1}{n} \ln \sum_{u \in \mathcal{A}^n} \mathbb{P}_n^2(u) \\ &= \limsup_{k \rightarrow \infty} \frac{1}{2k\ell} \sum_{u \in \mathcal{A}^{2k\ell}} \mathbb{P}_{2k\ell}^2(u) \\ &\leq \max \{ \gamma_+ + R_2, \gamma_+ + R_1 \} \\ &< \gamma_+, \end{aligned}$$

which contradicts the assumption, and thus proves the claim.

Proof of (5.4). By Remark B.5, since $\tau_n = O(1)$, we can assume without loss of generality that the sequence $(C_n)_{n \in \mathbb{N}}$ required for (3.1) and (3.2) to hold, is also bounded by C .

The event “there is no occurrence of w in the interval L_j ” can be written as $T^{-j\ell}U$, where $U := \{x : W_m(w, x) > \ell + 1 - m - \tau\} \in \mathcal{F}_{\ell - \tau}$. By the definition of A_k ,

$$A_k \subseteq \bigcup_{\substack{M \subseteq \{0, 1, \dots, 2k-1\} \\ |M|=k}} \bigcap_{j \in M} T^{-j\ell}U. \tag{5.6}$$

Using the bound (3.4) of Lemma 3.1, we obtain

$$\mathbb{P}(U) \leq (1 - C\mathbb{P}_m(w))^{\lfloor \frac{\ell+1-m-\tau}{m+\tau} \rfloor} \leq (1 - C\mathbb{P}_m(w))^{\frac{\ell}{m+\tau} - 2}.$$

Since $U \in \mathcal{F}_{\ell-\tau}$, we can apply **UD** recursively $k - 1$ times to the intersection in (5.6). This and a simple union bound (noting that the union in (5.6) contains $\binom{2k}{k}$ terms) yields

$$\mathbb{P}(A_k) \leq \binom{2k}{k} C^{k-1} (1 - C \mathbb{P}_m(w))^{\frac{k\ell}{m+\tau} - 2k}.$$

Since $\lim_{n \rightarrow \infty} \frac{1}{2k} \ln \binom{2k}{k} = \ln 2$, we have proved that $\limsup_{k \rightarrow \infty} \frac{1}{2k\ell} \ln \mathbb{P}(A_k) \leq R_1$. Combining this with the definition of γ_+ completes the proof of (5.4).

Proof of (5.5). First, let $a_n := \sup_{u \in \mathcal{A}^n} \ln \mathbb{P}_n(u)$. Clearly, the sequence $(a_n)_{n \in \mathbb{N}}$ is nonincreasing, and by **SLD**, for all $n, m \in \mathbb{N}$, there is $0 \leq \ell \leq \tau$ such that $a_{n+m} \geq a_{n+m+\ell} \geq a_n + a_m - \ln C$. By applying Fekete’s lemma to the sequence $(\ln C - a_n)_{n \in \mathbb{N}}$, we obtain $\gamma_+ = \sup_{n \in \mathbb{N}} \frac{a_n - \ln C}{n}$, so for every $x \in \Omega$ and $n \in \mathbb{N}$,

$$\ln \mathbb{P}_n(x_1^n) \leq n\gamma_+ + \ln C. \tag{5.7}$$

Let $x \in A_k^c$. By construction, we can choose k occurrences of w in $x_1^{2k\ell}$, which are pairwise separated by gaps of length at least τ (possible additional occurrences of w in $x_1^{2k\ell}$ do not play any role in the argument). In other words, we can write $x_1^{2k\ell} = u_0 w u_1 w u_2 \dots u_{k-1} w u_k$ where $|u_0| \geq 0$ and $|u_i| \geq \tau$ for all $1 \leq i \leq k$. We then apply **UD** before and after each occurrence of w of $x_1^{2k\ell}$.¹³ More precisely, in order to apply **UD**, we discard a prefix of length τ in each u_i with $1 \leq i \leq k$, and we discard a suffix of length τ in each u_i with $0 \leq i \leq k - 1$; the u_i whose length is insufficient for this to be done are discarded entirely. By using (5.7) to bound the probability of what remains of the u_i at the end of this process, we easily obtain the estimate

$$\mathbb{P}_{2k\ell}(x_1^{2k\ell}) \leq \mathbb{P}_m(w)^k C^{3k+1} e^{\gamma_+(2k\ell - km - 2k\tau)},$$

where the C^{3k+1} comes from the fact that we have used **UD** at most $2k$ times and (5.7) at most $k + 1$ times, and where $2k\ell - km - 2k\tau$ is the minimum cumulative length of the words to which we have applied (5.7). By the definition of R_2 , this establishes (5.5), and thus the proof is complete. \square

6. Relation to Almost Sure Convergence Results

In this section we discuss the complementarity of our LDPs and existing almost sure convergence results in the literature.

6.1. A brief review of results on almost sure convergence. Throughout this subsection, we consider two measures $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_{\text{inv}}$, and we only assume that \mathbb{Q} satisfies **UD** and **SLD**. Under these conditions, the limits

$$h_{\mathbb{P}}(x) := \lim_{n \rightarrow \infty} -\frac{1}{n} \ln \mathbb{P}_n(x_1^n) \tag{6.1}$$

and

$$h_{\mathbb{Q}}(x) := \lim_{n \rightarrow \infty} -\frac{1}{n} \ln \mathbb{Q}_n(x_1^n) \tag{6.2}$$

¹³ Remark that decoupling is not necessarily used near the boundaries of the intervals L_j .

exist for \mathbb{P} -almost all x . Moreover,

$$\int h_{\mathbb{P}}(x) d\mathbb{P}(x) = \lim_{n \rightarrow \infty} -\frac{1}{n} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \ln \mathbb{P}_n(u) =: h(\mathbb{P}) \tag{6.3}$$

and

$$\int h_{\mathbb{Q}}(x) d\mathbb{P}(x) = \lim_{n \rightarrow \infty} -\frac{1}{n} \sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \ln \mathbb{Q}_n(u) =: h_c(\mathbb{P}|\mathbb{Q}). \tag{6.4}$$

The relations (6.1) and (6.3) are the contents of the well-known SMB theorem. While often stated assuming ergodicity, the general form of the SMB theorem, which only requires shift invariance, is obtained as a special case of the Brin–Katok formula [BK83]; see also Theorem 9.3.1 of [VO16] for a direct proof.

On the other hand, (6.2) and (6.4) are much less general, and the literature on them is comparatively sparse. Even the limit defining $h_c(\mathbb{P}|\mathbb{Q})$ in (6.4) fails to exist in some cases; see e.g. [vEFS93, §A.5.2]. In our setup, the almost sure existence of the limit (6.2) and the validity of (6.4) are guaranteed by the assumption UD imposed on \mathbb{Q} . Indeed, as discussed in [CDEJR23a, §5], the assumption implies that the functions $(f_n)_{n \in \mathbb{N}}$ defined by $f_n(x) = \ln \mathbb{Q}_n(x_1^n)$ satisfy the gapped almost subadditivity condition of the adaptation of Kingman’s theorem given in [Raq23, §2].

Still assuming only that \mathbb{Q} satisfies UD and SLD (and that \mathbb{P} is merely shift invariant), the random variables studied in our main theorems are known to satisfy the following almost sure identities, which, as mentioned in the introduction, justify their role as entropy estimators: for \mathbb{P} -almost every x , we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln R_n(x) = h_{\mathbb{P}}(x) \tag{6.5}$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln V_n(x) = h_{\mathbb{P}}(x), \tag{6.6}$$

and, for $(\mathbb{P} \otimes \mathbb{Q})$ -almost every (x, y) , we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln W_n(x, y) = h_{\mathbb{Q}}(x). \tag{6.7}$$

The convergences expressed in (6.5) and (6.6) were first proved under the assumption that \mathbb{P} is ergodic, in which case $h_{\mathbb{P}}(x) = h(\mathbb{P})$ for \mathbb{P} -almost every x ; see [WZ89, OW93, Kon98]. However, combining Remark 1 of [Kon98, §2] with a generalization of Kac’s lemma, as found e.g. in [VO16, §1.2.2], shows that shift invariance suffices for (6.5). Since $R_n(x) \leq V_n(x) + n - 1$ and $V_n(x) = R_n(T^{m_n(x)-1}x) - (n - m_n(x))$ with $m_n(x) := \max\{1 \leq i \leq n : x_i^{i+n-1} = x_1^n\}$, it is straightforward to show that the bounds used in [Kon98, §2] to prove (6.5) via the Borel–Cantelli lemma also imply (6.6). The convergence expressed in (6.7) was first proved for Markov measures, and then under various strong mixing assumptions; see [WZ89, Shi93, MS95, Kon98]. However, it is known that (6.7) may fail for some mixing measures, even when $\mathbb{P} = \mathbb{Q}$; see [Shi93]. In the present setup, the SLD assumption satisfied by \mathbb{Q} and (6.2) have been shown to imply (6.7) in [CDEJR23a, §2].

6.2. *Zeros of the rate functions.* In this subsection, we discuss how the LDPs of Theorems 0 and A–C complement the above almost sure results. We assume throughout that the pair (\mathbb{P}, \mathbb{Q}) is admissible, but not necessarily that \mathbb{P} satisfies PA. The central role is held by the set

$$J := \{s : I_{\mathbb{Q}}(s) = 0\} \subseteq [0, \infty), \tag{6.8}$$

which, by convexity of $I_{\mathbb{Q}}$ (see Theorem 0.i), is an interval and coincides with the subdifferential of $q_{\mathbb{Q}}$ at 0. Since $I_{\mathbb{Q}}$ is lower semicontinuous, and since $\inf_{s \in \mathbb{R}} I_{\mathbb{Q}}(s) = 0$, only the three following cases are possible: $J = [a, b]$ for some $0 \leq a \leq b < \infty$, $J = [a, \infty)$ for some $a \geq 0$, and $J = \emptyset$, which happens when $I_{\mathbb{Q}}(s) > 0$ for all $s \in \mathbb{R}$ and $\lim_{s \rightarrow \infty} I_{\mathbb{Q}}(s) = 0$. When $\mathbb{Q} = \mathbb{P}$, we necessarily have $J = [a, b]$ since $I_{\mathbb{P}}$ is a good rate function (recall Theorem 0.iii).

Lemma 6.1. *Let (\mathbb{P}, \mathbb{Q}) be admissible. Then, the rate function I_W of (1.19) vanishes precisely on J . Moreover, if $\mathbb{Q} = \mathbb{P}$, the same is true of the rate functions I_V of (1.22) and I_R of (1.25).*

Proof. The statement for I_W is an immediate consequence of (1.19). Appealing to Remark 1.11, the statement for I_V is immediate. In order to extend the result to I_R , it remains to note that $I_R(0) = 0$ if and only if $\gamma_+ = 0$, which, by (1.15), happens if and only if $I_{\mathbb{P}}(0) = 0$, i.e. if and only if $0 \in J$. \square

We briefly comment on the simple case where J is a singleton, which occurs in particular in Examples 2.1–2.3, and which corresponds to the property of *exponential rates for entropy* in the terminology of [Shi96, Chap. III], [JB13] and [CRS18]. In this case, we obtain from Lemma 6.1 that the rate functions of Theorems 0 and A–C all vanish at a single point, say a . In this situation, a standard argument shows that, almost surely, all limits in (6.2) and (6.5)–(6.7) exist and are equal to a ; see e.g. [EII06, §II.6] and the argument in the proof of Proposition 6.2.i below. In this simple case, our LDPs thus imply the almost sure convergence results reviewed in Sect. 6.1.

