Large Deviations and Fluctuation Theorem for Selectively Decoupled Measures on Shift Spaces

Noé Cuneo, Vojkan Jakšić, Claude-Alain Pillet, Armen Shirikyan

▶ **To cite this version:**

Noé Cuneo, Vojkan Jakšić, Claude-Alain Pillet, Armen Shirikyan. Large Deviations and Fluctuation Theorem for Selectively Decoupled Measures on Shift Spaces. Reviews in Mathematical Physics, 2019, 31 (10), pp.1950036. 10.1142/S0129055X19500363 . hal-01672683v2

HAL Id: hal-01672683

https://hal.science/hal-01672683v2

Submitted on 13 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Large Deviations and Fluctuation Theorem for Selectively Decoupled Measures on Shift Spaces

Noé Cuneo[1,3], Vojkan Jakšić[1], Claude-Alain Pillet[2], Armen Shirikyan[1,3,4]

[1]Department of Mathematics and Statistics, McGill University
805 Sherbrooke Street West, Montreal, QC, H3A 2K6, Canada

[2]Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France

[3]Department of Mathematics, University of Cergy–Pontoise, CNRS UMR 8088
2 avenue Adolphe Chauvin, 95302 Cergy–Pontoise, France

[4]Centre de Recherches Mathématiques, CNRS UMI 3457
Université de Montréal, Montréal, QC, H3C 3J7, Canada

Abstract. We establish the Level-1 and Level-3 Large Deviation Principle (LDP) for invariant measures on shift spaces over finite alphabets under very general decoupling conditions for which the thermodynamic formalism does not apply. Such decoupling conditions arise naturally in multifractal analysis, in Gibbs states with hard-core interactions, and in the statistics of repeated quantum measurement processes. We also prove the LDP for the entropy production of pairs of such measures and derive the related Fluctuation Relation. The proofs are based on Ruelle–Lanford functions, and the exposition is essentially self-contained.

Contents

1 Introduction

This work concerns the Large Deviation Principle (LDP) for a class of invariant probability measures on shift spaces over finite alphabets. We prove:

1. The LDP for averages of continuous random variables (**Level-1 LDP**).
2. The LDP for empirical measures (**Level-3 LDP**).
3. The LDP for the entropy production of pairs of probability measures (**Fluctuation Theorem**) together with the corresponding symmetry (**Fluctuation Relation**).

The class of invariant probability measures we shall consider are characterized by certain decoupling properties that are described in Section 2.2. The nature and generality of these decoupling assumptions exclude the application of the thermodynamic formalism as in, for example, [16, 33, 22, 11, 8]. The technical route that proved effective is based on Ruelle–Lanford functions.

In this paper, we mean by Fluctuation Theorem (FT) the LDP for the entropy production observable, and Fluctuation Relation (FR) refers to the Gallavotti–Cohen symmetry (1.2) satisfied by the rate function governing the FT. The FT will be established for general pairs of measures (subject to decoupling assumptions), whereas the FR further requires the two measures to be related by some form of involution (including, but not limited to, time reversal; see Definition 2.5).

We now discuss how our results fit in the existing literature, with special emphasis on Part 3 above, which is the original motivation for this work and its most novel part. No knowledge of the works cited in this introduction is required to understand our results and their proofs, as our exposition is essentially self-contained starting from Section 2.

Part 3 extends and complements the results of [3] in the spirit of the recent work [8], and the reader may benefit from reading introductions of [3, 8] in parallel with this one. Both works [3, 8] concern the FT and FR in the context of dynamical systems (M, φ), where M is a compact metric space and $\varphi : M \to M$ a continuous map. However, the scope, details of the setting, the assumptions, and the technical aspects of the two works are quite different, and we start by describing them separately.

In [3] the metric space M was taken to be $\mathcal{A}^{\mathbb{N}}$, where $\mathcal{A}$ is a finite alphabet and φ is the left shift map. A φ-invariant probability measure $\mathbb{P}$ of interest arises through a repeated quantum measurement process generated by a quantum instrument on a finite-dimensional Hilbert space (we recall the precise setup in Example 2.26). The time-reversed instrument and measurement process yield another probability measure $\widehat{\mathbb{P}}$, and the object of study is the entropic distinguishability of the pair $(\mathbb{P}, \widehat{\mathbb{P}})$ that quantifies the emergence of the arrow of time in the repeated measurement process. Denoting by $\mathbb{P}_t$ and $\widehat{\mathbb{P}}_t$ the marginals of these measures on the first t coordinates of $\mathcal{A}^{\mathbb{N}}$, the entropic distinguishability is quantified by the sequence of entropy production observables

$$\sigma_t = \log \frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\widehat{\mathbb{P}}_t}, \qquad t \in \mathbb{N}.$$

The statement of the FT is the LDP for the sequence of random variables $(t^{-1}\sigma_t)_{t\geq 1}$ with respect to the measure $\mathbb{P}$. The main application of the FT concerns hypothesis testing of the pairs $(\mathbb{P}_t, \widehat{\mathbb{P}}_t)$ as $t \to \infty$. The corresponding error exponents (Stein, Chernoff, Hoeffding) quantify the emergence of the arrow of time. The proofs in [3] follow a strategy that goes back to [25] and are centered around the so-called *entropic pressure* defined by

$$e(\alpha) = \lim_{t\to\infty} \frac{1}{t} \log \left[\int \mathrm{e}^{-\alpha\sigma_t} \mathrm{d}\mathbb{P}_t \right], \qquad \alpha \in \mathbb{R}. \tag{1.1}$$

If the limit exists, and is finite and differentiable for all $\alpha \in \mathbb{R}$, then the FT follows from the Gärtner–Ellis theorem, with a rate function I that satisfies the FR

$$I(-s) = I(s) + s, \qquad s \in \mathbb{R}. \tag{1.2}$$

The difficulty with this strategy is that the measures $\mathbb{P}$ and $\widehat{\mathbb{P}}$ that arise through repeated quantum measurement processes often do not satisfy the usual Gibbsian-type conditions that allow the application of the thermodynamic formalism and ensure the existence and regularity of the entropic pressure defined in (1.1). In this case the Gibbsian-type conditions are naturally replaced by a decoupling condition motivated by [15, Proposition 2.8], which is generalized in Section 2.2 below under the name *selective lower decoupling*. Under those decoupling conditions the measures $\mathbb{P}$ and $\widehat{\mathbb{P}}$ can exhibit a very singular behavior from the thermodynamic formalism point of view. In [3] a restricted form of selective symmetric decoupling (see Section 2.2) has been employed to develop a subadditive thermodynamic formalism that leads to the proof of the existence and finiteness of the limit (1.1) for $\alpha \in [0, 1]$ and the differentiability of e on $(0, 1)$. That sufficed for the proof of the local LDP on the interval $\mathfrak{J} = (e'(0^+), e'(1^-))$ (via the local Gärtner–Ellis theorem), the validity of (1.2) for $s \in \mathfrak{J}$, and the development of hypothesis testing. It was however clear that this route cannot be used for the proof of the global LDP and FT since the assumptions of [3] allowed, for example, for situations where $e(\alpha) = +\infty$ for $\alpha \notin [0, 1]$ and $|e'(0^+)| = |e'(1^-)| < \infty$; see the *rotational instruments* in [2].

The work [8] concerned the FT and FR for general dynamical systems (M, φ) with a compact metric space M and a continuous map φ, under minimal chaoticity assumptions (expansiveness and specification). The proof of the FT bypassed the use of the Gärtner–Ellis theorem, hence lifting the regularity requirement on the entropic pressure e. Instead, a Level-3 LDP for empirical measures is established following a strategy that goes back to [16, 33] and has been used in a similar context in [11]. The key steps of the proof are based on application of the Shannon–McMillan–Breiman/Katok–Brin theorem in the ergodic setting and an entropic approximation argument. After that, a fluctuation relation for the rate function governing the Level-3 LDP is obtained,[1] and a contraction argument yields the Level-1 LDP and the familiar FR and FT. This route has proved to be very robust and allowed for the proof of the FR and FT in circumstances that were previously unreachable: in phase transition regime, for discontinuous entropy production observables, and in

[1] We also obtain this Level-3 fluctuation relation in (2.19), although in the present paper the usual FR (for the entropy production observable) is not derived from it.

the asymptotically additive thermodynamic formalism setting. However, in spite of their generality, the assumptions of [8] did not cover the decoupling conditions of [3]. An obvious question is whether the results of [8] could be extended to the setting of [3] with possibly an alternative technical approach. One of the goals of this work is to achieve that.

We shall work with pairs of invariant probability measures $(\mathbb{P}, \widehat{\mathbb{P}})$ on $M = \mathcal{A}^{\mathbb{N}}$ which are more general than those considered in [3], and not necessarily related to each other by time reversal. We shall establish LDPs and FRs that in particular extend those of [8] to the setting of [3].

The derivation of the LDPs in this work is very different from the one in [8]. Here we use the method of *Ruelle–Lanford (RL) functions* that goes back to [38, 24].[2] The method was then used in [1], and further developed in [27, 28, 26, 34]; see also [30]. The main ideas of the method are also exposed in [9, Section 4.1.2], although the terminology *Ruelle–Lanford* does not appear there. The method of RL functions provides a unified approach to the Level-1, Level-3, and entropy production LDPs, and no application of the contraction principle is needed (in other words, the different levels are independent, although their respective proofs have common threads).[3] We are not aware of any previous use of RL functions in the study of entropy production. After that, the FR is proved under the assumption that $\mathbb{P}$ and $\widehat{\mathbb{P}}$ are related by a suitable involution. It should be added that our application of the RL function method is specific to $M = \mathcal{A}^{\mathbb{N}}$ (with straightforward extensions to $M = \mathcal{A}^{\mathbb{N}^d}$ and $M = \mathcal{A}^{\mathbb{Z}^d}$), and at the moment the method does not extend to the general setting of [8].

The paper is organized as follows. In Section 2 we describe our general setting, state our assumptions and our main results, and discuss several examples. For reasons of space, the detailed discussion of examples related to quantum measurement processes is postponed to [2], which is a continuation of [3] and this work. The general construction leading to the proof of our main results together with a presentation of the method of RL functions is given in Section 3. The proofs of the main results are presented in Sections 4, 5, and 6. In the appendix we describe further applications (in particular to weak Gibbs measures, which do not fit directly into our assumptions) and develop a prototypical example of hidden Markov chain where the present method applies but not those of [8] and [3].

We finish with the following general remark. The RL functions method turned out to be surprisingly effective for our purposes. Although this method is both very powerful and natural, it appears to be a lesser used route to LDPs. Even in the cases where the respective LDPs are well known, this approach gives a new perspective on the results and their proofs. The assumptions under which the method is used here are different from the ones existing in the literature, and we hope that the essentially self-contained presentation given in this paper will facilitate its future applications.

Acknowledgments

We are grateful to C.-E. Pfister for helpful discussions and advice. This research was supported by the *Agence Nationale de la Recherche* through the grant NONSTOPS (ANR-17-CE40-0006-01, ANR-17-CE40-0006-02, ANR-17-CE40-0006-03), the CNRS collaboration grant *Fluctuation theorems in stochastic systems*, and the *Initiative d'excellence Paris-Seine*. NC was supported by Swiss National Science Foundation Grant 165057. VJ acknowledges the support of NSERC. The work of CAP has been carried out in the framework of the Labex Archimède (ANR-11-LABX-0033) and of the A*MIDEX project (ANR-11-IDEX-0001-02), funded by the "Investissements d'Avenir" French Government programme managed by the French National Research Agency (ANR). The

[2] See the introduction of [28] for a historical perspective.

[3] Of course, the Level-1 LDP can also be obtained from the Level-3 LDP by the contraction principle, and so can the Level-2 LDP, which we do not discuss except in Remark 2.17. We only include an explicit derivation of the Level-1 LDP for pedagogical purposes, as it is a simple illustration of the method. See also Remark 2.16.

research of AS was carried out within the MME-DII Center of Excellence (ANR-11-LABX-0023-01).

2 Preliminaries and main results

2.1 Setup and notation

Let $\mathcal{A}$ be a finite set and let[4] $\Omega = \mathcal{A}^{\mathbb{N}}$ be the set of sequences $\omega = (\omega_j)_{j\in\mathbb{N}}$ whose elements belong to $\mathcal{A}$. We denote by $\varphi : \Omega \to \Omega$ the left shift defined by $\varphi(\omega)_j = \omega_{j+1}$ for $j \in \mathbb{N}$. We write also $\Omega_t = \mathcal{A}^{[\![1,t]\!]}$, with $[\![1,t]\!] = [1,t] \cap \mathbb{N}$, and call $w = (w_1, \dots, w_t) \in \Omega_t$ a *word* of length $|w| = t$. Given a sequence $\omega \in \Omega$ and two integers $m \le n$, we set $\omega_{[m,n]} = (\omega_m, \dots, \omega_n)$, and similarly for $w \in \Omega_t$ if $m \le n \le t$. The set of words of finite length is denoted by $\Omega_{\text{fin}} = \bigcup_{t\in\mathbb{N}_0} \Omega_t$, with the convention that $\Omega_0 = \{\kappa\}$, where κ is the "empty word" ($|\kappa| = 0$). Given $u, v \in \Omega_{\text{fin}}$, $uv \in \Omega_{\text{fin}}$ denotes the natural concatenation of u and v, which satisfies $|uv| = |u| + |v|$. For the empty word κ and any $w \in \Omega_{\text{fin}}$, $w\kappa = \kappa w = w$.

The set Ω is endowed with the product topology and the corresponding Borel σ-algebra $\mathcal{F}$. We denote by $C(\Omega)$ the usual Banach space of real-valued continuous functions on Ω. The set $\mathcal{P}(\Omega)$ of Borel probability measures on Ω is endowed with the weak topology.[5] We shall write $\mathbb{Q}$ for generic elements of $\mathcal{P}(\Omega)$ and use the symbol $\mathbb{P}$ for the probability measure that will be fixed throughout. $\mathbb{Q} \in \mathcal{P}(\Omega)$ is φ-invariant, or invariant for short, whenever $\mathbb{Q} \circ \varphi^{-1} = \mathbb{Q}$. We denote by $\mathcal{P}_\varphi(\Omega)$ the set of invariant elements of $\mathcal{P}(\Omega)$. For $\mathbb{Q} \in \mathcal{P}(\Omega)$ and $f \in L^1(\mathbb{Q})$ we write

$$\langle f, \mathbb{Q} \rangle = \int f \mathrm{d}\mathbb{Q}.$$

Given a word $w \in \Omega_t$ with $t \in \mathbb{N}$, we introduce the cylinder set $\mathcal{C}_w = \{\omega \in \Omega : \omega_{[1,t]} = w\}$. We adopt the convention that $\mathcal{C}_\kappa = \Omega$ for the empty word κ, and we denote by $(\mathcal{F}_t)_{t\in\mathbb{N}_0}$ the filtration generated by the cylinder sets.[6]

For any $\mathbb{Q} \in \mathcal{P}(\Omega)$ and any $t \in \mathbb{N}$, $\mathbb{Q}_t$ denotes the restriction of $\mathbb{Q}$ to $\mathcal{F}_t$, which we identify with a function on Ω_t in the natural way:

$$\mathbb{Q}_t(w) = \mathbb{Q}(\mathcal{C}_w) =: \mathbb{Q}(w), \qquad w \in \Omega_t,$$

where the expression $\mathbb{Q}(w)$ is used by a slight abuse of notation. Consistently with the convention that $\mathcal{C}_\kappa = \Omega$, we have $\mathbb{Q}(\kappa) = \mathbb{Q}_0(\kappa) = 1$.

Throughout the paper, we fix an invariant probability measure $\mathbb{P} \in \mathcal{P}_\varphi(\Omega)$, which will be subject to some assumptions below. We write

$$\Omega^+ = \operatorname{supp} \mathbb{P} = \{\omega \in \Omega : \mathbb{P}(\omega_{[1,t]}) > 0 \text{ for all } t \in \mathbb{N}\} \tag{2.1}$$

and notice that Ω^+ is a *subshift, i.e.*, a closed subset of Ω satisfying $\varphi(\Omega^+) = \Omega^+$. For $t \in \mathbb{N}$, let $\Omega_t^+ = \{w \in \Omega_t : \mathbb{P}_t(w) > 0\}$ and set $\Omega_{\text{fin}}^+ = \bigcup_{t\in\mathbb{N}_0} \Omega_t^+ = \{w \in \Omega_{\text{fin}} : \mathbb{P}(w) > 0\}$. The set Ω_{fin}^+ is a *language*[7] in the sense that for each $w \in \Omega_{\text{fin}}^+$ the following holds: (1) each subword of w is also in Ω_{fin}^+ and (2) there exist non-empty words u, v such that $uwv \in \Omega_{\text{fin}}^+$.

Finally, we use throughout the conventions that $\log 0 = -\infty$, and $0 \log 0 = 0$.

[4] We adopt the convention that $\mathbb{N} = \{1, 2, 3, \dots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

[5] This topology is metrizable and $\mathcal{P}(\Omega)$ is a compact metric space.

[6] Notice that $\mathcal{F}_t$ is the finite algebra generated by the elements of $\{\mathcal{C}_w : w \in \Omega_t\}$.

[7] See for example [29]. The notions of subshift and language are not crucial in our study; they will only be used to discuss how (weak) Gibbs measures on subshifts fit into our assumptions (see Example 2.21 and Appendix A.3).

2.2 Assumptions

We now introduce a set of *decoupling* assumptions on $\mathbb{P}$ (which are not in force throughout). Without further saying, we shall always assume that the sequences $(\tau_t)_{t\in\mathbb{N}} \subset \mathbb{N}_0$ and $(c_t)_{t\in\mathbb{N}} \subset [0,\infty)$ which appear in these assumptions satisfy

$$\lim_{t\to\infty} \frac{c_t}{t} = \lim_{t\to\infty} \frac{\tau_t}{t} = 0,$$

which we write as of now $c_t = o(t)$ and $\tau_t = o(t)$.

The assumption that will play the central role in our work is

Selective Lower Decoupling (SLD). *For all $t \in \mathbb{N}$, all $u \in \Omega_t$ and all $v \in \Omega_{\mathrm{fin}}$, $|v| \geq 1$, there exists $\xi \in \Omega_{\mathrm{fin}}$, $|\xi| \leq \tau_t$, such that*

$$\mathbb{P}(u\xi v) \geq \mathrm{e}^{-c_t}\mathbb{P}(u)\mathbb{P}(v). \tag{2.2}$$

(Note that we take $|\xi| \leq \tau_t$ and not $|\xi| = \tau_t$; this is crucial).

In order to refine some of the results (see Theorem 2.13), we will sometimes also assume

Upper Decoupling (UD). *For all $t \in \mathbb{N}$, all $u \in \Omega_t$ and all $v \in \Omega_{\mathrm{fin}}$, $|v| \geq 1$,*

$$\sup_{\xi\in\Omega_{\tau_t}} \mathbb{P}(u\xi v) \leq \mathrm{e}^{c_t}\mathbb{P}(u)\mathbb{P}(v).$$

Some of our results involve a pair of measures $(\mathbb{P}, \widehat{\mathbb{P}})$, where $\mathbb{P}$ is as above, and $\widehat{\mathbb{P}} \in \mathcal{P}_\varphi(\Omega)$ is another invariant probability measure. When we consider a pair $(\mathbb{P}, \widehat{\mathbb{P}})$, we always assume the following absolute continuity condition:

$$\mathbb{P}_t \ll \widehat{\mathbb{P}}_t \quad \text{for all } t \in \mathbb{N}. \tag{2.3}$$

Interesting cases include when $\widehat{\mathbb{P}}$ is the uniform measure[8] on Ω, and when $\widehat{\mathbb{P}}$ is obtained from some transformation of $\mathbb{P}$ (see Definition 2.5). This leads to our final assumption that concerns the pair $(\mathbb{P}, \widehat{\mathbb{P}})$:

Selective Symmetric Decoupling (SSD). *For all $t \in \mathbb{N}$, all $u \in \Omega_t$ and all $v \in \Omega_{\mathrm{fin}}$, $|v| \geq 1$, there exists $\xi \in \Omega_{\mathrm{fin}}$, $|\xi| \leq \tau_t$, such that for both $\mathbb{P}^\sharp = \mathbb{P}$ and $\mathbb{P}^\sharp = \widehat{\mathbb{P}}$ we have*

$$\mathrm{e}^{-c_t}\mathbb{P}^\sharp(u)\mathbb{P}^\sharp(v) \leq \mathbb{P}^\sharp(u\xi v) \leq \mathrm{e}^{c_t}\mathbb{P}^\sharp(u)\mathbb{P}^\sharp(v). \tag{2.4}$$

(Note that this is the same ξ for both $\mathbb{P}$ and $\widehat{\mathbb{P}}$).

Remark 2.1. The same sequences $(c_t)_{t\in\mathbb{N}}$ and $(\tau_t)_{t\in\mathbb{N}}$ are used in the different conditions above. This results in no loss of generality. This is obvious for c_t, while for τ_t the argument is slightly more involved. It is immediate that for **(SLD)** and **(SSD)**, τ_t can always be replaced by some $\tau_t' \geq \tau_t$. The same is true of **(UD)**, since we have for all $\xi \in \Omega_{\tau_t'}$ that $\xi = \xi' b$ for some $\xi' \in \Omega_{\tau_t}$ and $b \in \Omega_{\mathrm{fin}}$, and then

$$\mathbb{P}(u\xi v) = \mathbb{P}(u\xi' bv) \leq \mathrm{e}^{c_t}\mathbb{P}(u)\mathbb{P}(bv) \leq \mathrm{e}^{c_t}\mathbb{P}(u)\mathbb{P}(v).$$

Remark 2.2. **(SLD)** is implied by the seemingly weaker condition that

$$\sum_{\xi\in\Omega_{\mathrm{fin}}\,:\,|\xi|\leq\tau_t} \mathbb{P}(u\xi v) \geq \mathrm{e}^{-c_t'}\mathbb{P}(u)\mathbb{P}(v)$$

[8]That is, $\widehat{\mathbb{P}}_t(w) = |\Omega_t|^{-1}$ for $w \in \Omega_t$.

for some $c'_t = o(t)$. In this case (2.2) is easily shown to hold[9] with $c_t = c'_t + \log(\sum_{i=0}^{\tau_t} |\mathcal{A}|^i) = o(t)$. Similarly, **(UD)** implies the seemingly stronger assumption that

$$\sum_{\xi \in \Omega_{\tau_t}} \mathbb{P}(u\xi v) \leq \mathrm{e}^{c'_t} \mathbb{P}(u)\mathbb{P}(v)$$

if we choose $c'_t = c_t + \log(|\mathcal{A}|^{\tau_t}) = o(t)$.

Remark 2.3. Unless $\tau_t \equiv 0$, **(SSD)** does not imply **(UD)**, since the upper bound in (2.4) has to be satisfied only for the "selected" ξ. **(SSD)** does, however, imply **(SLD)** for both $\mathbb{P}$ and $\widehat{\mathbb{P}}$, with the additional information that we can choose the *same* ξ for both $\mathbb{P}$ and $\widehat{\mathbb{P}}$. On the other hand, in order to have **(SSD)**, it is enough to have **(UD)** for both $\mathbb{P}$ and $\widehat{\mathbb{P}}$ as well as **(SLD)** for both $\mathbb{P}$ and $\widehat{\mathbb{P}}$ with the same ξ (in general c_t and τ_t have to be increased, see Lemma A.1).

Remark 2.4. The measure $\mathbb{P}$ is not assumed to be ergodic. One can show, however, that it is ergodic if, for example, **(SLD)** holds with $\sup_t \tau_t < \infty$ and $\sup_t c_t < \infty$ (see Lemma A.2).

One special case of interest is when $\widehat{\mathbb{P}}$ is related to $\mathbb{P}$ by a transformation defined as follows.