When J is an interval of positive length, our LDPs do not imply these almost sure convergence results anymore. Still, all displayed equations in Sect. 6.1 remain valid under our assumptions (see the references given there), and we show in the following two propositions that the limiting values are constrained by J .

Proposition 6.2. *Let (\mathbb{P}, \mathbb{Q}) be an admissible pair. Then, the following hold:*

- i. *If J is nonempty, then $h_{\mathbb{Q}}(x) \in J$ for \mathbb{P} -almost all x and $h_c(\mathbb{P}|\mathbb{Q}) \in J$.*
- ii. *If J is empty, then $h_{\mathbb{Q}}(x) = \infty$ for \mathbb{P} -almost all x and $h_c(\mathbb{P}|\mathbb{Q}) = \infty$.*
- iii. *If J is a singleton, then $J = \{h_c(\mathbb{P}|\mathbb{Q})\}$, $h_{\mathbb{Q}}(x) = h_c(\mathbb{P}|\mathbb{Q})$ for \mathbb{P} -almost all x , and the convergence in (6.2) (resp. (6.7)) occurs in $L^p(\mathbb{P})$ (resp. $L^p(\mathbb{P} \otimes \mathbb{Q})$) for all $p \geq 1$.*

Proof. We proceed with the proof, taking (6.4), (6.7) as well as the almost sure existence of the limit in (6.2) for granted.

- i. Suppose that the interval J is nonempty. Then, by convexity,

$$\inf_{s: \text{dist}(s, J) \geq \delta} I_{\mathbb{Q}}(s) > 0$$

for every $\delta > 0$. Therefore, by the large-deviation upper bound, the probability $\mathbb{P}\{x : \text{dist}(-\frac{1}{n} \ln \mathbb{Q}_n(x_1^n), J) \geq \delta\}$ decreases exponentially fast and a standard application of the Borel–Cantelli lemma, followed by taking the limit $\delta \rightarrow 0$, implies that $\inf J \leq h_{\mathbb{P}}(x) \leq \sup J$ for \mathbb{P} -almost every x . By (6.4), we obtain $h_c(\mathbb{P}|\mathbb{Q}) \in J$.

- ii. If $J = \emptyset$, a similar argument applies to $\mathbb{P}\{x : -\frac{1}{n} \ln Q_n(x_1^n) \leq K\}$ with K arbitrarily large, forcing $h_Q(x)$ to properly diverge to ∞ for \mathbb{P} -almost every x . By (6.4), we conclude that also $h_c(\mathbb{P}|Q) = \infty$.
- iii. If J is a singleton, by Part i we conclude that $J = \{h_c(\mathbb{P}|Q)\}$ and that $h_Q(x) = h_c(\mathbb{P}|Q)$ for \mathbb{P} -almost all x . We now claim that there exists $\alpha > 0$ such that $q_Q(\alpha) < \infty$. Indeed, if this were not the case, we would have $I_Q(s) = q_Q^*(s) = \sup_{\alpha \leq 0} (\alpha s - q_Q(\alpha))$ for all $s \in \mathbb{R}$, so I_Q would be nonincreasing, which contradicts the assumption that J is a singleton. For such α , note that the function $\phi : s \mapsto \exp(\alpha s^{1/p})$ grows superlinearly and that

$$\limsup_{n \rightarrow \infty} \int \phi \left(\left| \frac{1}{n} \ln Q_n \right|^p \right) d\mathbb{P} \leq \exp(q_Q(\alpha)) \tag{6.9}$$

by Jensen’s inequality applied to $\sqrt[p]{\cdot}$. Hence, de la Vallée Poussin’s criterion guarantees the sufficient uniform integrability requirement for the convergence in (6.2) to hold in $L^p(\mathbb{P})$. The exact same argument applies to (6.7): we just need to replace (6.9) with

$$\limsup_{n \rightarrow \infty} \int \phi \left(\left(\frac{1}{n} \ln W_n \right)^p \right) d(\mathbb{P} \otimes Q) \leq \exp(q_W(\alpha))$$

and to recall that, by (1.20) and since $\alpha > 0$, we have $q_W(\alpha) = q_Q(\alpha) < \infty$. □

In the next proposition, we consider the special case where $Q = \mathbb{P}$, which allows to include the limits (6.5) and (6.6) into the discussion. Notice also that even when discussing (6.5), the measure \mathbb{P} is not assumed to satisfy PA.

Proposition 6.3. *Assume \mathbb{P} satisfies UD and SLD, and let J be defined by (6.8) with $Q = \mathbb{P}$. Then, the following hold:*

- i. $J \subseteq [-\gamma_+, -\gamma_-]$.
- ii. $h_{\mathbb{P}}(x) \in J$ for \mathbb{P} -almost every x , and $h(\mathbb{P}) \in J$.
- iii. If J is a singleton, then $J = \{h(\mathbb{P})\}$ and $h_{\mathbb{P}}(x) = h(\mathbb{P})$ for \mathbb{P} -almost every x . Moreover, for all $p \geq 1$, the convergence in (6.1), (6.5) and (6.6) holds in $L^p(\mathbb{P})$, and that of (6.7) holds in $L^p(\mathbb{P} \otimes \mathbb{P})$.

Proof. Part i is immediate by (1.15) and Part ii is a special case of Proposition 6.2.i. All statements in Part iii except the one about (6.5) are either special cases of Proposition 6.2.iii or proved in exactly the same way (possibly using the simplification offered by (1.16)).

The statement about (6.5) requires additional considerations to lift PA. As discussed at the end of Sect. 4.2, the large-deviation upper bound of Theorem C still holds without assuming PA, and we have $\bar{q}_R \leq I_R^*$, with \bar{q}_R defined by taking the limit superior in (1.4). Then, (6.9) is simply replaced by

$$\limsup_{n \rightarrow \infty} \int \phi \left(\left(\frac{1}{n} \ln R_n \right)^p \right) d\mathbb{P} \leq \exp(\bar{q}_R(\alpha)).$$

Since $\bar{q}_R(\alpha) \leq I_R^*(\alpha) = q_{\mathbb{P}}(\alpha)$ for all $\alpha \geq 0$ by (1.26), the same argument applies once more. □

Remark 6.4. The \mathbb{P} -essential image of $h_{\mathbb{P}}(x)$ can be a strict subset of J .

Acknowledgement The authors would like to thank Tristan Benoist, Jean-René Chazottes, Vaughn Climenhaga, Bastien Fernandez, Jérôme Rousseau and Charles-Édouard Pfister for stimulating discussions, relevant references and valuable advice about the manuscript. They are specifically grateful to Vojkan Jakšić for introducing them to this problem in the framework of entropy information theory lectures at McGill University in 2020, for which RR was a course assistant. RR acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada and from the Fonds de recherche du Québec—Nature et technologies. Part of this work was done while RR was a post-doctoral researcher at CY Cergy Paris Université and supported by LabEx MME-DII (Investissements d’Avenir program of the French government). Part of this work was done during a stay of both authors at the Centre de recherches mathématiques and McGill University supported by the Agence Nationale de la Recherche of the French government through the grant NONSTOPS (ANR-17-CE40-0006) and by the Centre de recherches mathématiques through a CNRS-CRM “Appel à mobilité Québec–France” grant.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

A. Supports and Subshifts

In this short appendix, we collect some basic definitions and technical results from the theory of dynamical systems, which are used at several places in the paper.

We recall that a nonempty set $\Omega' \subseteq \Omega$ is called a *subshift* of Ω if it is closed (with respect to the product topology) and if it satisfies $T(\Omega') \subseteq \Omega'$. The *language* of Ω' is defined as $\mathcal{L} := \{x_n^m : 1 \leq n \leq m, x \in \Omega'\}$. The subshift Ω' is called *transitive* if, whenever $u \in \mathcal{L}$ and $v \in \mathcal{L}$, there exists $\xi \in \mathcal{L}$ such that $u\xi v \in \mathcal{L}$.

A nonempty set $\Omega' \subseteq \Omega$ is a subshift if and only if there exists a collection $\mathcal{W} \subset \Omega_{\text{fin}}$ of *forbidden words* such that Ω' consists of those $x \in \Omega$ such that no word $x_n^m, 1 \leq n \leq m$, belongs to \mathcal{W} . For a given subshift Ω' , the set \mathcal{W} is not uniquely determined, but if it can be chosen as a finite set, then Ω' is called a *subshift of finite type*. In the further specialized case where \mathcal{W} can be chosen as a subset of \mathcal{A}^2 , then Ω' is called a *Markov subshift*.

We now fix $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ and consider $\Omega' := \text{supp } \mathbb{P} = \{x \in \Omega : \mathbb{P}_n(x) > 0 \text{ for all } n \in \mathbb{N}\}$. It is clear that Ω' is a subshift whose language is given by $\mathcal{L} = \{u \in \Omega_{\text{fin}} : \mathbb{P}([u]) > 0\}$, and that actually $T(\Omega') = \Omega'$.

The topological entropy of Ω' is defined by

$$h_{\text{top}}(\Omega') := \lim_{n \rightarrow \infty} \frac{1}{n} \ln |\{x_1^n : x \in \Omega'\}| = \lim_{n \rightarrow \infty} \frac{1}{n} \ln |\text{supp } \mathbb{P}_n| \leq \ln |\mathcal{A}|, \quad (\text{A.1})$$

where we recall that $\text{supp } \mathbb{P}_n := \{u \in \mathcal{A}^n : \mathbb{P}_n(u) > 0\}$. The limits in (A.1) exist by a standard subadditivity argument (which does not require any decoupling assumption).

Lemma A.1. *Let $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$. Then, the bounds (1.13) hold.*

Proof. Obviously, $0 \leq -\gamma_+$. Let now $H(\mathbb{P}_n) := -\sum_{u \in \mathcal{A}^n} \mathbb{P}_n(u) \ln \mathbb{P}_n(u)$. By the definition of γ_+ we find $\ln \mathbb{P}_n(u) \leq n\gamma_+ + o(n)$ for every $u \in \mathcal{A}^n$ and so $-n\gamma_+ \leq H(\mathbb{P}_n) + o(n)$. On the other hand, $H(\mathbb{P}_n) \leq \ln |\text{supp } \mathbb{P}_n|$ (see e.g. Corollary 4.2.1 in [Wal82]). Finally, clearly $\min_{u \in \text{supp } \mathbb{P}_n} \mathbb{P}_n(u) \leq 1/|\text{supp } \mathbb{P}_n|$, so $\ln |\text{supp } \mathbb{P}_n| \leq -\min_{u \in \text{supp } \mathbb{P}_n} \ln \mathbb{P}_n(u)$. Dividing the above inequalities by n and taking $n \rightarrow \infty$ completes the proof of (1.13). Remark that the inequality $h(\mathbb{P}) \leq h_{\text{top}}(\text{supp } \mathbb{P})$ also follows from the much more general result (A.2) below. □

A.1. Specification properties. The following definitions will be especially helpful when combined with Lemmas B.7 and B.8 below. As in the statement of our decoupling conditions, we shall assume that $\tau_n = o(n)$, and there will be no loss of generality assuming that the various assumptions hold with the same sequence $(\tau_n)_{n \in \mathbb{N}}$. Again, the case where the sequence is bounded (written $\tau_n = O(1)$) will sometimes be singled out.