Definition 2.5. *For each $t \in \mathbb{N}$, let $\theta_t : \Omega_t \to \Omega_t$ be an involution. Assume that the sequence $\Theta = (\theta_t)_{t\in\mathbb{N}}$ is such that one of the following holds for some involution $u : \mathcal{A} \to \mathcal{A}$:*

1. $\theta_t(w_1, w_2, \dots, w_t) = (u(w_1), u(w_2), \dots, u(w_t))$ *for each $t \in \mathbb{N}$, $w \in \Omega_t$;*
2. $\theta_t(w_1, w_2, \dots, w_t) = (u(w_t), u(w_{t-1}), \dots, u(w_1))$ *for each $t \in \mathbb{N}$, $w \in \Omega_t$.*

For each $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$, we denote by $\Theta\mathbb{Q}$ the invariant measure on Ω obtained by extending the family[10] *of marginals $((\Theta\mathbb{Q})_t)_{t\in\mathbb{N}}$, where $(\Theta\mathbb{Q})_t = \mathbb{Q}_t \circ \theta_t$.*

We shall see below that when $\widehat{\mathbb{P}} = \Theta\mathbb{P}$ for some Θ as above, the FT rate function satisfies the celebrated Fluctuation Relation [13, 14, 19, 18, 17].

Remark 2.6. By the absolute continuity assumption (2.3), in order for the choice $\widehat{\mathbb{P}} = \Theta\mathbb{P}$ to be allowed, Θ and $\mathbb{P}$ must be so that $\mathbb{P}_t \ll (\Theta\mathbb{P})_t$ for all $t \in \mathbb{N}$. Since θ_t is an involution (and hence a bijection), the support of $\mathbb{P}_t$ and that of $(\Theta\mathbb{P})_t$ (as subsets of Ω_t) have the same cardinality, and hence $\widehat{\mathbb{P}} = \Theta\mathbb{P}$ implies that $\mathbb{P}_t$ and $\widehat{\mathbb{P}}_t$ are equivalent for all t.

We finish with several comments on the relation between the decoupling assumptions described in this section and those to be found in the literature.

- Our decoupling assumptions are related to those in [34, Definition 3.2] (restricted to one-sided shift spaces). In view of Remark 2.1, the upper decoupling assumption is the same. Our **(SLD)** condition is weaker than the lower decoupling condition in [34], as we allow $|\xi| \leq \tau_t$ instead of $|\xi| = \tau_t$. This weaker condition covers some important classes of measures (see the examples below), which are not covered by any result in the literature, as far as we are aware. The Ruelle–Lanford estimates, which are done in the spirit of [34], are noticeably complicated by the fact that we allow $|\xi| \leq \tau_t$ in the **(SLD)** condition.
- The main feature of our **(SLD)** assumption, *i.e.*, allowing $|\xi| \leq \tau_t$, is reminiscent of some variants of the specification property for subshifts of Ω, which allow for similar "flexibility" (see for example [36, 37, 40, 35]). Specification properties are conditions on the structure of the subshift (viewed as a metric space in itself), not on measures defined on it. We shall discuss Gibbs states and (weak) Gibbs measures whose supports satisfy such "flexible" forms of specification property in Examples 2.21, 2.23 and in Appendix A.3.

[9] We use that for any finite set A, $\sum_{x\in A} f(x) \leq |A| \max_{x\in A} f(x)$.

[10] One easily shows that both conditions imply that $\sum_{a\in\mathcal{A}} \mathbb{Q}_{t+1}\circ\theta_{t+1}(wa) = \sum_{a\in\mathcal{A}} \mathbb{Q}_{t+1}\circ\theta_{t+1}(aw) = \mathbb{Q}_t\circ\theta_t(w)$, which, by Kolmogorov's extension theorem, guarantees that such an invariant extension exists.

- A property similar to **(SLD)** (with τ_t and c_t independent of t) was observed to hold for some products of matrices in [15, Proposition 2.8], and some parts of our construction are similar to [15]. See Example 2.25 below.
- To the best of our knowledge, the only assumptions similar to **(SSD)** to be found in the literature are in [3], with c_t and τ_t not allowed to depend on t (see Assumptions (C) and (D) therein, and Example 2.26 below).

2.3 Main results

We endow $\mathbb{R}^d$ with the Euclidian structure and denote by $|\cdot|$ and $(\cdot,\cdot)$ the corresponding norm and inner product. Given a function $f : \Omega \to \mathbb{R}^d$, we write $\|f\| = \sup_{\omega\in\Omega} |f(\omega)|$ and introduce

$$S_t f(\omega) = \sum_{s=0}^{t-1} f(\varphi^s(\omega)), \qquad t \in \mathbb{N}.$$

Let us recall that in the standard LDP terminology, a *rate function* is always assumed to be lower semicontinuous, while a *good rate function* has, in addition, compact level sets. The next result follows from Propositions 4.2 and 4.3 below.

Theorem 2.7 (Level-1 LDP)**.** *Assume **(SLD)** and let $f \in C(\Omega,\mathbb{R}^d)$.*

1. *For all $\alpha \in \mathbb{R}^d$, the limit*

$$q_f(\alpha) := \lim_{t\to\infty} \frac{1}{t} \log \left\langle \mathrm{e}^{(\alpha, S_t f)}, \mathbb{P} \right\rangle$$

exists, is finite, and the mapping $\alpha \mapsto q_f(\alpha)$ is convex and $\|f\|$-Lipschitz.

2. *The sequence of random variables $(\frac{1}{t}S_t f)_{t\in\mathbb{N}}$ satisfies the LDP with a good convex rate function I_f, in the sense that for every open set $O \subset \mathbb{R}^d$ and every closed set $\Gamma \subset \mathbb{R}^d$,*

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{1}{t} S_t f \in O\right) \geq -\inf_{x\in O} I_f(x), \tag{2.5}$$

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{1}{t} S_t f \in \Gamma\right) \leq -\inf_{x\in \Gamma} I_f(x). \tag{2.6}$$

Moreover, I_f is the Fenchel–Legendre transform q_f^* of q_f, i.e.,

$$I_f(x) = q_f^*(x) = \sup_{\alpha\in\mathbb{R}^d} ((\alpha, x) - q_f(\alpha)), \qquad x \in \mathbb{R}^d.$$

We define the *entropy production observable* over the time interval $[\![1,t]\!]$ by

$$\sigma_t = \log \frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\widehat{\mathbb{P}}_t}, \tag{2.7}$$

which is $\mathcal{F}_t$-measurable and well defined $\mathbb{P}_t$-almost surely since $\mathbb{P}_t \ll \widehat{\mathbb{P}}_t$. The next result follows from Propositions 5.1 and 5.2.

Theorem 2.8 (LDP for entropy production)**.** *Assume **(SSD)**.*

1. *For all $\alpha \in \mathbb{R}$, the limit*[11]

$$q(\alpha) := \lim_{t\to\infty} \frac{1}{t} \log \langle \mathrm{e}^{\alpha\sigma_t}, \mathbb{P}\rangle \tag{2.8}$$

[11] Note that the sign of α in (2.8) is different from that in (1.1). See Remark 2.9.

exists and defines a closed proper convex function[12] *on* $\mathbb{R}$ *taking its values in* $(-\infty, \infty]$. *In particular* $q(0) = 0$ *and* $q(-1) \leq 0$, *so that* q *is non-positive (and hence finite) on* $[-1, 0]$.

2. *The sequence of random variables* $(\frac{1}{t}\sigma_t)_{t\in\mathbb{N}}$ *satisfies the LDP with a convex rate function* I *in the sense that for every open set* $O \subset \mathbb{R}$ *and every closed set* $\Gamma \subset \mathbb{R}$,

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{1}{t}\sigma_t \in O\right) \geq -\inf_{s\in O} I(s), \tag{2.9}$$

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{1}{t}\sigma_t \in \Gamma\right) \leq -\inf_{s\in \Gamma} I(s). \tag{2.10}$$

Moreover, I *is the Fenchel–Legendre transform* q^* *of* q, i.e.,

$$I(s) = q^*(s) = \sup_{\alpha\in\mathbb{R}}(\alpha s - q(\alpha)), \qquad s \in \mathbb{R}. \tag{2.11}$$

3. *If* q *is finite in a neighborhood of* 0, *then* I *is a good rate function.*

4. *If* $\widehat{\mathbb{P}} = \Theta\mathbb{P}$ *for some* Θ *as in Definition 2.5,*[13] *then* q *satisfies the symmetry*

$$q(-\alpha) = q(\alpha - 1), \qquad \alpha \in \mathbb{R}, \tag{2.12}$$

and I *satisfies the Fluctuation Relation (also known as the* Gallavotti–Cohen *symmetry)*

$$I(-s) = I(s) + s, \qquad s \in \mathbb{R}. \tag{2.13}$$

Remark 2.9. In the physics literature, it is more common to work with $e(\alpha) = q(-\alpha)$ as in (1.1). Then (2.11) and (2.12) read respectively

$$I(s) = -\inf_{\alpha\in\mathbb{R}}(\alpha s + e(\alpha)), \qquad s \in \mathbb{R},$$

$$e(\alpha) = e(1 - \alpha), \qquad \alpha \in \mathbb{R}.$$

This is relevant for the applications to hypothesis testing that are discussed in Section 2.4.

The following remark gives a sufficient condition for $q(\alpha)$ to be finite for all $\alpha \in \mathbb{R}$.

Remark 2.10. Assume that, in addition to **(SSD)**, we have for all $t \in \mathbb{N}$, all $u \in \Omega_t$ and all $v \in \Omega_{\text{fin}}$, $|v| \geq 1$, that both $\mathbb{P}^\sharp = \mathbb{P}$ and $\mathbb{P}^\sharp = \widehat{\mathbb{P}}$ satisfy

$$\mathbb{P}^\sharp(uv) \geq \mathrm{e}^{-c_t}\mathbb{P}^\sharp(u)\mathbb{P}^\sharp(v). \tag{2.14}$$

Let moreover c be the minimum of all the non-zero values achieved by $\mathbb{P}(a)$ and $\widehat{\mathbb{P}}(a)$ for $a \in \mathcal{A}$. Then, for all $w \in \Omega_t^+$, we find $\mathbb{P}^\sharp(w) \geq \mathrm{e}^{-(t-1)c_1}\mathbb{P}^\sharp(w_1)\cdots\mathbb{P}^\sharp(w_t) \geq \mathrm{e}^{-t(c_1 - \log c)}$. As a consequence, we obtain that $\sup_{t\in\mathbb{N}}\sup_{w\in\Omega_t^+}|t^{-1}\sigma_t| < \infty$, which implies in particular that $q(\alpha)$ is finite for all $\alpha \in \mathbb{R}$.

Remark 2.11. If $\widehat{\mathbb{P}}$ is the uniform measure, then $\frac{1}{t}\sigma_t = \frac{1}{t}\log\mathbb{P}_t + \log|\mathcal{A}|$. Assuming **(SSD)** (note that (2.4) trivially holds for $\widehat{\mathbb{P}}$), Parts 1-2 of the theorem above apply. We then have

$$r(\alpha) := \lim_{t\to\infty}\frac{1}{t}\log\left\langle \mathrm{e}^{\alpha\log\mathbb{P}_t}, \mathbb{P}\right\rangle = q(\alpha) - \alpha\log|\mathcal{A}|, \qquad \alpha \in \mathbb{R},$$

[12] *i.e.*, it is convex, lower semicontinuous and not everywhere infinite.

[13] The **(SSD)** assumption of the theorem is still in force.

and r inherits the properties of existence, convexity and lower semicontinuity of q. Moreover, $\frac{1}{t}\log \mathbb{P}_t$ satisfies the LDP with convex rate function $J(s) = I(s + \log|\mathcal{A}|)$, which can be identified with the Fenchel–Legendre transform of r. Part 3 extends in an obvious way to J and r. In order to apply the discussion of Remark 2.10, it suffices to verify (2.14) for $\mathbb{P}$, since (2.14) is trivially satisfied for the uniform measure $\widehat{\mathbb{P}}$. Note also that r is related to the Rényi entropy of $\mathbb{P}$, since

$$r(\alpha) = \lim_{t\to\infty} \frac{1}{t} \log \sum_{w\in\Omega_t^+} (\mathbb{P}(w))^{1+\alpha}.$$

Remark 2.12. By the Shannon–McMillan–Breiman (SMB) theorem,[14] the limit

$$H(\omega) := -\lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}_t(\omega_{[1,t]})$$

exists $\mathbb{P}$-almost surely and in $L^1(\mathbb{P})$, $H \circ \varphi = H$, and $\int_\Omega H \mathrm{d}\mathbb{P} = h(\mathbb{P})$, where $h(\mathbb{P})$ is the Kolmogorov–Sinai entropy of $\mathbb{P}$. Thus, in the case when $\widehat{\mathbb{P}}$ is the uniform measure, Theorem 2.8 provides the LDP counterpart to the SMB theorem, and establishes the result that was originally intended for the fourth installment in the series of papers initiated by [3].

We now turn to the Level-3 LDP. The sequence of *empirical measures* $(\mu_t)_{t\in\mathbb{N}}$ is defined by

$$\mu_t(\omega) = \frac{1}{t}\sum_{s=0}^{t-1} \delta_{\varphi^s(\omega)} \in \mathcal{P}(\Omega), \qquad \omega\in\Omega,\ t\in\mathbb{N}.$$

We also recall that the relative entropy of two probability measures $\mathbb{Q}$ and $\mathbb{Q}'$ on a measurable space $(X, \mathcal{F})$ is given by

$$\operatorname{Ent}(\mathbb{Q}' \,|\, \mathbb{Q}) = \begin{cases} \displaystyle\int_X \log\Big(\frac{\mathrm{d}\mathbb{Q}'}{\mathrm{d}\mathbb{Q}}\Big)\mathrm{d}\mathbb{Q}' & \text{if } \mathbb{Q}' \ll \mathbb{Q}, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.15}$$

In what follows, we always assume that $\mathcal{P}(\Omega)$ is endowed with the weak topology and the corresponding Borel σ-algebra. The next result follows from Propositions 4.2, 6.1, 6.3, and 6.4.

Theorem 2.13 (Level-3 LDP)**.** *Assume* ***(SLD)****.*

1. *For all $f \in C(\Omega, \mathbb{R})$, the limit*

$$Q(f) := \lim_{t\to\infty} \frac{1}{t} \log \left\langle \mathrm{e}^{S_t f}, \mathbb{P}\right\rangle \tag{2.16}$$

   exists and defines a convex, 1-Lipschitz function on $C(\Omega, \mathbb{R})$.

2. *The sequence of random variables $(\mu_t)_{t\in\mathbb{N}}$ satisfies the LDP on the space $\mathcal{P}(\Omega)$ with some good[15] convex rate function $\mathbb{I}$, i.e., for every open set $O \subset \mathcal{P}(\Omega)$, and every closed set $\Gamma \subset \mathcal{P}(\Omega)$,*

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}(\mu_t \in O) \geq -\inf_{\mathbb{Q}\in O} \mathbb{I}(\mathbb{Q}),$$
$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}(\mu_t \in \Gamma) \leq -\inf_{\mathbb{Q}\in \Gamma} \mathbb{I}(\mathbb{Q}).$$

[14] For this result no assumptions are needed; the SMB theorem holds for any $\mathbb{P} \in \mathcal{P}_\varphi(\Omega)$.
[15] Notice that goodness follows from lower semicontinuity since $\mathcal{P}(\Omega)$ is compact.

Moreover, $\mathbb{I}$ is the restriction of the Fenchel–Legendre transform Q^* of Q to $\mathcal{P}(\Omega)$, i.e.,

$$\mathbb{I}(\mathbb{Q}) = \sup_{f \in C(\Omega,\mathbb{R})} \big(\langle f, \mathbb{Q}\rangle - Q(f)\big), \qquad \mathbb{Q} \in \mathcal{P}(\Omega), \tag{2.17}$$

and satisfies $\mathbb{I}(\mathbb{Q}) = +\infty$ for $\mathbb{Q} \in \mathcal{P}(\Omega) \setminus \mathcal{P}_\varphi(\Omega)$.

3. *Assuming **(UD)** (in addition to **(SLD)**), we have for any $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$ that*

$$\mathbb{I}(\mathbb{Q}) = \lim_{t\to\infty} \frac{1}{t} \mathrm{Ent}(\mathbb{Q}_t | \mathbb{P}_t), \tag{2.18}$$

and $\mathbb{I}$ is an affine function of $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$.

4. *Assume again **(UD)** and **(SLD)**. Assume moreover that $\widehat{\mathbb{P}} = \Theta\mathbb{P}$ for some Θ as in Definition 2.5. Then, for any $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$ such that $(\Theta\mathbb{Q})_t$ and $\mathbb{Q}_t$ are equivalent for all t, $\mathbb{I}(\mathbb{Q}) < +\infty$, and $\mathbb{I}(\Theta\mathbb{Q}) < +\infty$, the following* Level-3 fluctuation relation *holds:*

$$\mathbb{I}(\Theta\mathbb{Q}) - \mathbb{I}(\mathbb{Q}) = \lim_{t\to\infty} \frac{1}{t} \langle \sigma_t, \mathbb{Q}\rangle. \tag{2.19}$$

Remark 2.14. Note that in general, the identification (2.18) is not possible for $\mathbb{Q} \in \mathcal{P}(\Omega) \setminus \mathcal{P}_\varphi(\Omega)$; for such measures the left-hand side is infinite, but the right-hand side need not be.

Remark 2.15. The right-hand side of (2.19) is interpreted as the mean entropy production of the pair $(\mathbb{P}, \widehat{\mathbb{P}})$ w.r.t. $\mathbb{Q}$. For a discussion of the Level-3 fluctuation relation (2.19), we refer the reader to [8, 7].

Remark 2.16. As mentioned, the Level-1 LDP of Theorem 2.7 can of course be retrieved from the Level-3 LDP by using the contraction principle. In our proofs, however, the Level-1 LDP is established first, and then the Level-3 LDP is proved independently (although the two proofs have many common points). A natural question is whether, as in [8], the LDP for the entropy production can be retrieved by "approximate" contraction from the Level 3, and if then (2.13) follows from (2.19). We are not aware of a way of doing so at the level of generality of **(SSD)**, as σ_t may be highly "non-additive" (see Example 2.24). A contraction argument along the lines of [8] is, however, possible if **(SSD)** holds with $\tau_t \equiv 0$, since in this case, in the terminology of [8], σ_t is *asymptotically additive*.

Remark 2.17. Assuming **(SLD)**, the following *Level-2 LDP* holds: the sequence

$$\left(\frac{1}{t}\sum_{s=0}^{t-1} \delta_{\omega_s}\right)_{t\in\mathbb{N}} \subset \mathcal{P}(\mathcal{A})$$

satisfies the LDP with respect to a good convex rate function I_2, which can be expressed as

$$I_2(\nu) = \sup_{f\in C(\mathcal{A})} \left(\sum_{a\in\mathcal{A}} f(a)\nu(a) - Q(f)\right), \qquad \nu \in \mathcal{P}(\mathcal{A}),$$

where $Q(f)$ is as in (2.16) (viewing $f \in C(\mathcal{A})$ as a function on Ω depending only on ω_1). This result can be obtained in three different ways: (1) by applying the Level-1 LDP to some well-chosen $\mathbb{R}^{\mathcal{A}}$-valued function, (2) from the Level-3 LDP by the contraction principle, or (3) independently of the others LDPs by applying the Ruelle-Lanford functions method directly. We shall not discuss the Level-2 LDP any further.

Remark 2.18. Although the work [3] was focused on the LDP for entropy production for pairs of probability measures $\mathbb{P}$ and $\Theta\mathbb{P}$ obtained from repeated quantum measurement processes, with Θ as in Case 2 of Definition 2.5, the method of proof extends to the general setting of this paper and yields the following: (a) Theorems 2.7 and 2.13 hold[16] assuming **(UD)** and **(SLD)** with τ_t and c_t that do not depend on t; (b) a local version of Theorem 2.8 holds assuming **(SSD)** and **(UD)** (for both $\mathbb{P}$ and $\widehat{\mathbb{P}}$) with c_t and τ_t that do not depend on t; (c) Theorem 2.8 holds assuming **(SSD)** with $\tau_t \equiv 0$ and with c_t that does not depend on t. In this context, see Example 2.24.

2.4 Hypothesis testing

An important application of Theorem 2.8 concerns asymptotic hypothesis testing of the pairs of measures $(\mathbb{P}_t, \widehat{\mathbb{P}}_t)$ as $t \to \infty$. The discussion of this point is nearly identical to the one presented in Section 2.9 of [3] (see also [8, 7, 6]), and we shall only briefly comment on a few changes that are needed due to the generality of our setting.

Unless $\widehat{\mathbb{P}} = \Theta\mathbb{P}$, the function $\alpha \mapsto e(\alpha) = q(-\alpha)$ that controls the Chernoff and Hoeffding error exponents does not need to satisfy the symmetry $e(\alpha) = e(1-\alpha)$. In this case the upper and lower Chernoff exponents $\bar{c}$ and $\underline{c}$ satisfy[17]

$$\bar{c} = \underline{c} = \min_{\alpha\in[0,1]} e(\alpha).$$

The formula for the Hoeffding error exponents (Theorem 2.13 in [3]) remains unchanged. We also remark that although the analysis of the Chernoff and Hoeffding error exponents presented in [20] (and used in [3]) required the function e to be differentiable on the interval $(0,1)$, this assumption was used only through the application of the induced local LDP for the entropy production.[18] In case e exists but is not necessarily differentiable, and if the required LDP with a convex rate function is established by other means, as is the case in Theorem 2.8, then the analysis of [20] carries through without changes; see [6] for details.

The interpretation of all three types of exponents (Stein, Chernoff, Hoeffding) in terms of hypothesis testing and support separation of the pairs $(\mathbb{P}_t, \widehat{\mathbb{P}}_t)$ as $t \to \infty$ as presented in [3] remains unchanged. Obviously, the support separation is linked to the emergence of the arrow of time only in Case 2 of Definition 2.5.

2.5 Examples

We start with five examples where our results apply, but for which the conclusions are well known and have been reached in the literature by other means. We believe however that in all these cases the Ruelle–Lanford functions method presented here offers a different perspective on the resulting LDPs.

Example 2.19. Bernoulli measures. Let P be a probability measure on $\mathcal{A}$, and define[19] the measure $\mathbb{P}$ by $\mathbb{P}_t(w) = \prod_{i=1}^t P(w_i)$, $w \in \Omega_t$. Then obviously **(UD)** and **(SLD)** are satisfied with $\tau_t \equiv 0$ and $c_t \equiv 0$. If $\widehat{\mathbb{P}}$ is defined similarly for some probability measure $\widehat{P}$ on $\mathcal{A}$, then **(SSD)** also holds with the same sequences $(\tau_t)_{t\in\mathbb{N}}$ and $(c_t)_{t\in\mathbb{N}}$, and all our results apply provided that $P \ll \widehat{P}$.

[16] By adapting the proof of [3, Theorem 2.5] to the case where σ_n is replaced by $S_n f$ for any function f depending on finitely many variables, one verifies the assumptions of [22, Theorem 2.1], which yield the Level-3 LDP.

[17] If the symmetry holds, as in [3], then obviously $\bar{c} = \underline{c} = e(1/2)$.

[18] The local LDP followed by an application of the Gärtner–Ellis theorem.

[19] See [24, Section A.4] for a pedagogical exposition of this case.

Example 2.20. Irreducible Markov processes. Let $\mathbb{P} \in \mathcal{P}_\varphi(\Omega)$ be a Markov process. Then **(UD)** holds. Assume furthermore that it is irreducible (*i.e.*, that for all $a, b \in \mathcal{A}$, there exists $\xi^{(a,b)} \in \Omega_{\text{fin}}$ such that $\mathbb{P}(a\xi^{(a,b)}b) > 0$). Then **(SLD)** holds. If, in addition, $\widehat{\mathbb{P}} \in \mathcal{P}_\varphi(\Omega)$ is another Markov process such that $\mathbb{P}_2 \ll \widehat{\mathbb{P}}_2$ (hence $\widehat{\mathbb{P}}$ is also irreducible), then **(SSD)** holds. See Lemma A.3 for the proof of these claims. Note that no aperiodicity condition is required; if the Markov process $\mathbb{P}$ is irreducible and *aperiodic*, then (2.2) also holds with the condition $|\xi| \leq \tau_t$ strengthened to $|\xi| = \tau_t$.

The next example consists of (weak) Gibbs measures, which have been studied extensively, and for which the LDPs and FR have been obtained via the thermodynamic formalism (see for example [44, 42, 43, 35, 7, 8]).