Definition A.2. For a subshift $\Omega' \subset \Omega$ with language \mathcal{L} , we define the following properties:

- the subshift Ω' satisfies the *flexible specification property* if for every $n \in \mathbb{N}$, $u \in \mathcal{L} \cap \mathcal{A}^n$, and every $v \in \mathcal{L}$, there exists $0 \leq \ell \leq \tau_n$ and $\xi \in \mathcal{A}^\ell$ such that $u\xi v \in \mathcal{L}$;
- the subshift Ω' satisfies the *abundant periodic orbit property* if for each $n \in \mathbb{N}$ and each $u \in \mathcal{L} \cap \mathcal{A}^n$, the cylinder $[u]$ contains at least one periodic point of period no larger than $n + \tau_n$, i.e. if there exists a word ξ with $0 \leq |\xi| \leq \tau_n$ such that $u\xi u\xi u\xi \cdots \in \Omega'$.

Remark A.3. Obviously, if \mathbb{P} satisfies SLD, then $\text{supp } \mathbb{P}$ satisfies the flexible specification property (with the same sequence $(\tau_n)_{n \in \mathbb{N}}$). In the same way, if a subshift Ω' satisfies Bowen’s specification property [Bow74], then the two properties in (A.2) are satisfied with $\tau_n = O(1)$. The same is true when Ω' is a transitive subshift of finite type: the flexible specification property (with $\tau_n = O(1)$) is quite immediate, and the abundant periodic orbit property is then obtained using Lemma A.4 below. The interested reader can gather other sufficient conditions by combining ideas from [Jun11, KŁO16, CT21].

Lemma A.4. *If the flexible specification property holds with $\tau_n = O(1)$, then the abundant periodic orbit property holds with some (possibly larger) $\tau'_n = O(1)$.*

Proof. Assume that the flexible specification property holds with $\tau := \sup_{n \in \mathbb{N}} \tau_n < \infty$. Following [Ber88] and [Jun11, §3], this implies the existence of a *synchronizing word* w , i.e. a word $w \in \mathcal{L}$ such that whenever $uw \in \mathcal{L}$ and $wv \in \mathcal{L}$, then $uvw \in \mathcal{L}$. Using the assumption twice, for every $u \in \mathcal{L}$, we can find ξ_1 and ξ_2 , each of length at most τ , such that $w\xi_1 u\xi_2 w \in \mathcal{L}$. Note that $|\xi_1 w \xi_2| \leq 2\tau + |w|$. But by the synchronizing property of w , we must have $u\xi_2 w\xi_1 u\xi_2 w\xi_1 u\xi_2 w\xi_1 \cdots \in \Omega'$, so the abundant periodic orbit property holds with $\tau'_n := 2\tau + |w| = O(1)$. \square

A.2. Measures of maximal entropy. Let Ω' be a subshift. We denote by $\mathcal{P}_{\text{inv}}(\Omega')$ the set of shift-invariant probability measures on Ω' , and we identify the elements of $\mathcal{P}_{\text{inv}}(\Omega')$ with those of $\{\mu \in \mathcal{P}_{\text{inv}}(\Omega) : \mu(\Omega') = 1\}$ in the obvious way.

By the variational principle (see e.g. Theorem 8.6 in [Wal82]) applied to the dynamical system (Ω', T) , we have

$$h_{\text{top}}(\Omega') = \max_{\mu \in \mathcal{P}_{\text{inv}}(\Omega')} h(\mu), \tag{A.2}$$

where the right-hand side is indeed a maximum, since the map h is upper semicontinuous (see e.g. Theorem 8.2 in [Wal82]). In particular $h(\mu) \leq h_{\text{top}}(\Omega')$ for all $\mu \in \mathcal{P}_{\text{inv}}(\Omega')$, and μ is called a *measure of maximal entropy* (MME) on Ω' if $h(\mu) = h_{\text{top}}(\Omega')$.

Given a continuous potential $\varphi : \Omega' \rightarrow \mathbb{R}$, we denote by $p_{\text{top}}(\varphi)$ the topological pressure of φ with respect to the dynamical system (Ω', T) , that is,

$$p_{\text{top}}(\varphi) := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \sum_{u \in \mathcal{A}^n \cap \mathcal{L}} e^{\sup_{x \in [u] \cap \Omega'} \sum_{i=0}^{n-1} \varphi(T^i x)},$$

which reduces to (2.10) when $\Omega' = \Omega$. With this in mind, noting that $h_{\text{top}}(\Omega')$ is nothing but $p_{\text{top}}(\varphi_0)$ for the potential $\varphi_0 : \Omega' \rightarrow \mathbb{R}$ that vanishes identically, we conclude that $\mu \in \mathcal{P}_{\text{inv}}(\Omega')$ is a MME on Ω' if and only if $p_{\text{top}}(\varphi_0) = h(\mu)$, that is, if and only if μ is an equilibrium measure for φ_0 .

Proposition A.5. *Let $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$, and let $\Omega' := \text{supp } \mathbb{P}$.*

- i. Assume \mathbb{P} satisfies SLD. Then, the measure \mathbb{P} is a MME on Ω' if and only if $\gamma_- = \gamma_+$.*
- ii. Assume \mathbb{P} satisfies SLD with $\tau_n = O(1)$. Then, the dynamical system (Ω', T) is intrinsically ergodic, i.e. there is exactly one MME on Ω' .*

Proof. Recall that γ_{\pm} are defined in (1.12).

- i. Assume \mathbb{P} is a MME on Ω' . As discussed above, \mathbb{P} is then an equilibrium measure for the identically vanishing potential φ_0 defined on Ω' . By SLD, Ω' enjoys the flexible specification property (Remark A.3) and thus we can apply Theorem 2.1 in [PS20], which guarantees that \mathbb{P} is weak Gibbs with respect to φ_0 , i.e. that for all $x \in \Omega'$,

$$\mathbb{P}_n(x_1^n) = e^{\sum_{i=0}^{n-1} \varphi_0(T^i x) - np_{\text{top}}(\varphi_0) + o(n)} = e^{-nh_{\text{top}}(\Omega') + o(n)}.$$

From this we obtain $\gamma_+ = \gamma_- = h_{\text{top}}(\Omega')$. Conversely, if $\gamma_- = \gamma_+$, then (1.13) implies that $h(\mathbb{P}) = h_{\text{top}}(\Omega')$, so \mathbb{P} is a MME.

- ii. Under the stated assumption, Ω' satisfies the flexible specification property with $\tau_n = O(1)$ (Remark A.3) and the conclusion follows from e.g. Theorem 3.2 in [CT21]. □

B. Technical Results About Decoupling Assumptions

This appendix complements the discussion of our assumptions in Sect. 1.1 and uses the notations therein.

B.1. Alternative definitions. The following lemma is used in Remark 1.3.

Lemma B.1. *If any of the assumptions UD, SLD or JSLD holds with some sequences $(C_n)_{n \in \mathbb{N}}$ and $(\tau_n)_{n \in \mathbb{N}}$, then it holds with any choice of sequences $(C'_n)_{n \in \mathbb{N}}$ and $(\tau'_n)_{n \in \mathbb{N}}$ satisfying $C'_n \geq C_n$ and $\tau'_n \geq \tau_n$ for all $n \in \mathbb{N}$.*

Proof. The only nontrivial part of the statement is that we may replace τ_n with τ'_n in UD, which we now prove. Let $n, m \in \mathbb{N}$, $u \in \mathcal{A}^n$, $v \in \mathcal{A}^m$ and $\xi \in \Omega^{\tau'_n}$. UD then implies, with $\xi' = \xi_{\tau'_n+1}^{\tau'_n}$ (possibly the empty word if $\tau'_n = \tau_n$) that $\mathbb{P}_{n+\tau'_n+m}(u\xi v) \leq C_n \mathbb{P}_n(u) \mathbb{P}_{m+\tau'_n-\tau_n}(\xi' v) \leq C_n \mathbb{P}_n(u) \mathbb{P}_m(v)$. □

The next two lemmas justify (3.1) and (3.2) respectively.

Lemma B.2. *If $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ satisfies UD, then*

$$\mathbb{P}(A \cap T^{-n-\tau_n} B) \leq \tilde{C}_n \mathbb{P}(A) \mathbb{P}(B)$$

for every $n \in \mathbb{N}$, $A \in \mathcal{F}_n$ and $B \in \mathcal{F}$, with $\tilde{C}_n := C_n |\mathcal{A}|^{\tau_n} = e^{o(n)}$.

Proof. First, since the semi-algebra of cylinder sets generates \mathcal{F} , and since $B \mapsto \mathbb{P}(A \cap T^{-n-\tau_n} B)$ defines a finite measure, it suffices to consider the case where $B \in \mathcal{F}_m$ for some $m \in \mathbb{N}$ (see e.g. Lemma 1.9.4 in [Bog07]). But then, by **UD**,

$$\begin{aligned} \mathbb{P}(A \cap T^{-n-\tau_n} B) &= \sum_{\substack{u \in \mathcal{A}^n: [u] \subseteq A \\ v \in \mathcal{A}^m: [v] \subseteq B \\ \xi \in \mathcal{A}^{\tau_n}}} \mathbb{P}_{n+\tau_n+m}(u\xi v) \\ &\leq C_n \sum_{\substack{u \in \mathcal{A}^n: [u] \subseteq A \\ v \in \mathcal{A}^m: [v] \subseteq B \\ \xi \in \mathcal{A}^{\tau_n}}} \mathbb{P}_n(u)\mathbb{P}_m(v) \\ &= C_n \sum_{\xi \in \mathcal{A}^{\tau_n}} \mathbb{P}(A)\mathbb{P}(B), \end{aligned}$$

from which the result follows, since the sum over ξ contains $|\mathcal{A}|^{\tau_n}$ terms. □

Lemma B.3. *If $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ satisfies **SLD**, then*

$$\max_{0 \leq \ell \leq \tau_n} \mathbb{P}(A \cap T^{-n-\ell} B) \geq \widehat{C}_n^{-1} \mathbb{P}(A)\mathbb{P}(B)$$

for every $n \in \mathbb{N}$, $A \in \mathcal{F}_n$ and $B \in \mathcal{F}$, with $\widehat{C}_n := (\tau_n + 1)C_n = e^{o(n)}$.