Example 2.21. Gibbs and weak Gibbs measures on subshifts. Assume that the subshift Ω^+ satisfies the following *weak specification property*:[20] for all $u, v \in \Omega^+_{\text{fin}}$, there exists $\xi \in \Omega^+_{\text{fin}}$, $|\xi| \leq \tau_{|u|}$ such that $u\xi v \in \Omega^+_{\text{fin}}$. Assume moreover that $\mathbb{P}$ is a Gibbs measure for some potential $f \in C(\Omega^+)$, *i.e.*, that for some $p \in \mathbb{R}$, $d \geq 0$ and all $\omega \in \Omega^+$,

$$\mathrm{e}^{-d+S_t f(\omega)-tp} \leq \mathbb{P}_t(\omega_1, \dots, \omega_t) \leq \mathrm{e}^{d+S_t f(\omega)-tp}. \tag{2.20}$$

Then it is easy to realize that **(UD)** and **(SLD)** are satisfied with τ_t as in the above specification property and $c_t = 3d + \tau_t\|f - p\|$. Moreover, **(SSD)** is satisfied if one of the following conditions holds: (a) $\widehat{\mathbb{P}} = \Theta\mathbb{P}$ with Θ as in Definition 2.5 and $\theta_t(\Omega_t^+) = \Omega_t^+$; (b) $\widehat{\mathbb{P}}$ is also a Gibbs measure (*i.e.*, satisfies (2.20) for some $\hat{f}$ and $\hat{p}$, and all ω in the support Ω^+ of $\mathbb{P}$). More generally, we say that $\mathbb{P}$ is a *weak Gibbs measure* if (2.20) holds with d replaced by $d_t = o(t)$. In this case, the decoupling assumptions above do not hold in general.[21] However, we show in Appendix A.3 that our results can easily be adapted to this case.

An interesting special case of Example 2.21 is:

Example 2.22. β-shifts. Consider the β-shift for some $\beta > 1$ (see [37] and references therein). The weak specification property described in Example 2.21 is satisfied for Lebesgue-almost all $\beta > 1$ (see the discussion after Corollary 5.1 in [37]; the quantity defined in equation (5.9) therein plays the role of τ_t), and hence for such β's our results apply to any (weak) Gibbs measure.

We next turn to Gibbs states. Such measures satisfy at the same time our decoupling assumptions and the weak Gibbs condition.

Example 2.23. Gibbs states in 1D. Let $\mathbb{P}^*$ be an invariant Gibbs state (in the Dobrushin–Lanford–Ruelle sense, see for example [39, 41, 11, 28]) for some absolutely summable interaction on the full two-sided shift $\mathcal{A}^{\mathbb{Z}}$. Then the marginal $\mathbb{P}$ of $\mathbb{P}^*$ on the one-sided shift $\mathcal{A}^{\mathbb{N}}$ satisfies **(UD)** and **(SLD)** with $\tau_t \equiv 0$ [28, Lemma 2.9]. If one considers also hard-core interactions, *i.e.*, if $\mathbb{P}^*$ is an invariant Gibbs state on a subshift M of $\mathcal{A}^{\mathbb{Z}}$, then the proof can be adapted provided M satisfies the following condition:[22] for all $\eta, \omega \in M$, and all $t \in \mathbb{N}$, there exists $\eta' \in M$ such that $\omega_{[1,t]}$ appears in $\eta'_{[1,t+\tau_t]}$, and such that $\eta_i = \eta'_i$ for all $i \in \mathbb{Z} \setminus [\![1, t + \tau_t]\!]$. The discussion of **(SSD)** in this setup is similar to the (weak) Gibbs case discussed in Appendix A.3.

[20] A typical example would be a subshift of $\{0, 1, 2\}^{\mathbb{N}}$ where the only restriction is that for each $t \in \mathbb{N}$, every occurrence of the word 01^t0 must be followed by the word $2^{\lfloor\sqrt{t+2}\rfloor}$. Then the weak specification property is satisfied with $\tau_t = \lfloor\sqrt{t}\rfloor$.

[21] As discussed in Appendix A.3, our decoupling assumptions are not comparable with the weak Gibbs property.

[22] This condition is slightly more "flexible" than Condition D in [39, Section 4.1] in the sense that the position where $\omega_{[1,t]}$ appears in $\eta'_{[1,t+\tau_t]}$ may depend on ω (for fixed t).

The LDPs and FR for Gibbs states have also been obtained using the thermodynamic formalism [11, 16, 33, 8, 7], and the proofs therein do not require $\mathbb{P}$ to be φ-invariant. The condition on M spelled out in Example 2.23 seems to be slightly more general than those found in the literature on Gibbs states.

We now turn to examples which genuinely require the full generality of our assumptions.

Example 2.24. A class of hidden Markov chains. In Appendix A.2 we describe a prototypical pair of hidden Markov chains, which satisfies **(SSD)** with $\tau_t \equiv 1$ and $\sup_t c_t < \infty$, and for which the function q defined in (2.8) displays different types of singularities. Depending on the parameters of the model, one can have that:

- q is finite but not differentiable everywhere on $\mathbb{R}$;
- there exists $\alpha_* \geq 0$ such that q is finite (and even analytic) on $(-\infty, \alpha_*)$ and infinite on (α_*, ∞), with either
  - $\lim_{\alpha \uparrow \alpha_*} q(\alpha) = q(\alpha_*) = +\infty$;
  - $q(\alpha_*) < \infty$ and $q'(\alpha_*^-) = +\infty$;[23] or
  - $q(\alpha_*) < \infty$ and $q'(\alpha_*^-) < \infty$.

This leads to situations where either or both [8] and [3] fail to apply, or to give the global LDP in Theorem 2.8. This example illustrates how "non-additive" (or "non-extensive" in physical terms) σ_t can be under our assumptions, in the sense that the sequence $(t^{-1}\sigma_t(\omega))_{t\in\mathbb{N}}$ may be unbounded for some $\omega \in \Omega$. A closely related, and physically relevant, example of *rotational quantum instrument* will be discussed in [2].

The following example arises naturally in multifractal analysis, see [31, 32].

Example 2.25. Matrix product probability measures. Let $M : \mathcal{A} \to \mathbb{M}_N(\mathbb{R})$ be a map taking values in the algebra of real $N \times N$ matrices that satisfies the following assumptions:

(A1) The entries of $M(a)$ are non-negative for all $a \in \mathcal{A}$.

(A2) The matrix $S = \sum_{a\in\mathcal{A}} M(a)$ and its transpose S^T satisfy

$$Sv = \lambda v, \qquad S^T w = \lambda w$$

for some $\lambda > 0$ and vectors $v, w \in \mathbb{R}^N$ with strictly positive entries.

For each $t \in \mathbb{N}$ we define a probability measure $\mathbb{P}_t$ on Ω_t by

$$\mathbb{P}_t(\omega_1, \ldots, \omega_t) = \frac{1}{\lambda^t}\, (w, M(\omega_1) \cdots M(\omega_t) v)\,.$$

One easily verifies that there exists a unique $\mathbb{P} \in \mathcal{P}_\varphi(\Omega)$ whose family of marginals is given by $(\mathbb{P}_t)_{t\in\mathbb{N}}$. We shall call such $\mathbb{P}$ the matrix product measure associated with the triple (M, v, w). We have:

- **(UD)** holds for $\mathbb{P}$ with $\tau_t \equiv 0$, and $\sup_t c_t < \infty$.
- If the entries of $M(a)$ are strictly positive for all $a \in \mathcal{A}$, then **(SLD)** holds with $\tau_t \equiv 0$ and $\sup_t c_t < \infty$.
- If for some $a \in \mathcal{A}$ all the entries of $M(a)$ are strictly positive, then **(SLD)** holds with $\tau_t \equiv 1$ and $\sup_t c_t < \infty$ (by taking $\xi = a$ in (2.2)).

[23] Here and below $q'(\alpha_*^-)$ denotes the left-derivative of q at α_*.

- Suppose that for all $a \in \mathcal{A}$ some entries of $M(a)$ are vanishing. If the matrix S is irreducible, *i.e.*, for some $r \in \mathbb{N}$ the matrix $(I+S)^r$ has strictly positive entries (here I denotes the identity matrix), then **(SLD)** holds with $\tau_t \equiv r$ and $\sup_t c_t < \infty$. Note that if S is irreducible, then **(A2)** automatically holds and $\mathbb{P}$ is ergodic. It is easy to construct examples of M for which **(A1)** and **(A2)** hold, S is not irreducible, and **(SLD)** fails.
- Let $\widehat{M} : \mathcal{A} \to \mathbb{M}_N(\mathbb{R})$ be another map satisfying **(A1)** and **(A2)**, and let $\widehat{\mathbb{P}}$ be the induced probability measure. If the matrix $\sum_{a\in\mathcal{A}} M(a) \otimes \widehat{M}(a)$ acting on $\mathbb{R}^N \otimes \mathbb{R}^N$ is irreducible, then **(SSD)** holds for the pair $(\mathbb{P}, \widehat{\mathbb{P}})$ with τ_t and c_t that do not depend on t (by an adaptation of the proof of Proposition 2.6 in [3]; see also [2]).
- If Θ is as in Definition 2.5 and $\widehat{\mathbb{P}} = \Theta\mathbb{P}$, then in Case 1 (of Definition 2.5), the measure $\widehat{\mathbb{P}}$ is the matrix product measure associated with $(\widehat{M}, v, w)$, where $\widehat{M}(a) = M(u(a))$, and in Case 2 the measure $\widehat{\mathbb{P}}$ corresponds to $(\widehat{M}, w, v)$ (note the order of w and v), where $\widehat{M}(a) = M^T(u(a))$.
- The quantity $q(\alpha)$ in (2.8) is finite for all α. Indeed, since all the non-zero entries of the matrices at hand are bounded below by some constant $c > 0$, the integrand in (2.8) increases at most exponentially in t on the support of $\mathbb{P}$.

For reasons of space we postpone the detailed discussion of various concrete examples of matrix product probability measures to [2].

As a final example, we recall here the setup of the quantum instruments studied [3], as these were our initial motivation. We note that any matrix product measure can also be obtained by a well-chosen positive instrument (see [2]).

Example 2.26. Positive instruments. Let $\mathcal{H}$ be a finite-dimensional complex Hilbert space and denote by $\mathcal{C} = \mathcal{B}(\mathcal{H})$ the $*$-algebra of all linear maps $A : \mathcal{H} \to \mathcal{H}$ equipped with the inner product $(A, B) = \operatorname{tr}(A^*B)$. Let $\Phi : \mathcal{A} \to \mathcal{B}(\mathcal{C})$ be a map satisfying the following assumptions:

(B1) The map $\Phi(a)$ is positive[24] for all $a \in \mathcal{A}$.

(B2) The map $S = \sum_{a\in\mathcal{A}} \Phi(a)$ and its adjoint S^* satisfy

$$S[\nu] = \lambda\nu, \qquad S^*[\rho] = \lambda\rho$$

for some $\lambda > 0$ and strictly positive $\nu, \rho \in \mathcal{C}$.

For each $t \in \mathbb{N}$ we define a probability measure $\mathbb{P}_t$ on Ω_t by

$$\mathbb{P}_t(\omega_1, \dots, \omega_t) = \frac{1}{\lambda^t \operatorname{tr}(\rho\nu)} \operatorname{tr}(\rho(\Phi(\omega_1) \circ \cdots \circ \Phi(\omega_t))[\nu]).$$

One easily verifies that there exists a unique $\mathbb{P} \in \mathcal{P}_\varphi(\Omega)$ whose family of marginals is given by $(\mathbb{P}_t)_{t\in\mathbb{N}}$. We shall call such $\mathbb{P}$ the positive instrument process associated with the positive instrument (Φ, ν, ρ). If $\Phi(a)$ is completely positive[25] for all $a \in \mathcal{A}$, ν is the identity map, and $\lambda = 1$, then (Φ, ρ) is called a *quantum instrument* and $\mathbb{P}$ describes the statistics of the repeated quantum measurement process generated by (Φ, ρ); see [3] for additional information and references regarding quantum instruments and induced processes.

We have:

- **(UD)** holds for $\mathbb{P}$ with $\sup_t c_t < \infty$, [3, Lemma 3.4].

[24] $\Psi \in \mathcal{B}(\mathcal{C})$ is positive if $\Psi[X] \geq 0$ for any $X \geq 0$.

[25] $\Psi \in \mathcal{B}(\mathcal{C})$ is completely positive if for all $k \in \mathbb{N}$ the map $\mathrm{id}_k \otimes \Psi \in \mathcal{B}(\mathcal{B}(\mathbb{C}^k) \otimes \mathcal{C})$ is positive, where id_k is the identity map on $\mathcal{B}(\mathbb{C}^k)$.

- If $\Phi(a)$ is positivity improving[26] for all $a \in \mathcal{A}$, then **(SLD)** holds with $\tau_t \equiv 0$, $\sup_t c_t < \infty$.
- If $\Phi(a)$ is positivity improving for some $a \in \mathcal{A}$, then **(SLD)** holds with $\tau_t \equiv 1$, (by taking $\xi = a$ in (2.2)), and $\sup_t c_t < \infty$.
- Suppose that none of the $\Phi(a)$'s is positivity improving. If the map S is irreducible, *i.e.*, for some $r \in \mathbb{N}$ the map $(\imath + S)^r$ is positivity improving[27] ($\imath$ denotes the identity map on $\mathcal{B}(\mathcal{C})$), then **(SLD)** holds with $\tau_t \equiv r$ and $\sup_t c_t < \infty$. We remark that if S is irreducible, then **(B2)** automatically holds and $\mathbb{P}$ is ergodic.
- Let $\widehat{\Phi} : \mathcal{A} \to \mathcal{B}(\mathcal{C})$ be another map satisfying **(B1)** and **(B2)**, and let $\widehat{\mathbb{P}}$ be the induced positive instrument process. If the map $\sum_{a\in\mathcal{A}} \Phi(a) \otimes \widehat{\Phi}(a)$ acting on $\mathcal{C} \otimes \mathcal{C}$ is irreducible, then **(SSD)** holds for the pair $(\mathbb{P}, \widehat{\mathbb{P}})$ with τ_t and c_t that do not depend on t; see [3, Proposition 2.6].[28]
- If Θ is as in Definition 2.5 and $\widehat{\mathbb{P}} = \Theta\mathbb{P}$, then in Case 1 the measure $\widehat{\mathbb{P}}$ is the positive instrument process associated with $(\widehat{\Phi}, \nu, \rho)$, where $\widehat{\Phi}(a) = \Phi(u(a))$. In Case 2, $\widehat{\mathbb{P}}$ is the positive instrument process associated with $(\widehat{\Phi}, \rho, \nu)$, where $\widehat{\Phi}(a) = \Phi^*(u(a))$.
- Unlike for matrix product measures, we do not have in general that $q(\alpha) < \infty$ for all $\alpha \in \mathbb{R}$. See the *rotational instruments* in [2].

Again, for details and discussion of concrete examples we refer the reader to [2].

3 General constructions and abstract LDP

We start with some further notation and conventions that will be used throughout the paper. A function f on Ω is $\mathcal{F}_t$-measurable if and only if $f(\omega)$ depends only on $\omega_1, \ldots, \omega_t$. We identify the space of $\mathcal{F}_t$-measurable functions and the space[29] $C(\Omega_t)$ in the obvious way, and for such functions we write $f(\omega)$ and $f(\omega_1, \ldots, \omega_t)$ interchangeably. The space $C_{\text{fin}}(\Omega)$ consisting of all functions which are $\mathcal{F}_t$-measurable for some t is dense in $C(\Omega)$.

By this identification, any function $f \in C(\Omega_t)$ is associated with a $\mathcal{F}_t$-measurable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Conversely, any $\mathcal{F}_t$-measurable real-valued random variable f on $(\Omega, \mathcal{F}, \mathbb{P})$ is associated with a function $f \in C(\Omega_t^+)$, and can be extended to a function in $C(\Omega_t)$ by defining $f(w)$ arbitrarily for $w \in \Omega_t \setminus \Omega_t^+$. We note that with this convention, $\langle f, \mathbb{P}\rangle = \sum_{w\in\Omega_t^+} f(w)\mathbb{P}(w)$. These considerations extend to $\mathbb{R}^d$-valued functions, and the corresponding spaces are denoted by $C(\Omega, \mathbb{R}^d)$ and $C_{\text{fin}}(\Omega, \mathbb{R}^d)$.

Following these conventions, the quantity σ_t defined in (2.7) can be expressed as

$$\sigma_t(w) = \log \frac{\mathbb{P}_t(w)}{\widehat{\mathbb{P}}_t(w)}, \qquad w \in \Omega_t^+, \tag{3.1}$$

which is well defined since $\mathbb{P}_t \ll \widehat{\mathbb{P}}_t$.

[26] $\Psi \in \mathcal{B}(\mathcal{C})$ is positivity improving if $\Psi[X] > 0$ for any non-zero $X \geq 0$.
[27] See [12, Section 2.1].
[28] It is easy to realize that Assumption (C) in [2] together with **(UD)** imply **(SSD)**, see also [2].
[29] Since Ω_t is endowed with the discrete topology, all functions on Ω_t are continuous.

3.1 Construction of $\psi_{n,t}$

For any pair $(t,n)\in\mathbb{N}^2$ with $t\geq n$, we define[30]

$$N=N(t,n)=2\left\lfloor\frac{t}{2(n+\tau_n)}\right\rfloor \qquad\text{and}\qquad t'=t'(t,n)=Nn, \tag{3.2}$$

where $(\tau_n)_{n\in\mathbb{N}}$ is the integer sequence introduced in Section 2.2. Observe that N is even. An important inequality following from the above definition is

$$\frac{t}{1+\tau_n/n}-2n\leq t'(t,n)\leq t,$$

which implies that

$$\lim_{n\to\infty}\limsup_{t\to\infty}\left|\frac{t'(t,n)}{t}-1\right|\leq\lim_{n\to\infty}\limsup_{t\to\infty}\left(\frac{2n}{t}+\frac{\tau_n}{n}\frac{1}{1+\tau_n/n}\right)=0. \tag{3.3}$$

For each $n\in\mathbb{N}$, we define the decoupled measure $\mathbb{P}^{(n)}=(\mathbb{P}_n)^{\times\mathbb{N}}$ (which is φ^n-invariant, but not φ-invariant). For $t=mn+j$ with $0\leq j<n$, the marginal $\mathbb{P}_t^{(n)}$ is given by

$$\mathbb{P}_t^{(n)}(w)=\left(\prod_{k=0}^{m-1}\mathbb{P}_n(w_{[kn+1,(k+1)n]})\right)\mathbb{P}_j(w_{[mn+1,mn+j]}),$$

where the last term is 1 if $j=0$. We also define

$$\Lambda_{t'}=(\Omega_n^+)^N=\{w\in\Omega_{t'}:\mathbb{P}_n(w_{[kn+1,(k+1)n]})>0,\ k=0,1,\dots,N-1\}, \tag{3.4}$$

which is the support of $\mathbb{P}_{t'}^{(n)}$. Note that obviously $\Omega_{t'}^+\subset\Lambda_{t'}$.

The main result of this subsection is the following proposition that provides a way to compare the two discrete probability spaces $(\Omega_{t'},\mathbb{P}_{t'}^{(n)})$ and $(\Omega_t,\mathbb{P}_t)$.

Proposition 3.1. *Assume* **(SLD)***. For any pair $(t,n)\in\mathbb{N}^2$ with $t\geq n$, and with N and t' defined by* (3.2)*, there exists a map $\psi_{n,t}:\Omega_{t'}\to\Omega_t$ such that the following holds.*

1. *There exists $g(n,t)\geq 0$ such that*

$$\mathbb{P}_{t'}^{(n)}\circ\psi_{n,t}^{-1}\leq e^{g(n,t)}\,\mathbb{P}_t, \tag{3.5}$$

$$\lim_{n\to\infty}\limsup_{t\to\infty}\frac{1}{t}g(n,t)=0. \tag{3.6}$$

2. *Assume furthermore* **(SSD)***. Then $\psi_{n,t}$ can be chosen so that, in addition to the above,*

$$\lim_{n\to\infty}\limsup_{t\to\infty}\frac{1}{t}\sup_{w\in\Lambda_{t'}}\left|\sigma_t(\psi_{n,t}(w))-\sum_{k=0}^{N-1}\sigma_n(w_{[kn+1,(k+1)n]})\right|=0. \tag{3.7}$$

Remark 3.2. For further reference, we make the immediate observation that (3.5) is equivalent to the fact that for all $A\subset\Omega_{t'}$,

$$\mathbb{P}_{t'}^{(n)}(A)\leq e^{g(n,t)}\mathbb{P}_t(\psi_{n,t}(A)), \tag{3.8}$$

and to the fact that for each function $h:\Omega_t\to[0,\infty)$ we have

$$\sum_{w\in\Omega_{t'}}h(\psi_{n,t}(w))\mathbb{P}_{t'}^{(n)}(w)\leq e^{g(n,t)}\sum_{w\in\Omega_t}h(w)\mathbb{P}_t(w). \tag{3.9}$$

[30] $\lfloor x\rfloor=\max\{n\in\mathbb{Z}:n\leq x\}$ denotes the floor function.

Remark 3.3. It follows from (3.8) that $\psi_{n,t}(\Lambda_{t'}) \subset \Omega_t^+$. In particular, all the quantities in (3.7) are well defined (see (3.1)). Note that the map $\psi_{n,t}$ is in general neither injective nor surjective. There will be two contributions in $g(n,t)$: one coming from the ratio $\mathbb{P}_t(\psi_{n,t}(w))/\mathbb{P}_{t'}^{(n)}(w)$, and one coming from the maximal number of points $w \in \Omega_{t'}$ which share the same image $\psi_{n,t}(w) \in \Omega_t$.

Remark 3.4. The structure of $\psi_{n,t}$ here is very similar to a construction used in Section 2 of [15] in the context of products of matrices.

We start with two technical lemmas.

Lemma 3.5. *There exists a constant C such that the following holds. For all $t, k \in \mathbb{N}$ and all $v \in \Omega_t$, there exists $b \in \Omega_k$ such that*

$$\mathbb{P}(bv) \geq \mathbb{P}(v)\mathrm{e}^{-Ck}. \tag{3.10}$$

Assuming **(SSD)***, b can be chosen so that, in addition to the above,*

$$\widehat{\mathbb{P}}(bv) \geq \widehat{\mathbb{P}}(v)\mathrm{e}^{-Ck}. \tag{3.11}$$

Proof. The first statement holds with $C = \log |\mathcal{A}|$, since for all $v \in \Omega_t$ we have

$$\mathbb{P}(v) = \sum_{b\in\Omega_k} \mathbb{P}(bv) \leq |\mathcal{A}|^k \max_{b\in\Omega_k} \mathbb{P}(bv).$$

The second statement is less trivial because (3.10) and (3.11) have to hold for the same b. Fix a symbol $a \in \mathcal{A}$ such that $\mathbb{P}_1(a)\widehat{\mathbb{P}}_1(a) > 0$ (which is possible by the absolute continuity condition (2.3)). We claim that the result holds with $C = c_1 - \log(\mathbb{P}_1(a) \wedge \widehat{\mathbb{P}}_1(a))$. Assume first that $k = 1$. Then for all $v \in \Omega_t$, there exists $\xi \in \Omega_{\text{fin}}$ such that $|\xi| \leq \tau_1$ and

$$\mathbb{P}^\sharp(a\xi v) \geq \mathrm{e}^{-c_1}\mathbb{P}^\sharp(a)\mathbb{P}^\sharp(v) \geq \mathrm{e}^{-C}\mathbb{P}^\sharp(v)$$

for both $\mathbb{P}^\sharp = \mathbb{P}$ and $\mathbb{P}^\sharp = \widehat{\mathbb{P}}$. Now, let $b = a$ if $|\xi| = 0$, and $b = \xi_1$ otherwise. We then have

$$\mathbb{P}^\sharp(bv) \geq \mathbb{P}^\sharp(a\xi v) \geq \mathrm{e}^{-C}\mathbb{P}^\sharp(v),$$

which shows that both (3.10) and (3.11) hold in the case $k = 1$. The general statement follows by induction on k. □

The following lemma is immediate.

Lemma 3.6. *Let (X_1, P_1) and (X_2, P_2) be two discrete probability spaces (each with its discrete σ-algebra). Let $\psi : X_1 \to X_2$ be a mapping which is at most r-to-one, and assume that $P_1(\omega) \leq cP_2(\psi(\omega))$ for some $c > 0$ and all $\omega \in X_1$. Then*

$$P_1 \circ \psi^{-1} \leq crP_2.$$

We can now prove the main result of this subsection.