Proof. Consider the finite measure μ defined by $\mu(B) = \sum_{\ell=0}^{\tau_n} \mathbb{P}(A \cap T^{-n-\ell} B)$. If $B \in \mathcal{F}_m$ for some $m \in \mathbb{N}$, we obtain by **SLD**,

$$\begin{aligned} \mu(B) &= \sum_{\substack{u \in \mathcal{A}^n: [u] \subseteq A \\ v \in \mathcal{A}^m: [v] \subseteq B \\ \xi \in \bigcup_{\ell=0}^{\tau_n} \mathcal{A}^\ell}} \mathbb{P}_{n+|\xi|+m}(u\xi v) \\ &\geq C_n^{-1} \sum_{\substack{u \in \mathcal{A}^n: [u] \subseteq A \\ v \in \mathcal{A}^m: [v] \subseteq B}} \mathbb{P}_n(u)\mathbb{P}_m(v) \\ &= C_n^{-1} \mathbb{P}(A)\mathbb{P}(B). \end{aligned}$$

As in Lemma B.2, this extends to any $B \in \mathcal{F}$. Since

$$\mu(B) \leq (\tau_n + 1) \max_{0 \leq \ell \leq \tau_n} \mathbb{P}(A \cap T^{-n-\ell} B)$$

by definition, the proof is complete. □

Remark B.4. It is obvious that the condition stated in Lemma B.2 is actually equivalent to **UD**, by considering $A = [u]$ and $B = [v]$. In the same way, the condition stated in Lemma B.2 is equivalent to **SLD**, although going back to the original statement of **SLD** costs an additional (harmless) factor of $|\mathcal{A}|^{\tau_n}$, since the inequality $\mathbb{P}([u] \cap T^{-n-\ell}[v]) \geq \widehat{C}_n^{-1} \mathbb{P}([u])\mathbb{P}([v])$ only implies that there exists some $\xi \in \mathcal{A}^\ell$ such that $\mathbb{P}([u] \cap T^{-n}[\xi] \cap T^{-n-\ell}[v]) \geq \widehat{C}_n^{-1} |\mathcal{A}|^{-\ell} \mathbb{P}([u])\mathbb{P}([v])$.

Remark B.5. In general, boundedness of the sequence $(C_n)_{n \in \mathbb{N}}$ does not imply boundedness of the sequences $(\widetilde{C}_n)_{n \in \mathbb{N}}$ and $(\widehat{C}_n)_{n \in \mathbb{N}}$. However, this implication holds when $\tau_n = O(1)$.

B.2. Sufficient conditions. We start with a sufficient condition for **PA** to hold. In a nutshell, the sufficient condition is that for every $p \in \mathbb{N}$, $v \in \mathcal{A}^p$, there is some $\xi \in \Omega_{\text{fin}}$ such that the probability of $(v\xi)^{n-1}v$ satisfies a lower bound in terms of $\mathbb{P}_p(v)^n$. We stress that **SLD** does not imply such a lower bound in general. Indeed, applying **SLD** a total of $n - 1$ times recursively yields a word $v\xi^{(n-1)}v \dots v\xi^{(2)}v\xi^{(1)}v$ whose probability is bounded below by $C_p^{n-1}\mathbb{P}_p(v)^n$, but there is no reason for the $\xi^{(i)}$ to be equal in general, unless of course $\tau_n = 0$ for all n , in which case **SLD** does imply the assumptions of Lemma **B.6** and thus **PA**.

Lemma B.6. *Let $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ and assume that there exists a family $(c_{p,n})_{p \in \mathbb{N}, n \in \mathbb{N}}$ satisfying*

$$\lim_{p \rightarrow \infty} \frac{1}{p} \limsup_{n \rightarrow \infty} \frac{\ln c_{p,n}}{n} = 0, \tag{B.1}$$

and such that for every $p \in \mathbb{N}$ and $v \in \mathcal{A}^p$, there exists $\xi \in \Omega_{\text{fin}}$ such that

$$\mathbb{P}_{np+(n-1)|\xi|} \left((v\xi)^{n-1}v \right) \geq c_{p,n}(\mathbb{P}_p(v))^n \tag{B.2}$$

*for all $n \in \mathbb{N}$. Then, the measure \mathbb{P} satisfies **PA**.*

Proof. Let $\epsilon > 0$ and $P \in \mathbb{N}$ be arbitrary. By definition of γ_+ , there exists $p \geq P$ and $v \in \mathcal{A}^p$ such that $\mathbb{P}_p(v) > e^{p(\gamma_+ - \frac{1}{2}\epsilon)}$. With $\xi \in \Omega_{\text{fin}}$ as in (B.2), we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n(p + |\xi|)} \ln \mathbb{P}_{n(p+|\xi|)}((v\xi)^n) &\geq \liminf_{n \rightarrow \infty} \frac{1}{np} \ln \mathbb{P}_{n(p+|\xi|)}((v\xi)^n) \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{np} \ln \mathbb{P}_{(n+1)p+n|\xi|}((v\xi)^n v) \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{np} ((n+1) \ln \mathbb{P}_p(v) + \ln c_{p,n+1}) \\ &> \gamma_+ - \frac{\epsilon}{2} + \frac{1}{p} \liminf_{n \rightarrow \infty} \frac{\ln c_{p,n}}{n}. \end{aligned}$$

By (B.1), we can choose P —and thus p —large enough so that the right-hand side is bounded below by $\gamma_+ - \epsilon$, so the choice $u = v\xi$ satisfies the requirement in the definition of **PA**. \square

We now give a sufficient condition for **UD** which is used in the setup of Sects. 2.3 and 2.4.

Lemma B.7. *Assume that there exists an $e^{o(n)}$ -sequence $(C_n)_{n \in \mathbb{N}}$ such that*

$$\mathbb{P}_{n+m}(x_1^{n+m}) \leq C_n \mathbb{P}_n(x_1^n) \mathbb{P}_m(x_{n+1}^{n+m}) \tag{A.3}$$

*for all $x \in \text{supp } \mathbb{P}$ and all $n, m \in \mathbb{N}$. Then, the measure \mathbb{P} satisfies **UD** with $\tau_n = 0$.*

Proof. Let u, v be as in the definition of **UD**. First, if $\mathbb{P}_{n+m}(uv) > 0$, there exists $x \in [uv] \cap \text{supp } \mathbb{P}$, and in this case (A.3) implies that $\mathbb{P}_{n+m}(uv) \leq C_n \mathbb{P}_n(u) \mathbb{P}_m(v)$, as required. The same inequality trivially holds if $\mathbb{P}_{n+m}(uv) = 0$. \square

The next lemma gives sufficient conditions for **SLD**, **JSLD** and **PA** to hold in terms of the properties introduced in Definition A.2. This is again particularly useful in the setup of Sects. 2.3 and 2.4.

Lemma B.8. *Assume that there exists an $e^{o(n)}$ -sequence $(C_n)_{n \in \mathbb{N}}$ such that*

$$\mathbb{P}_{n+m}(x_1^{n+m}) \geq C_n^{-1} \mathbb{P}_n(x_1^n) \mathbb{P}_m(x_{n+1}^{n+m}) \tag{A.4}$$

for all $x \in \text{supp } \mathbb{P}$ and all $n, m \in \mathbb{N}$. Then, the following hold:

- i. If $\text{supp } \mathbb{P}$ satisfies the flexible specification property, then \mathbb{P} satisfies *SLD*.
- ii. If $\text{supp } \mathbb{P}$ satisfies the abundant periodic orbit property, then \mathbb{P} satisfies *PA*.
- iii. If $\text{supp } \mathbb{P}$ satisfies the flexible specification property, if $\mathbb{Q} \in \mathcal{P}_{\text{inv}}$ is another measure satisfying (A.4), and if $\text{supp } \mathbb{P} = \text{supp } \mathbb{Q}$, then the pair (\mathbb{P}, \mathbb{Q}) satisfies *JSLD*.

Proof. First note the following consequence of repeated applications of (A.4): for all $u, v \in \mathcal{L}$ such that $uv \in \mathcal{L}$, we have

$$\mathbb{P}([uv]) \geq \delta^{|u|} \mathbb{P}([v]), \tag{A.5}$$

with $\delta = C_1^{-1} \inf_{a \in \text{supp } \mathbb{P}_1(a)} \mathbb{P}_1(a) \in (0, 1]$.

- i. Let u, v be as in the definition of *SLD*. We assume that $u, v \in \mathcal{L}$, for otherwise there is nothing to prove. By the flexible specification property, there exists ξ with $|\xi| \leq \tau_n$ such that $u\xi v \in \mathcal{L}$. Then, applying (A.4) to any $x \in [u\xi v]$ implies that $\mathbb{P}_{n+|\xi|+m}(u\xi v) \geq C_n^{-1} \mathbb{P}_n(u) \mathbb{P}_{|\xi|+m}(\xi v) \geq C_n^{-1} \delta^{\tau_n} \mathbb{P}_n(u) \mathbb{P}_m(v)$, where we have used (A.5). We have thus proved *SLD* with $C'_n := C_n \delta^{-\tau_n} = e^{o(n)}$.
- ii. Let $p \in \mathbb{N}$ and $v \in \mathcal{A}^p$. By the abundant periodic orbit property, there exists ξ with $|\xi| \leq \tau_p$ such that $(v\xi)^{n-1} v \in \mathcal{L}$ for all $n \in \mathbb{N}$. By iterating the argument in the proof of Part i, we obtain (B.2) with $c_{n,p} := (C'_p)^{1-n}$, so Lemma B.6 establishes *PA*.
- iii. Under the stated assumptions, \mathbb{P}_n and \mathbb{Q}_n are equivalent, and since the choice of ξ in the proof of Part i only depends on the properties of the support, the same ξ can be chosen to prove *SLD* for both \mathbb{P} and \mathbb{Q} , so *JSLD* holds. \square

Remark B.9. In the same way that we derived (A.5), we conclude that if (A.4) holds, then $\mathbb{P}([u]) \geq \delta^{|u|}$ for all $u \in \mathcal{L}$, so the map $x \mapsto -\frac{1}{n} \ln \mathbb{P}_n(x_1^n)$ is bounded on $\text{supp } \mathbb{P}$. Thus, in this situation, the limit defining $q_{\mathbb{P}}$ is finite whenever it exists. In the same way, if $\mathbb{Q}, \mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ have common support and if \mathbb{Q} satisfies (A.4), then $q_{\mathbb{Q}}$ is finite whenever the limit exists.

Remark B.10. If both (A.3) and (A.4) hold for all $x \in \text{supp } \mathbb{P}$, then there exists a continuous potential φ on $\text{supp } \mathbb{P}$ such that the weak Gibbs condition (2.16) holds for all $x \in \text{supp } \mathbb{P}$, by Theorem 1.2 of [Cun20] applied to $f_n(x) = \ln \mathbb{P}_n(x_1^n)$. Unfortunately, the weak Gibbs condition does not seem to imply (A.3) and (A.4), due to the lack of uniformity in m , so Lemmas B.7 and B.8 do not apply to weak Gibbs measures in general.

B.3. Decoupling properties of g-measures. The following lemma is used in Sect. 2.4.2 in order to show that g -measures satisfy our decoupling assumptions.