Proof of Proposition 3.1. The map $\psi_{n,t} : \Omega_{t'} \to \Omega_t$ is constructed as follows. For $w \in \Omega_{t'}$, we write $w = w^1 w^2 \dots w^N$ with $w^i \in \Omega_n$, and define

$$\psi_{n,t}(w) = bw^1\xi^1 w^2\xi^2 \dots w^{N-1}\xi^{N-1}w^N \tag{3.12}$$

for some $\xi^i, b \in \Omega_{\text{fin}}$ to be chosen below that will satisfy $|\xi^i| \leq \tau_n$ and

$$|b| = \delta := t - t' - \sum_{i=1}^{N-1} |\xi^i|$$

(which may be zero), so that $|\psi_{n,t}(w)| = t$. Observe that

$$t - t' \geq \delta \geq t - N(n + \tau_n) \geq 0. \tag{3.13}$$

Using **(SLD)**, we first choose ξ^{N-1} such that $|\xi^{N-1}| \leq \tau_n$ and

$$\mathbb{P}\left(w^{N-1}\xi^{N-1}w^N\right) \geq \mathrm{e}^{-c_n}\mathbb{P}\left(w^{N-1}\right)\mathbb{P}\left(w^N\right).$$

Next, we choose ξ^{N-2} such that $|\xi^{N-2}| \leq \tau_n$ and

$$\begin{aligned}\mathbb{P}\left(w^{N-2}\xi^{N-2}w^{N-1}\xi^{N-1}w^N\right) &\geq \mathrm{e}^{-c_n}\mathbb{P}\left(w^{N-2}\right)\mathbb{P}\left(w^{N-1}\xi^{N-1}w^N\right)\\ &\geq \mathrm{e}^{-2c_n}\mathbb{P}\left(w^{N-2}\right)\mathbb{P}\left(w^{N-1}\right)\mathbb{P}\left(w^N\right).\end{aligned}$$

Continuing this process, we choose $\xi^{N-3}, \dots, \xi^1$ such that $|\xi^i| \leq \tau_n$ and

$$\mathbb{P}\left(w^1\xi^1w^2\xi^2\dots w^{N-1}\xi^{N-1}w^N\right) \geq \mathrm{e}^{-(N-1)c_n}\mathbb{P}\left(w^1\right)\mathbb{P}\left(w^2\right)\cdots\mathbb{P}\left(w^N\right) = \mathrm{e}^{-(N-1)c_n}\mathbb{P}^{(n)}(w).$$

Finally, if $\delta \geq 1$, we choose $b \in \Omega_\delta$ so that (3.10) holds with $v = w^1\xi^1w^2\xi^2\dots w^{N-1}\xi^{N-1}w^N$ and $k = \delta$, so that

$$\mathbb{P}(\psi_{n,t}(w)) \geq \mathrm{e}^{-(N-1)c_n - C\delta}\mathbb{P}^{(n)}(w).$$

If $\delta = 0$, we choose b as the empty word, and the above also holds. Next, (3.13) implies that

$$(N-1)c_n + C\delta \leq (N-1)c_n + (t - t')C =: g_1(n,t),$$

and so

$$\mathbb{P}(\psi_{n,t}(w)) \geq \mathrm{e}^{-g_1(n,t)}\mathbb{P}^{(n)}(w). \tag{3.14}$$

The mapping $\psi_{n,t}$ is not injective. In order to retrieve $w \in \Omega_{t'}$ from $\psi_{n,t}(w)$, it suffices to know the length of $\xi^1, \dots, \xi^{N-1}$, and there are at most $(\tau_n + 1)^{N-1}$ possibilities. Thus, $\psi_{n,t}$ is at most $(\tau_n + 1)^{N-1}$-to-one. By Lemma 3.6, we obtain (3.5) with

$$g(n,t) = g_1(n,t) + (N-1)\log(\tau_n + 1) \leq g_1(n,t) + N\tau_n.$$

To finish the proof of Part 1, observe that since $t \geq t' = nN$, we have

$$\frac{g(n,t)}{t} \leq \frac{c_n}{n} + C\left(1 - \frac{t'(t,n)}{t}\right) + \frac{\tau_n}{n},$$

which by (3.3), shows that (3.6) also holds.

To prove Part 2 of the proposition, assume **(SSD)** and let $w \in \Lambda_{t'}$. We then proceed exactly as above. By **(SSD)** one can choose $\xi^1, \dots, \xi^{N-1}$ such that

$$\begin{aligned}\mathrm{e}^{-(N-1)c_n}\mathbb{P}^\sharp\left(w^1\right)\mathbb{P}^\sharp\left(w^2\right)\cdots\mathbb{P}^\sharp\left(w^N\right) &\leq \mathbb{P}^\sharp\left(w^1\xi^1w^2\xi^2\dots w^{N-1}\xi^{N-1}w^N\right)\\ &\leq \mathrm{e}^{(N-1)c_n}\mathbb{P}^\sharp\left(w^1\right)\mathbb{P}^\sharp\left(w^2\right)\cdots\mathbb{P}^\sharp\left(w^N\right)\end{aligned}$$

for both $\mathbb{P}^\sharp = \mathbb{P}$ and $\mathbb{P}^\sharp = \widehat{\mathbb{P}}$. Note that all quantities here are positive, since $w \in \Lambda_{t'}$ implies, by definition, that all the w^i are in the support of $\mathbb{P}_n$, and hence in that of $\widehat{\mathbb{P}}_n$ by (2.3).

Defining δ as above and choosing $b \in \Omega_\delta$ as in Lemma 3.5 (with $b = \kappa$ if $\delta = 0$), we obtain that $\psi_{n,t}(w)$ defined by (3.12) satisfies

$$\mathrm{e}^{-(N-1)c_n - C\delta}\mathbb{P}^\sharp\left(w^1\right)\cdots\mathbb{P}^\sharp\left(w^N\right) \le \mathbb{P}^\sharp(\psi_{n,t}(w)) \le \mathrm{e}^{(N-1)c_n}\mathbb{P}^\sharp\left(w^1\right)\cdots\mathbb{P}^\sharp\left(w^N\right).$$

Recalling definition of σ_t and σ_n and using the inequality

$$(N-1)c_n \le (N-1)c_n + C\delta \le g_1(n,t) \le g(n,t),$$

we finally obtain

$$\left|\sigma_t(\psi_{n,t}(w)) - \sum_{k=1}^N \sigma_n(w^k)\right| \le 2g(n,t),$$

which implies (3.7). This completes the proof of Proposition 3.1. □

3.2 Ruelle–Lanford functions

Let X be a locally convex Hausdorff topological vector space endowed with its Borel σ-algebra. Let $\mathcal{N}_0$ be a neighborhood basis of $0 \in X$, so that $\mathcal{N}_x = \mathcal{N}_0 + x$ is a neighborhood basis of $x \in X$. Given a sequence $(z_t)_{t\in\mathbb{N}}$ of X-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, we define the following two non-decreasing set functions on the Borel sets of X:

$$\begin{aligned}\underline{s}(A) &= \liminf_{t\to\infty} \frac{1}{t}\log\mathbb{P}\left(\frac{1}{t}z_t \in A\right),\\ \overline{s}(A) &= \limsup_{t\to\infty} \frac{1}{t}\log\mathbb{P}\left(\frac{1}{t}z_t \in A\right).\end{aligned} \tag{3.15}$$

Definition 3.7. *Assume that for all $x \in X$, we have*

$$\inf_{G\in\mathcal{N}_x} \underline{s}(G) = \inf_{G\in\mathcal{N}_x} \overline{s}(G). \tag{3.16}$$

Then, the function $s : X \to [-\infty, 0]$, whose value $s(x)$ is defined by the two expressions in (3.16) for each $x \in X$, is called the Ruelle–Lanford function *of the sequence $(z_t)_{t\in\mathbb{N}}$.*[31]

When defined, the function s is upper semicontinuous. Indeed, for all $x \in X$ and $\varepsilon > 0$, there exists $G \in \mathcal{N}_x$ such that $\overline{s}(G) \le s(x) + \varepsilon$, and for each $x' \in G$ there exists $G' \in \mathcal{N}_{x'}$ such that $G' \subset G$. It follows that $s(x') \le \overline{s}(G') \le \overline{s}(G) \le s(x) + \varepsilon$.

We now give sufficient conditions for the Ruelle–Lanford function to exist.

Definition 3.8. *We say that the sequence $(z_t)_{t\in\mathbb{N}}$ is* admissible *if for all $x_1, x_2 \in X$ and for every neighborhood G of $x := \frac{1}{2}x_1 + \frac{1}{2}x_2$, there exist $G_1 \in \mathcal{N}_{x_1}$ and $G_2 \in \mathcal{N}_{x_2}$ such that*

$$\underline{s}(G) \ge \frac{1}{2}\overline{s}(G_1) + \frac{1}{2}\overline{s}(G_2). \tag{3.17}$$

Proposition 3.9. *Let the sequence $(z_t)_{t\in\mathbb{N}}$ be admissible. Then (3.16) holds for all $x \in X$, so that the Ruelle–Lanford function s is well defined. Moreover, $(\frac{1}{t}z_t)_{t\in\mathbb{N}}$ satisfies the* weak *LDP with convex rate function $-s$, in the sense that for every open set $O \subset X$,*

$$\underline{s}(O) \ge \sup_{x\in O} s(x) \tag{3.18}$$

[31] The two infima in (3.16) are independent of the choice of the neighborhood basis $\mathcal{N}_x$ of x, and hence so is $s(x)$.

and that for every compact *set* $\Gamma \subset X$*,*

$$\overline{s}(\Gamma) \leq \sup_{x \in \Gamma} s(x). \tag{3.19}$$

If, in addition, the laws of $(\frac{1}{t} z_t)_{t \in \mathbb{N}}$ *form an exponentially tight family*[32]*, then* $-s$ *is a good rate function, and* $(\frac{1}{t} z_t)_{t \in \mathbb{N}}$ *satisfies the LDP, i.e., (3.19) holds for any closed set* $\Gamma \subset X$*.*

Proof. For the reader's convenience, we include a complete proof, although this is a classical result (see [34, Proposition 3.5] or [9, Lemmas 4.1.11 and 4.1.21]). First, the special case $x = x_1 = x_2$ in (3.17) immediately implies that the two infima in (3.16) are equal, so that s is well defined. Next, if $x = \frac{1}{2}x_1 + \frac{1}{2}x_2$, then (3.17) yields $s(x) \geq \frac{1}{2}(s(x_1) + s(x_2))$. Since s is upper semicontinuous, this inequality implies that s is concave (by a bisection method).

We now turn to the LDP. The lower bound (3.18) is immediate. Indeed, for any open set $O \subset X$ and every $x \in O$ we have $x \in G \subset O$ for some $G \in \mathcal{N}_x$, so that $\underline{s}(O) \geq \underline{s}(G) \geq s(x)$. Since this holds for all $x \in O$, we obtain (3.18).

The upper bound (3.19) is more involved. Let $\Gamma \subset X$ be closed, and let $\varepsilon > 0$. It suffices to prove (3.19) in the following two cases.

- Case 1: the laws of $(\frac{1}{t} z_t)_{t \in \mathbb{N}}$ are exponentially tight. Then, there exists a compact set K such that $\overline{s}(K^c) \leq -1/\varepsilon$. We let then $G_0 = \Gamma \cap K^c$. Thus, for each $x \in G_0$, we have $s(x) \leq -1/\varepsilon$.
- Case 2: Γ is compact. Then we let $G_0 = \emptyset$.

Observe that in both cases $\Gamma \setminus G_0$ is compact. For each $x \in \Gamma \setminus G_0$, there exists $G(x) \in \mathcal{N}_x$ such that $\overline{s}(G(x)) \leq s(x) + \varepsilon \leq \sup_{y \in \Gamma} s(y) + \varepsilon$. Now, $\{G(x) : x \in \Gamma \setminus G_0\}$ is an open cover of $\Gamma \setminus G_0$, and by compactness one can extract a finite subcover $\{G_i : i = 1, \ldots, n\}$. Since $\Gamma \subset \bigcup_{i=0}^n G_i$, one has

$$\overline{s}(\Gamma) \leq \limsup_{t \to \infty} \frac{1}{t} \log \left(\sum_{i=0}^{n} \mathbb{P}\left(\frac{z_t}{t} \in G_i \right) \right) \leq \max_{i=0,\ldots,n} \overline{s}(G_i) \leq \max(-1/\varepsilon, \sup_{x \in \Gamma} s(x) + \varepsilon).$$

Sending $\varepsilon \to 0$ completes the proof of (3.19).