Lemma B.11. *Let Ω' be a subshift of Ω . Suppose that $\mathbb{P} \in \mathcal{P}_{\text{inv}}(\Omega)$ is such that $\text{supp } \mathbb{P} = \Omega'$ and that the sequence of functions on Ω' defined by*

$$g_n(x) := \frac{\mathbb{P}_n(x_1^n)}{\mathbb{P}_{n-1}(x_2^n)}$$

converges uniformly to some continuous function $g : \Omega' \rightarrow (0, 1]$. Then, there exists an $e^{o(n)}$ -sequence $(K_n)_{n \in \mathbb{N}}$ such that

$$K_n^{-1} \prod_{i=0}^{n-1} g(T^i x) \leq \mathbb{P}_n(x_1^n) \leq K_n \prod_{i=0}^{n-1} g(T^i x) \tag{A.6}$$

and

$$K_n^{-2} \mathbb{P}_n(x_1^n) \mathbb{P}_m(x_{n+1}^{n+m}) \leq \mathbb{P}_{n+m}(x_1^{n+m}) \leq K_n^2 \mathbb{P}_n(x_1^n) \mathbb{P}_m(x_{n+1}^{n+m}) \tag{A.7}$$

for all $x \in \Omega'$. In particular, (A.3) and (A.4) hold for all $x \in \Omega'$ with $C_n = K_n^2$, and \mathbb{P} is weak Gibbs on Ω' for the potential $\varphi = \ln g$, i.e. (2.16) holds for all $x \in \Omega'$.

Proof. Throughout the proof, we use the convention that $\mathbb{P}_0(x_1^0) = 1$. Let $\Delta_n(x) := \ln g_n(x) - \ln g(x)$ for $n \in \mathbb{N}$ and $x \in \Omega'$. By compactness, positivity and uniform convergence, we find that $\Delta_n^* := \sup_{n' \geq n} \sup_{x \in \Omega'} |\Delta_{n'}(x)|$ defines a bounded sequence that converges to 0 as $n \rightarrow \infty$. As a consequence, $\frac{1}{n} \sum_{k=1}^n \Delta_k^* \rightarrow 0$, and $K_n := \exp \sum_{k=1}^n \Delta_k^*$ defines an $e^{o(n)}$ -sequence. For all $x \in \Omega'$, a telescoping argument yields

$$\mathbb{P}_n(x_1^n) = \prod_{k=1}^n g_{n-k+1}(T^{k-1}x) = e^{\sum_{k=1}^n \Delta_k(T^{n-k}x)} \prod_{k=1}^n g(T^{k-1}x), \tag{A.8}$$

which establishes (A.6). In order to prove (A.7), we remark that by another telescoping argument,

$$\begin{aligned} \mathbb{P}_{n+m}(x_1^{n+m}) &= \mathbb{P}_m(x_{n+1}^{n+m}) \prod_{k=1}^n g_{n+m-k+1}(T^{k-1}x) \\ &= \mathbb{P}_m(x_{n+1}^{n+m}) e^{\sum_{k=1}^n \Delta_{k+m}(T^{n-k}x)} \prod_{k=1}^n g(T^{k-1}x). \end{aligned}$$

Combining the above with (A.8) and noting that $\Delta_{k+m}^* \leq \Delta_k^*$, we obtain (A.7). □

B.4. Relation to mixing. We recall that a measure $\mathbb{P} \in \mathcal{P}_{\text{inv}}$ is said to be ψ -mixing if its ψ -mixing coefficients $(\psi_{\mathbb{P}}(\tau))_{\tau \in \mathbb{N} \cup \{0\}}$, defined by

$$\psi_{\mathbb{P}}(\tau) := \sup_{n, m \in \mathbb{N}} \max \left\{ \left| \frac{\mathbb{P}([u] \cap T^{-n-\tau}[v])}{\mathbb{P}_n(u)\mathbb{P}_m(v)} - 1 \right| : u \in \text{supp } \mathbb{P}_n, v \in \text{supp } \mathbb{P}_m \right\},$$

satisfy $\lim_{\tau \rightarrow \infty} \psi_{\mathbb{P}}(\tau) = 0$.

Lemma B.12. *If \mathbb{P} is ψ -mixing, then \mathbb{P} satisfies SLD and UD with $\tau_n = O(1)$ and $C_n = O(1)$.*

Proof. By assumption, there exists $\tau \in \mathbb{N} \cup \{0\}$ such that $\psi_{\mathbb{P}}(\tau) < 1$. But then UD holds with $\tau_n^+ := \tau$ and $C_n^+ := 1 + \psi_{\mathbb{P}}(\tau)$, and SLD holds with $\tau_n^- := \tau$ and $C_n^- := \frac{|\mathcal{A}|^\tau}{1 - \psi_{\mathbb{P}}(\tau)}$. Both UD and SLD thus hold with $C_n := \max\{C_n^-, C_n^+\}$ and $\tau_n := \tau$. □

Remark B.13. We do not claim that ψ -mixing alone implies PA in general. It does, however, when $\psi_{\mathbb{P}}(0) < 1$. In this case, Lemma B.12 shows that SLD holds with $\tau_n = 0$ for all n , which implies PA; recall Remark 1.5.¹⁴ In the same way, if \mathbb{P} and \mathbb{Q} are two ψ -mixing measures with common support, we do not claim that JS�D holds in general, except once again if $\psi_{\mathbb{P}}(0), \psi_{\mathbb{Q}}(0) < 1$; recall Remark 1.2.

C. From Weak LDPs to Full LDPs

C.1. General results for nonnegative random variables. We gather in the following proposition a collection of results allowing to establish Legendre–Fenchel duality and promote the weak LDP to a full one in the case of nonnegative random variables. They rely on standard techniques, and we provide the details not only for completeness, but also because our setup is slightly unusual: we do not assume the full LDP nor goodness of the rate function, but we rely on the one-dimensional and convex nature of the problem, in the same spirit as the proof of Proposition 5.2 of [CJPS19]. We again use the notation of the beginning of Sect. 4.

Proposition C.1. *Let $(Z_n)_{n \in \mathbb{N}}$ be a sequence of $[0, \infty)$ -valued random variables defined on a probability space (Ω_*, \mathbb{P}_*) . Assume $(Z_n)_{n \in \mathbb{N}}$ satisfies the weak LDP with a rate function I , and set*

$$\underline{q}(\alpha) := \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \int e^{\alpha n Z_n} d\mathbb{P}_* \quad \text{and} \quad \bar{q}(\alpha) := \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \int e^{\alpha n Z_n} d\mathbb{P}_* \quad (\text{C.1})$$

for $\alpha \in \mathbb{R}$. When $\bar{q}(\alpha) = \underline{q}(\alpha)$, let $q(\alpha)$ denote their common value. Then, the following hold:

- i. q, \bar{q} are nondecreasing and satisfy $\underline{q}(0) = \bar{q}(0) = 0$.
- ii. $\bar{I}(s) = \infty$ for all $s < 0$.
- iii. $\underline{q} \geq I^*$.
- iv. $\bar{q}(\alpha) \leq I^*(\alpha)$ for every α satisfying the tail condition

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \int e^{\alpha n Z_n} 1_{Z_n > K} d\mathbb{P}_* = -\infty. \quad (\text{C.2})$$

- v. For every $\alpha < 0$, $q(\alpha)$ exists and $q(\alpha) = I^*(\alpha)$.
- vi. If $\bar{q}(\alpha) < \infty$ for some $\alpha > 0$, then the full LDP holds, I is a good rate function and the family $(Z_n)_{n \in \mathbb{N}}$ is exponentially tight.
- vii. Assume I is convex. If $q(\alpha)$ exists (in $[-\infty, \infty]$) and $q(\alpha) = I^*(\alpha)$ for every $\alpha \in \mathbb{R}$, then $I = q^*$ and the full LDP holds.

Remark C.2. Part v does not extend to $\alpha = 0$ automatically, because the weak LDP alone does not guarantee that $I^*(0) = 0$. Indeed, we could in principle have, for example, $I(s) = 1 + \infty \cdot 1_{s < 0}$ and $q(\alpha) = -1_{\alpha < 0} + \infty \cdot 1_{\alpha > 0}$.

Remark C.3. The convexity assumption in Part vii is required not only to derive $I = q^*$, but also in order to prove the full LDP. In the situation where Part vii applies but Part vi does not, that is when I is convex but $q(\alpha) = \infty$ for all $\alpha > 0$, we shall obtain in the proof that, due to convexity, $\lim_{s \rightarrow \infty} I(s) = 0$, which makes the large-deviation upper bound for unbounded sets somewhat trivial. In particular, in this situation I is not a good rate function and $(Z_n)_{n \in \mathbb{N}}$ is not an exponentially tight family; recall Fig. 4.

¹⁴ The reader familiar with the terminology in e.g. [Bra05, §2.1] will note that $\psi'_{\mathbb{P}}(0) > 0$ suffices.

Proof. In this proof, we denote by μ_n the distribution of Z_n under \mathbb{P}_* , that is $\mu_n = \mathbb{P}_* \circ Z_n^{-1}$, which obviously satisfies $\mu_n([0, \infty)) = 1$. Parts i and ii are immediate consequences of the definitions (C.1) and the fact that $Z_n \geq 0$.

iii. This is the standard lower bound of Varadhan’s lemma; see e.g. Lemma 4.3.4 in [DZ09]. We include the proof for completeness. For any $s, \alpha \in \mathbb{R}$ and $\epsilon > 0$, we have by the weak LDP

$$\begin{aligned} \underline{q}(\alpha) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \int_{B(s, \epsilon)} e^{\alpha n s'} d\mu_n(s') \\ &\geq (\alpha s - |\alpha| \epsilon) + \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mu_n(B(s, \epsilon)) \\ &\geq \alpha s - |\alpha| \epsilon - I(s). \end{aligned}$$

Letting $\epsilon \rightarrow 0$, we obtain $\underline{q}(\alpha) \geq \alpha s - I(s)$. Since α and s are arbitrary, we indeed have $\underline{q} \geq I^*$.

iv. We give here the classical covering argument used in the proof of the upper bound of Varadhan’s lemma; see e.g. Lemma 4.3.6 in [DZ09] for the same result under slightly different assumptions. Let α verify (C.2) and fix $\epsilon > 0$ and $K > 0$. For every $s \in [0, K]$, there exists an open neighborhood $G_s \ni s$ such that

$$\inf_{s' \in \overline{G}_s} I(s') \geq \min \left\{ (I(s) - \epsilon), \epsilon^{-1} \right\} \quad \text{and} \quad \sup_{s' \in G_s} \alpha s' \leq \alpha s + \epsilon.$$

These neighborhoods cover the compact set $[0, K]$ and we can extract a finite subcover $\{G_{s_1}, G_{s_2}, \dots, G_{s_r}\}$. Now, by the weak LDP,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \int_{G_{s_i}} e^{\alpha n s'} d\mu_n(s') &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \ln (e^{\alpha n s_i + \epsilon n} \mu_n(G_{s_i})) \\ &\leq \alpha s_i + \epsilon - \min \left\{ (I(s_i) - \epsilon), \epsilon^{-1} \right\} \\ &= \max \left\{ \alpha s_i - I(s_i) + 2\epsilon, \alpha s_i + \epsilon - \epsilon^{-1} \right\} \\ &\leq \max \left\{ I^*(\alpha) + 2\epsilon, |\alpha| K + \epsilon - \epsilon^{-1} \right\} \end{aligned}$$

for each $i = 1, 2, \dots, r$, and it follows that

$$\begin{aligned} \overline{q}(\alpha) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \left(\int_{(K, \infty)} e^{\alpha n s} d\mu_n(s) + \sum_{i=1}^r \int_{G_{s_i}} e^{\alpha n s} d\mu_n(s) \right) \\ &\leq \max \{ R_K, I^*(\alpha) + 2\epsilon, |\alpha| K + \epsilon - \epsilon^{-1} \}, \end{aligned}$$

where $R_K = \limsup_{n \rightarrow \infty} \int_{(K, \infty)} e^{\alpha n s} d\mu_n(s)$. Sending $\epsilon \rightarrow 0$ shows that

$$\overline{q}(\alpha) \leq \max \{ R_K, I^*(\alpha) \}.$$

Finally, sending $K \rightarrow \infty$ and using (C.2) yields $\overline{q}(\alpha) \leq I^*(\alpha)$, as claimed.

v. When $\alpha < 0$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \int_{(K, \infty)} e^{\alpha ns} d\mu_n(s) \leq \alpha K + \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \int_{(K, \infty)} d\mu_n(s) \leq \alpha K,$$

so Part iv ensures that $\bar{q}(\alpha) \leq I^*(q)$. Since also $\underline{q}(\alpha) \geq I^*(q)$ by Part iii, we conclude that $q(\alpha) = I^*(\alpha)$.

vi. If $\bar{q}(\alpha) < \infty$ for some $\alpha > 0$, then a standard application of Chebyshev’s exponential inequality yields

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mu_n((K, \infty)) \leq \bar{q}(\alpha) - \alpha K,$$

which establishes exponential tightness, and thus the full LDP and the goodness of the rate function (in fact $I(s) \geq \alpha s - \bar{q}(\alpha)$), see e.g. Lemma 1.2.18 in [DZ09].

vii. Since I is convex and lower semicontinuous, we obtain $I = I^{**} = q^*$. In view of Part vi, we need only prove the full LDP in the case where $q(\alpha) = \infty$ for all $\alpha > 0$. In this situation, $I(s) = q^*(s) = \sup_{\alpha \leq 0} (\alpha s - q(\alpha))$ for all $s \in \mathbb{R}$, so clearly the function I is nonincreasing. Thus, since $\inf_{s \geq 0} I(s) = -I^*(0) = -q(0) = 0$, we must have $\lim_{s \rightarrow \infty} I(s) = 0$. In order to extend the weak LDP to a full one, consider a closed set $\Gamma \subseteq \mathbb{R}$. Since $\mu_n((-\infty, 0)) = 0$ and $I(s) = \infty$ when $s < 0$, it suffices to prove the large-deviation upper bound for $\Gamma' := \Gamma \cap [0, \infty)$. If Γ' is compact, the weak LDP provides the desired upper bound. If Γ' is not compact, then it is unbounded, and since $\lim_{s \rightarrow \infty} I(s) = 0$ we conclude that $\inf_{s \in \Gamma} I(s) = 0$, making the large-deviation upper bound trivial. \square

C.2. Proof of Theorem 0. The proof relies on the weak LDP established in [CJPS19], which we shall promote, using Proposition C.1, to a full LDP with suitable Legendre–Fenchel duality relations.

Before we proceed, it should be noted that the arguments below can be simplified in the special case where $\mathbb{Q} = \mathbb{P}$. Indeed, in this case, exponential tightness is immediate (see Remark 2.8), and one can apply Theorem 2.8 of [CJPS19] to the pair $(\mathbb{P}, \widehat{\mathbb{P}})$, with $\widehat{\mathbb{P}}$ the uniform measure. Indeed, in this case, $\frac{1}{n} \sigma_n(x) = \frac{1}{n} \ln \mathbb{P}_n(x_1^n) + \ln |\mathcal{A}|$; see Remark 2.11 in [CJPS19].

Returning to the general case, we note that the combination of UD and JSLD implies, in view of Lemma A.1 of [CJPS19], that the pair (\mathbb{P}, \mathbb{Q}) satisfies assumption SSD of [CJPS19]. By an immediate adaptation of Proposition 3.1.2 of [CJPS19], we conclude that the random variables $x \mapsto -\ln \mathbb{Q}_n(x_1^n)$ are ψ -compatible in the sense of Definition 3.10 of [CJPS19] (the map $\psi_{n,t}$ is constructed in Proposition 3.1.1 therein). Thus, Proposition 3.11 of [CJPS19] implies that the sequence $(Z_n)_{n \in \mathbb{N}}$ defined by $Z_n(x) := -\frac{1}{n} \ln \mathbb{Q}_n(x_1^n)$ satisfies the weak LDP with a convex rate function $I_{\mathbb{Q}}$. More specifically, Proposition 3.11.2 of [CJPS19] ensures that there exists a sequence $(\gamma_n)_{n \in \mathbb{N}}$ with $\gamma_n \rightarrow 0$ such that for all $\epsilon > 0$, all $n \in \mathbb{N}$ and all $s \in \mathbb{R}$,

$$\frac{1}{n} \ln \mu_n(B(s, \epsilon)) \leq \gamma_n - \inf_{y \in B(s, \epsilon + (1+|s|)\gamma_n)} I_{\mathbb{Q}}(y), \tag{C.3}$$

where μ_n is the distribution of Z_n with respect to \mathbb{P} . Moreover, Lemma 3.12 in [CJPS19] implies that the limit defining $q_{\mathbb{Q}}$ exists for all $\alpha \in \mathbb{R}$ and defines a lower semicontinuous, convex function $q_{\mathbb{Q}} : \mathbb{R} \rightarrow (-\infty, \infty]$. We are thus in a position to exploit Proposition C.1, with $\Omega_* := \Omega$, Z_n as above, and $I := I_{\mathbb{Q}}$.

Proof of Parts i–ii. In view of the above discussion and of Proposition C.1.i–ii, it only remains to promote the weak LDP to a full one and to establish the Legendre–Fenchel duality relations (1.14). These two conclusions will follow from Proposition C.1.vii, once we have established the identity $q_{\mathbb{Q}} = I_{\mathbb{Q}}^*$.

Since Parts iii and v of Proposition C.1 ensure that $q_{\mathbb{Q}}(\alpha) \geq I_{\mathbb{Q}}^*(\alpha)$ for all $\alpha \in \mathbb{R}$ and that $q_{\mathbb{Q}}(\alpha) = I_{\mathbb{Q}}^*(\alpha)$ for all $\alpha < 0$, it remains to prove that $q_{\mathbb{Q}}(\alpha) \leq I_{\mathbb{Q}}^*(\alpha)$ for all $\alpha \geq 0$. Values of $\alpha \geq 0$ will be split into two regions, separated by the limit

$$\alpha_+ := \lim_{s \rightarrow \infty} \frac{I_{\mathbb{Q}}(s)}{s},$$

which exists in $[0, \infty]$ by convexity of $I_{\mathbb{Q}}$.

Case 1: $0 \leq \alpha < \alpha_+$. Let $\delta = \frac{1}{2} \min\{1, \alpha_+ - \alpha\}$. Then, there exists $c > 0$ such that

$$I_{\mathbb{Q}}(s) \geq \alpha s + \delta s - c$$

for all $s \geq 0$. Using this and (C.3), we deduce that

$$\frac{1}{n} \ln \mu_n(B(k, 1)) \leq \gamma_n - \inf_{y \in B(k, 1 + (1+k)\gamma_n)} I_{\mathbb{Q}}(y) \leq -\alpha k - \delta k + c' + c''k\gamma_n$$

for all $k \geq 0$, where $\gamma_n \rightarrow 0$ as $n \rightarrow \infty$, while the constants c' and c'' are independent of n and k . It follows that

$$\mu_n(B(k, 1)) \leq \exp\left(n\left(-\alpha k - \frac{\delta k}{2} + c'\right)\right)$$

for all n large enough so that $c''\gamma_n \leq \frac{\delta}{2}$. We thus obtain, for all $K > 0$,

$$\int_{(K, \infty)} e^{\alpha ns} d\mu_n(s) \leq \sum_{j=1}^{\infty} e^{\alpha n(K+j+1)} \mu_n(B(K+j, 1)) \leq \sum_{j=1}^{\infty} e^{-n(K+j)\frac{\delta}{2} + n(c'+\alpha)}.$$

By summing the geometric series, we conclude that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \int_{(K, \infty)} e^{\alpha ns} d\mu_n(s) \leq -(K+1)\frac{\delta}{2} + c' + \alpha.$$

Therefore, the condition (C.2) is satisfied and Proposition C.1.iv yields the bound $q_{\mathbb{Q}}(\alpha) \leq I_{\mathbb{Q}}^*(\alpha)$.

Case 2: $\alpha \geq \alpha_+$. Of course, this is only to be considered if $\alpha_+ < \infty$. Since $I_{\mathbb{Q}}(s) = \infty$ for all $s < 0$, we have $I_{\mathbb{Q}}^*(\alpha) = \sup_{s \geq 0} (\alpha s - I_{\mathbb{Q}}(s))$ for all $\alpha \in \mathbb{R}$, so the function $I_{\mathbb{Q}}^*$ is nondecreasing. Thus, both $q_{\mathbb{Q}}$ and $I_{\mathbb{Q}}^*$ are nondecreasing, lower semicontinuous functions, which implies that they are left continuous. It follows that the inequality $q_{\mathbb{Q}}(\alpha) \leq I_{\mathbb{Q}}^*(\alpha)$, which we have now established for all $\alpha \in (-\infty, \alpha_+)$, extends to $\alpha = \alpha_+$.¹⁵

Finally, if $\alpha > \alpha_+$, then obviously $I_{\mathbb{Q}}^*(\alpha) \geq \lim_{s \rightarrow \infty} (\alpha s - I_{\mathbb{Q}}(s)) = \infty$ by the definition of α_+ , so the inequality $q_{\mathbb{Q}}(\alpha) \leq I_{\mathbb{Q}}^*(\alpha)$ is trivial.

¹⁵ While this extension to α_+ may seem to be a technical detail, lower semicontinuity of $q_{\mathbb{Q}}$, which is ensured by a subadditivity argument in Lemma 3.12 of [CJPS19], plays an important role here, preventing for example the situation described in Remark C.2 (in that example $\alpha_+ = 0$).

Combining the two cases, we have proved that $q_{\mathbb{Q}} \leq I_{\mathbb{Q}}^*$, and thus the proof of Theorem 0.i–ii is complete.

Proof of Part iii. In this part $\mathbb{Q} = \mathbb{P}$. The four claims are proved in an order which differs from that of the statement.

c. By the rightmost expression in (1.7),

$$q_{\mathbb{P}}(1) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln |\text{supp } \mathbb{P}_n(u)| = h_{\text{top}}(\text{supp } \mathbb{P}), \tag{C.4}$$

so (1.16) is a consequence of monotonicity in α .

b. We need to prove that

$$\gamma_+ = \lim_{n \rightarrow \infty} \frac{1}{n} \sup_{u \in \mathcal{A}^n} \ln \mathbb{P}_n(u) \quad \text{and} \quad \gamma_- = \lim_{n \rightarrow \infty} \frac{1}{n} \inf_{u \in \text{supp } \mathbb{P}_n} \ln \mathbb{P}_n(u). \tag{C.5}$$

By **UD**, the sequence $(a_n)_{n \in \mathbb{N}}$ defined by $a_n := \sup_{u \in \mathcal{A}^n} \ln \mathbb{P}_n(u)$ satisfies $a_{n+m+\tau_n} \leq a_n + a_m + \ln C_n$ for all $n, m \in \mathbb{N}$, and a variant of Fekete’s subadditive lemma applies [Raq23, §2]. This proves the statement about γ_+ .

To prove the statement about γ_- , set this time $a_n := \inf_{u \in \text{supp } \mathbb{P}_n} \ln \mathbb{P}_n(u)$ and let $n, m \in \mathbb{N}$ be arbitrary. Pick $u \in \mathcal{A}^n$ such that $\ln \mathbb{P}_n(u) = a_n$ and $v \in \mathcal{A}^m$ such that $\ln \mathbb{P}_m(v) = a_m$. Since $\mathbb{P}_m(v) > 0$, there exists $b \in \mathcal{A}^{\tau_n}$ such that $\mathbb{P}_{\tau_n+m}(bv) > 0$, and by **SLD**, there exist $0 \leq \ell \leq \tau_n$ and $\xi \in \mathcal{A}^\ell$ such that $\mathbb{P}_{n+\ell+\tau_n+m}(u\xi bv) > 0$. But then, in view of **UD**,

$$0 < \mathbb{P}_{n+\ell+\tau_n+m}(u\xi bv) \leq C_n \mathbb{P}_n(u) \mathbb{P}_{m+\ell}(b_{\tau_n-\ell+1}^{\tau_n} v) \leq C_n \mathbb{P}_n(u) \mathbb{P}_m(v),$$

with the convention that $\xi_{\tau_n+1}^{\tau_n}$ is the empty word. Therefore, $u\xi bv \in \text{supp } \mathbb{P}_{n+\ell+\tau_n+m}$ and thus

$$a_{n+\ell+m+\tau_n} \leq \ln \mathbb{P}_{n+\ell+\tau_n+m}(u\xi bv) \leq a_n + a_m + \ln C_n.$$

Since the sequence $(a_n)_{n \in \mathbb{N}}$ is nonincreasing, we conclude that $a_{n+m+2\tau_n} \leq a_n + a_m + \ln C_n$, so the same variant of Fekete’s subadditive lemma yields the claim.

a. In view of (1.16), Proposition C.1.vi applies, so $(Z_n)_{n \in \mathbb{N}}$ is exponentially tight and $I_{\mathbb{P}}$ is a good rate function. We now prove (1.15). First, by (C.4) and since $q_{\mathbb{P}}(1) = I_{\mathbb{P}}^*(1) = \sup_{s \in \mathbb{R}} (s - I_{\mathbb{P}}(s))$, we conclude that $I_{\mathbb{P}}(s) \geq s - h_{\text{top}}(\text{supp } \mathbb{P})$ for all $s \in \mathbb{R}$. By the Ruelle–Lanford representation of $I_{\mathbb{P}}$, and since

$$\mu_n(B(s, \epsilon)) \leq e^{-n(s-\epsilon)} \left| \left\{ u \in \mathcal{A}^n : -\frac{1}{n} \ln \mathbb{P}_n(u) \in B(s, \epsilon) \right\} \right|$$

with a similar lower bound (with $e^{-n(s+\epsilon)}$), we obtain

$$I_{\mathbb{P}}(s) = s - \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \left| \left\{ u \in \mathcal{A}^n : -\frac{1}{n} \ln \mathbb{P}_n(u) \in B(s, \epsilon) \right\} \right|. \tag{C.6}$$

The limit superior can be replaced by a limit inferior. Then, since the logarithm is either nonnegative or equal to $-\infty$, we conclude that $I_{\mathbb{P}}(s)$ is either bounded above by s or infinite. If $s < -\gamma_+$ or $s > -\gamma_-$, then the definition of γ_{\pm} implies that for $\epsilon > 0$ small enough and n large enough there is no $u \in \mathcal{A}^n$ such that $-\frac{1}{n} \ln \mathbb{P}_n(u) \in B(s, \epsilon)$, so $I_{\mathbb{P}}(s) = \infty$. Next, by (C.5), we have for all $\epsilon > 0$ that the set on the right-hand side

of (C.6) with $s = -\gamma_+$ is nonempty for all n large enough, so $I_{\mathbb{P}}(-\gamma_+) \leq -\gamma_+$. When $\gamma_- > -\infty$, a symmetric argument shows that $I_{\mathbb{P}}(-\gamma_-) \leq -\gamma_-$, so by convexity, we indeed obtain $I_{\mathbb{P}}(s) \leq s$ for all $s \in [-\gamma_+, -\gamma_-]$. Now, if $\gamma_- = -\infty$, then by (C.5) we can pick, for each $n \in \mathbb{N}$, some $u^{(n)} \in \mathcal{A}^n$ so that $s_n := -\frac{1}{n} \ln \mathbb{P}_n(u^{(n)}) \rightarrow \infty$ as $n \rightarrow \infty$. By (C.3) with $\epsilon = 1$, we obtain

$$\inf_{s' \in B(s_n, 1 + (1 + s_n)\gamma_n)} I_{\mathbb{P}}(s') \leq -\frac{1}{n} \ln \mu_n(B(s_n, 1)) + \gamma_n \leq s_n + \gamma_n.$$

As a consequence, for each n , there is some $s'_n \in B(s_n, 1 + (1 + s_n)\gamma_n)$ such that $I_{\mathbb{P}}(s'_n) \leq s_n + \gamma_n + 1 < \infty$, and thus $I_{\mathbb{P}}(s'_n) \leq s'_n$ by (C.6). Noting that $s'_n \rightarrow \infty$, and recalling that $I_{\mathbb{P}}(-\gamma_+) \leq -\gamma_+$, convexity allows to conclude that $I_{\mathbb{P}}(s) \leq s$ for all $s \geq -\gamma_+$. The proof of (1.15) is complete.

- d. If $\gamma_- > -\infty$, then (1.15) yields $q_{\mathbb{P}}(\alpha) = \sup_{s \in [-\gamma_+, \gamma_-]} (\alpha s - I_{\mathbb{P}}(s)) < \infty$ for all $\alpha \in \mathbb{R}$. If on the contrary $\gamma_- = -\infty$, then (1.15) implies that $I_{\mathbb{P}}(s) \leq s$ for all $s \in [-\gamma_+, \infty)$, so that, for $\alpha > 1$, we obtain $q_{\mathbb{P}}(\alpha) \geq \lim_{s \rightarrow \infty} (\alpha s - I_{\mathbb{P}}(s)) = \infty$. \square

References

- [AACG23] Abadi, M., Amorim, V.G., Chazottes, J.-R., Gallo, S.: Return-time L^q -spectrum for equilibrium states with potentials of summable variation. *Ergodic Theor. Dyn. Syst.* **43**(8), 2489–2515 (2023)
- [AAG21] Abadi, M., Amorim, V.G., Gallo, S.: Potential well in Poincar  recurrence. *Entropy* **23**(3), 1–26 (2021)
- [AC15] Abadi, M., Card no, L.: R nyi entropies and large deviations for the first match function. *IEEE Trans. Inf. Theor.* **61**(4), 1629–1639 (2015)
- [AG01] Abadi, M., Galves, A.: Inequalities for the occurrence times of rare events in mixing processes. The state of the art. In Saada, E., Gobron, T., Dunlop, F. (eds.), *Inhomogeneous Random Systems*, Cergy-Pontoise, volume 7 of *Markov Process. Relat. Fields*, pp. 97–112. *Polymat* (2001)
- [AV08] Abadi, M., Vaienti, S.: Large deviations for short recurrence. *Discrete Contin. Dyn. Syst. Ser. A* **21**(3), 729–747 (2008)
- [BCJP21] Benoist, T., Cuneo, N., Jak i , V., Pillet, C.-A.: On entropy production of repeated quantum measurements II. Examples. *J. Stat. Phys.* **182**(3), 1–71 (2021)
- [Ber88] Bertrand, A.: Specification, synchronisation, average length. In Cohen, G., Godlewski, P. (eds) *Coding Theory and Applications*, volume 311 of *Lect. Notes Comput. Sci.*, pp. 86–95. Springer, (1988)
- [BEVELN18] Bissacot, R., Endo, E.O., van Enter, A.C., Le Ny, A.: Entropic repulsion and lack of the g -measure property for Dyson models. *Commun. Math. Phys.* **363**, 767–788 (2018)
- [BJPP18] Benoist, T., Jak i , V., Pautrat, Y., Pillet, C.-A.: On entropy production of repeated quantum measurements I. General theory. *Commun. Math. Phys.* **357**(1), 77–123 (2018)
- [BK83] Brin, M., Katok, A.: On local entropy. In Palis J., Jr (ed) *Geometric Dynamics*, Rio de Janeiro, Brasil 1981, volume 1007 of *Lect. Notes Math.*, pp. 30–38. Springer (1983)
- [Bog07] Bogachev, V.I.: *Measure Theory*, volume 1. Springer (2007)
- [Bow74] Bowen, R.: Some systems with unique equilibrium states. *Math. Syst. Theor.* **8**(3), 193–202 (1974)
- [Bra05] Bradley, R.C.: Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2**, 107–144 (2005)
- [CDEJR23a] Cristadoro, G., Degli Esposti, M., Jak i , V., Raqu pas, R.: On a waiting-time result of Kontoyiannis: mixing or decoupling? *Stoch. Proc. Appl.* **166**, 104222 (2023)
- [CDEJR23b] Cristadoro, G., Degli Esposti, M., Jak i , V., Raqu pas, R.: Recurrence times, waiting times and universal entropy production estimators. *Lett. Math. Phys.* **113**(1) (2023)
- [CFM19] Caby, T., Faranda, D., Mantica, G., Vaienti, S., Yiou, P.: Generalized dimensions, large deviations and the distribution of rare events. *Phys. D: Nonlinear Phenom.* **400**, 132143 (2019)
- [CG05] Chazottes, J.-R., Gabrielli, D.: Large deviations for empirical entropies of g -measures. *Nonlinearity* **18**(6), 2545–2563 (2005)

- [CGS99] Collet, P., Galves, A., Schmitt, B.: Repetition times for Gibbsian sources. *Nonlinearity* **12**(4), 1225 (1999)
- [CJPS19] Cuneo, N., Jakšić, V., Pillet, C.-A., Shirikyan, A.: Large deviations and fluctuation theorem for selectively decoupled measures on shift spaces. *Rev. Math. Phys.* **31**(10), 1950036 (2019)
- [CO00] Chazottes, J.-R., Olivier, E.: Relative entropy, dimensions and large deviations for g-measures. *J. Phys. A* **33**(4), 675 (2000)
- [Com09] Comman, H.: Strengthened large deviations for rational maps and full-shifts, with unified proof. *Nonlinearity* **22**(6), 1413–1429 (2009)
- [CRS18] Coutinho, A., Rousseau, J., Saussol, B.: Large deviation for return times. *Nonlinearity* **31**(11), 5162 (2018)
- [CT06] Cover, T.M., Thomas, J.A.: *Elements of Information Theory*, 2nd edn. Wiley (2006)
- [CT13] Climenhaga, V., Thompson, D.J.: Equilibrium states beyond specification and the Bowen property. *J. London Math. Soc.* **87**(2), 401–427 (2013)
- [CT16] Climenhaga, V., Thompson, D.J.: Unique equilibrium states for flows and homeomorphisms with non-uniform structure. *Adv. Math.* **303**, 745–799 (2016)
- [CT21] Climenhaga, V., Thompson, D.J.: Beyond Bowen’s specification property. In Pollicot, M., Vaienti, S. (eds.) *Thermodynamic Formalism*, volume 2290 of *Lect. Notes Math.*, pp. 3–82. Springer (2021)
- [CU05] Chazottes, J.-R., Ugalde, E.: Entropy estimation and fluctuations of hitting and recurrence times for Gibbsian sources. *Discrete Contin. Dyn. Syst. Ser. B* **5**(3), 565–586 (2005)
- [CU11] Chazottes, J.-R., Ugalde, E.: On the preservation of Gibbsianness under symbol amalgamation. In Marcus, B., Petersen, K., Weissman, T. (eds.) *Entropy of Hidden Markov Processes and Connections to Dynamical Systems* *London Math. Soc. Lect. Notes*, pp. 72–97. Cambridge University Press (2011)
- [Cun20] Cuneo, N.: Additive, almost additive and asymptotically additive potential sequences are equivalent. *Commun. Math. Phys.* **377**, 2579–2595 (2020)
- [DS89] Deuschel, J.-D., Stroock, D.W.: *Large Deviations*. Academic Press (1989)
- [DZ09] Dembo, A., Zeitouni, O.: *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*, 2nd edn. Springer (2009)
- [EKW94] Eizenberg, A., Kifer, Y., Weiss, B.: Large deviations for \mathbb{Z}^d -actions. *Commun. Math. Phys.* **164**(3), 433–454 (1994)
- [Ell06] Ellis, R.S.: *Entropy, large deviations, and statistical mechanics*. *Classics in mathematics*. Springer (2006)
- [FGM11] Fernández, R., Gallo, S., Maillard, G.: Regular g-measures are not always Gibbsian. *Electron. Commun. Probab.* **16**, 732–740 (2011)
- [FW01] Feng, D.-J., Wu, J.: The Hausdorff dimension of recurrent sets in symbolic spaces. *Nonlinearity* **14**(1), 81–86 (2001)
- [GBC16] Goodfellow, I., Bengio, Y., Courville, A.: *Deep learning*. MIT Press (2016)
- [GS97] Galves, A., Schmitt, B.: Inequalities for hitting times in mixing dynamical systems. *Rand. Comput. Dyn.* **5**(4), 337–348 (1997)
- [HSV99] Hirata, M., Saussol, B., Vaienti, S.: Statistics of return times: a general framework and new applications. *Commun. Math. Phys.* **206**(1), 33–55 (1999)
- [Hul06] Hulse, P.: An example of non-unique g-measures. *Ergodic Theor. Dyn. Syst.* **26**(2), 439–445 (2006)
- [HV10] Haydn, N., Vaienti, S.: The Rényi entropy function and the large deviation of short return times. *Ergodic Theor. Dyn. Syst.* **30**(1), 159–179 (2010)
- [JB13] Jain, S., Bansal, R.K.: On large deviation property of recurrence times. In: 2013 IEEE International Symposium on Information Theory, pp. 2880–2884 (2013)
- [Joh06] Johnson, O.: A central limit theorem for non-overlapping return times. *J. Appl. Probab.* **43**(1), 32–47 (2006)
- [Jun11] Jung, U.: On the existence of open and bi-continuing codes. *Trans. Amer. Math. Soc.* **363**(3), 1399–1417 (2011)
- [Kel98] Keller, G.: *Equilibrium states in ergodic theory*. *London Mathematical Society Student Texts*, vol. 42. Cambridge University Press, Cambridge (1998)
- [Kif90] Kifer, Y.: Large deviations in dynamical systems and stochastic processes. *Trans. Amer. Math. Soc.* **321**(2), 505–524 (1990)
- [KŁO16] Kwietniak, D., Łącka, M., Oprocha, P.: A panorama of specification-like properties and their consequences. In Kolyad, S., Möller, M., Moree, P., Ward, T. (eds.) *Dynamics and Numbers*, volume 669 of *Contemp. Math.*, pp. 155–186. *Amer. Math. Soc.*, (2016)
- [Kon98] Kontoyannis, I.: Asymptotic recurrence and waiting times for stationary processes. *J. Theor. Probab.* **11**(3), 795–811 (1998)

- [LMV98] Lőrinczi, J., Maes, C., Velde, K.V.: Transformations of Gibbs measures. *Probab. Theory Relat. Fields* **112**(1), 121–147 (1998)
- [LPS94] Lewis, J.T., Pfister, C.-É., Sullivan, W.G.: Large deviations and the thermodynamic formalism: a new proof of the equivalence of ensembles. In Fannes, M., Maes, C., Verbeure, A. (eds.) *On Three Levels. Micro-, Meso-, and Macro-Approaches in Physics*, volume 324 of NATO ASI Series, pp. 183–192. Springer (1994)
- [LPS95] Lewis, J.T., Pfister, C.-É., Sullivan, W.G.: Entropy, concentration of probability and conditional limit theorems. *Markov Proc. Relat. Fields* **1**(3), 319–386 (1995)
- [MS95] Marton, K., Shields, P.C.: Almost-sure waiting time results for weak and very weak Bernoulli processes. *Ergodic Theor. Dyn. Syst.* **15**(5), 951–960 (1995)
- [Mur22] Murphy, K.P.: *Probabilistic Machine Learning: An Introduction*. MIT Press (2022)
- [Ols03] Olsen, L.: First return times: multifractal spectra and divergence points. *Discrete Contin. Dyn. Syst.* **10**(3), 635–656 (2003)
- [OW93] Ornstein, D.S., Weiss, B.: Entropy and data compression schemes. *IEEE Trans. Inf. Theory* **39**(1), 78–83 (1993)
- [Par64] Parry, W.: Intrinsic Markov chains. *Trans. Amer. Math. Soc.* **112**(1), 55–66 (1964)
- [Pfi02] Pfister, C.-É.: Thermodynamical aspects of classical lattice systems. In Sidoravicius, V. (ed.) *In and Out of Equilibrium: Probability with a Physics Flavor*, volume 51 of *Prog. Probab.*, pp. 393–472. Birkhäuser (2002)
- [PPW78] Palmer, M.R., Parry, W., Walters, P.: Large sets of endomorphisms and of g -measures. In Markley, N.G., Martin, J.C., Perrizo, W. (eds.) *The Structure of Attractors in Dynamical Systems*, volume 559 of *Lect. Notes Math.*, pp. 191–210. Springer (1978)
- [PS17] Pfister, C.-É., Sullivan, W.G.: Weak Gibbs measures and large deviations. *Nonlinearity* **31**(1), 49–53 (2017)
- [PS20] Pfister, C.-É., Sullivan, W.G.: Asymptotic decoupling and weak Gibbs measures for finite alphabet shift spaces. *Nonlinearity* **33**(9), 4799–4817 (2020)
- [Raq23] Raquépas, R.: A gapped generalization of Kingman’s subadditive ergodic theorem. *J. Math. Phys.* **64**(6), 06270 (2023)
- [Roc70] Rockafellar, R.T.: *Convex Analysis*. Princeton University Press (1970)
- [Rud87] Rudin, W.: *Real and Complex Analysis*, 3rd edn. McGraw Hill (1987)
- [Rue04] Ruelle, D.: *Thermodynamic Formalism*. Cambridge University Press (2004)
- [Shi93] Shields, P.C.: Waiting times: positive and negative results on the Wyner-Ziv problem. *J. Theor. Probab.* **6**(3), 499–519 (1993)
- [Shi96] Shields, P.C.: *The ergodic theory of discrete sample paths*, volume 13 of *Graduate Studies in Mathematics*. American Mathematical Society (1996)
- [Sim93] Simon, B.: *The Statistical Mechanics of Lattice Gases*, volume 1. Princeton University Press (1993)
- [Szp93] Szpankowski, W.: A generalized suffix tree and its (un) expected asymptotic behaviors. *SIAM J. Comput.* **22**(6), 1176–1198 (1993)
- [Var12] Varandas, P.: Non-uniform specification and large deviations for weak Gibbs measures. *J. Stat. Phys.* **146**(2), 330–358 (2012)
- [vEFS93] van Enter, A.C., Fernández, R., Sokal, A.D.: Regularity properties and pathologies of position-space renormalization-group transformations: Scope and limitations of Gibbsian theory. *J. Stat. Phys.* **72**(5/6), 879–1167 (1993)
- [Ver11] Verbitskiy, E.: Thermodynamics of hidden Markov processes. In Marcus, B., Petersen, K., Weissman, T. (eds.) *Entropy of Hidden Markov Processes and Connections to Dynamical Systems*, *London Math. Soc. Lect. Notes*, pp. 258–272. Cambridge University Press (2011)
- [VO16] Viana, M., Oliveira, K.: *Foundations of Ergodic Theory*, volume 151 of *Cambridge studies in Advanced Mathematics*. Cambridge University Press (2016)
- [Wal75] Walters, P.: Ruelle’s operator theorem and g -measures. *Trans. Amer. Math. Soc.* **214**, 375–387 (1975)
- [Wal82] Walters, P.: *An Introduction to Ergodic Theory*. Springer (1982)
- [Wal01] Walters, P.: Convergence of the Ruelle operator for a function satisfying Bowen’s condition. *Trans. Amer. Math. Soc.* **353**(1), 327–347 (2001)
- [Wal05] Walters, P.: Regularity conditions and Bernoulli properties of equilibrium states and g -measures. *J. London Math. Soc.* **71**(2), 379–396 (2005)
- [WZ89] Wyner, A.D., Ziv, J.: Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inf. Theory* **35**(6), 1250–1258 (1989)
- [You90] Young, L.-S.: Some large deviation results for dynamical systems. *Trans. Amer. Math. Soc.* **318**(2), 525–543 (1990)

[Yur02] Yuri, M.: Weak Gibbs measures and the local product structure. *Ergodic Theor. Dyn. Syst.* **22**(6), 1933–1955 (2002)

Communicated by C. Liverani