Finally, we show that exponential tightness implies the goodness of the rate function $I := -s$ (see for example [9, Lemma 1.2.18]). Let $a \in \mathbb{R}$, and let $L_a = \{x \in X : I(x) \leq a\}$ be the corresponding level set (which is closed by lower semicontinuity of I). Assuming exponential tightness, there is a compact set K such that $\overline{s}(K^c) < -a$, and applying (3.18) to $O = K^c$ yields $\inf_{x \in K^c} I(x) > a$. Thus, $L_a \subset K$, and hence L_a is compact. □

3.3 Compatible observables

In this subsection we assume **(SLD)**, so that the map $\psi_{n,t}$ is well defined and Part 1 of Proposition 3.1 holds. Moreover, $N = N(t, n)$ and $t' = t'(t, n)$ are as in (3.2) and $\Lambda_{t'}$ is as in (3.4). Finally, for $x \in X = \mathbb{R}^d$, we choose the neighborhood basis $\mathcal{N}_x = \{B(x, \varepsilon)\}_{\varepsilon > 0}$, where $B(x, \varepsilon)$ denotes the open ball of radius ε in X, centered at x.

Definition 3.10. *Let* $X = \mathbb{R}^d$ *and let* $(z_t)_{t \in \mathbb{N}}$ *be a sequence of* X*-valued random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$*. We say that* $(z_t)_{t \in \mathbb{N}}$ *is* ψ*-compatible if* z_t *is* $\mathcal{F}_t$*-measurable for each* t*, and the quantity*

$$h(n, t) := \frac{1}{t} \sup_{w \in \Lambda_{t'}} \left| z_t(\psi_{n,t}(w)) - \sum_{k=0}^{N-1} z_n(w_{[kn+1,(k+1)n]}) \right|$$

[32] This means that for any $\epsilon > 0$ there exists a compact set K_ϵ such that $\overline{s}(K_\epsilon^c) \leq -1/\epsilon$.

satisfies

$$\lim_{n\to\infty}\limsup_{t\to\infty} h(n,t) = 0. \tag{3.20}$$

Proposition 3.11. *Let $(z_t)_{t\in\mathbb{N}}$ be a sequence of ψ-compatible $\mathbb{R}^d$-valued random variables on $(\Omega,\mathcal{F},\mathbb{P})$. Then the following holds.*

1. *For all $x=\frac{1}{2}x_1+\frac{1}{2}x_2\in\mathbb{R}^d$ and $0<\varepsilon<\varepsilon'$, we have*

$$\underline{s}(B(x,\varepsilon'))\geq \frac{1}{2}\overline{s}(B(x_1,\varepsilon))+\frac{1}{2}\overline{s}(B(x_2,\varepsilon)).$$

In particular, $(z_t)_{t\in\mathbb{N}}$ is admissible and the conclusions of Proposition 3.9 hold.

2. *There exists a sequence $(\gamma_t)_{t\in\mathbb{N}}$ with $\gamma_t\to 0$ such that for all $\varepsilon>0$, all $t\in\mathbb{N}$ and all $x\in\mathbb{R}^d$,*[33]

$$\frac{1}{t}\log\mathbb{P}\left(\frac{z_t}{t}\in B(x,\varepsilon)\right)\leq \gamma_t+\sup_{y\in B(x,\varepsilon+(1+|x|)\gamma_t)} s(y), \tag{3.21}$$

where s is as in Definition 3.7 and Proposition 3.9.

Proof. We shall prove that there exists a sequence $(\gamma_t)_{t\in\mathbb{N}}$ with $\gamma_t\to 0$ such that for all $\varepsilon>0$, all $t\in\mathbb{N}$ and all $x=\frac{1}{2}x_1+\frac{1}{2}x_2\in\mathbb{R}^d$, we have

$$\frac{1}{2t}\log\mathbb{P}\left(\frac{z_t}{t}\in B(x_1,\varepsilon)\right)+\frac{1}{2t}\log\mathbb{P}\left(\frac{z_t}{t}\in B(x_2,\varepsilon)\right)\leq \underline{s}(B(x,\varepsilon+(1+|x|)\gamma_t))+\gamma_t. \tag{3.22}$$

This relation yields both Part 1 and Part 2 of the proposition. Part 1 follows by taking the limit $t\to\infty$ and using that, for fixed x, we have

$$(1+|x|)\gamma_t\leq\varepsilon'-\varepsilon$$

for t large enough. For Part 2, we take $x_1=x_2=x$ in (3.22), and by (3.19) we obtain (3.21) with the ball $B(x,\varepsilon+(1+|x|)\gamma_t)$ replaced by its closure in the right-hand side. Replacing γ_t with γ_t+t^{-1}, we then obtain (3.21).

To prove (3.22), let $x=\frac{1}{2}x_1+\frac{1}{2}x_2\in\mathbb{R}^d$ and $\varepsilon>0$. In the following, $j(k)=1$ if k is odd, $j(k)=2$ if k is even, and we write $I_k=[kn+1,(k+1)n]$. By assumption, for any $w\in\Lambda_{t'}$,

$$\begin{aligned}\left|\frac{1}{t}z_t(\psi_{n,t}(w))-x\right| &\leq \left|\frac{1}{t}\sum_{k=0}^{N-1}z_n(w_{I_k})-x\right|+h(n,t)\\ &\leq \left|\frac{n}{t}\sum_{k=0}^{N-1}\left(\frac{z_n(w_{I_k})}{n}-x\right)\right|+h(n,t)+|x|\left(1-\frac{t'}{t}\right)\\ &= \left|\frac{n}{t}\sum_{k=0}^{N-1}\left(\frac{z_n(w_{I_k})}{n}-x_{j(k)}\right)\right|+h(n,t)+|x|\left(1-\frac{t'}{t}\right),\end{aligned}$$

where the last equality holds because N is even and $x=\frac{1}{2}x_1+\frac{1}{2}x_2$. Using further that $\frac{n}{t}\leq\frac{1}{N}$ leads to

$$\left|\frac{1}{t}z_t(\psi_{n,t}(w))-x\right|\leq\frac{1}{N}\sum_{k=0}^{N-1}\left|\frac{z_n(w_{I_k})}{n}-x_{j(k)}\right|+h(n,t)+|x|\left(1-\frac{t'}{t}\right).$$

[33] Note that the bound is uniform in both t and x. This will be crucial in the proof of Proposition 5.2 below.

Let $u_n = n^{-1} + \limsup_{t\to\infty} \max\left(h(n,t), 1-\frac{t'}{t}\right)$. By (3.3) and (3.20) we have $u_n \to 0$. In addition, for each fixed n, there exists $t_0(n)$ such that for all $t \geq t_0(n)$,

$$\left|\frac{1}{t} z_t(\psi_{n,t}(w)) - x\right| \leq \frac{1}{N}\sum_{k=0}^{N-1}\left|\frac{z_n(w_{I_k})}{n} - x_{j(k)}\right| + (1+|x|)u_n,$$

and hence

$$\psi_{n,t}\left(\bigcap_{k=0}^{N-1}\left\{w \in \Lambda_{t'} : \frac{1}{n} z_n(w_{I_k}) \in B(x_{j(k)}, \varepsilon)\right\}\right) \subset \left\{w \in \Omega_t^+ : \frac{1}{t} z_t(w) \in B'\right\},$$

where $B' = B(x, \varepsilon + (1+|x|)u_n)$. Using (3.8) and translation invariance, for all $t \geq t_0(n)$ we derive

$$\begin{aligned}
\mathbb{P}\left(\frac{z_t}{t} \in B'\right) &\geq \mathbb{P}_t\left(\psi_{n,t}\left(\bigcap_{k=0}^{N-1}\left\{w \in \Lambda_{t'} : \frac{1}{n} z_n(w_{I_k}) \in B(x_{j(k)}, \varepsilon)\right\}\right)\right) \\
&\geq \mathrm{e}^{-g(n,t)} \mathbb{P}_{t'}^{(n)}\left(\bigcap_{k=0}^{N-1}\left\{w \in \Lambda_{t'} : \frac{1}{n} z_n(w_{I_k}) \in B(x_{j(k)}, \varepsilon)\right\}\right) \\
&= \mathrm{e}^{-g(n,t)}\left(\mathbb{P}_n\left(\frac{z_n}{n} \in B(x_1, \varepsilon)\right)\right)^{\frac{N}{2}}\left(\mathbb{P}_n\left(\frac{z_n}{n} \in B(x_2, \varepsilon)\right)\right)^{\frac{N}{2}}.
\end{aligned}$$

Using also that $\frac{N}{t} \leq \frac{1}{n}$, and sending $t \to \infty$, we obtain

$$\underline{s}(B(x, \varepsilon + (1+|x|)u_n)) \geq \frac{1}{2n}\log \mathbb{P}\left(\frac{z_n}{n} \in B(x_1, \varepsilon)\right) + \frac{1}{2n}\log \mathbb{P}\left(\frac{z_n}{n} \in B(x_2, \varepsilon)\right) - u_n',$$

where $u_n' = \limsup_{t\to\infty} \frac{g(n,t)}{t}$. By (3.6) we have $u_n' \to 0$. Defining $\gamma_n = \max(u_n, u_n')$ completes the proof of (3.22). □

Lemma 3.12. *Let $(z_t)_{t\in\mathbb{N}}$ be a sequence of ψ-compatible $\mathbb{R}^d$-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then, for all $\alpha \in \mathbb{R}^d$, the limit*

$$q(\alpha) := \lim_{t\to\infty} \frac{1}{t}\log\left\langle \mathrm{e}^{(\alpha, z_t)}, \mathbb{P}\right\rangle$$

exists and takes value in $(-\infty, \infty]$. Moreover, the function $\mathbb{R}^d \ni \alpha \mapsto q(\alpha)$ is convex and lower semicontinuous.

Proof. Let $h(n,t)$ be as in Definition 3.10, and consider

$$A_t(\alpha) = \left\langle \mathrm{e}^{(\alpha, z_t)}, \mathbb{P}\right\rangle.$$

For each finite t, the map $\alpha \mapsto A_t(\alpha)$ is continuous. Recall that by definition of ψ-compatibility, z_n is $\mathcal{F}_n$-measurable. Thus, by invariance and (3.9),

$$\begin{aligned}
(A_n(\alpha))^N &= \sum_{w\in\Lambda_{t'}} \exp\left\{\sum_{k=0}^{N-1}\left(\alpha, z_n(w_{[kn+1,(k+1)n]})\right)\right\}\mathbb{P}_{t'}^{(n)}(w) \\
&\leq \mathrm{e}^{|\alpha| t h(n,t)} \sum_{w\in\Lambda_{t'}} \mathrm{e}^{(\alpha, z_t(\psi_{n,t}(w)))}\mathbb{P}_{t'}^{(n)}(w) \\
&\leq \mathrm{e}^{|\alpha| t h(n,t) + g(n,t)} \sum_{w\in\Omega_t^+} \mathrm{e}^{(\alpha, z_t(w))}\mathbb{P}_t(w) \\
&= \mathrm{e}^{|\alpha| t h(n,t) + g(n,t)} A_t(\alpha).
\end{aligned}$$

It follows that

$$\frac{1}{t}\log A_t(\alpha) \geq \frac{N}{t}\log A_n(\alpha) - |\alpha| h(n,t) - \frac{g(n,t)}{t} = \frac{1}{n}\frac{t'}{t}\log A_n(\alpha) - |\alpha| h(n,t) - \frac{g(n,t)}{t}.$$

By (3.3), (3.6) and (3.20), there exists $\delta_n \to 0$ such that

$$\liminf_{t\to\infty} \frac{1}{t}\log A_t(\alpha) \geq \frac{1}{n}(1+\delta_n)\log A_n(\alpha) - (1+|\alpha|)\delta_n\,. \tag{3.23}$$

Taking now the $\limsup$ as $n \to \infty$ yields

$$\liminf_{t\to\infty} \frac{1}{t}\log A_t(\alpha) \geq \limsup_{n\to\infty} \frac{1}{n}\log A_n(\alpha),$$

and so $q(\alpha)$ exists. Combining this with (3.23), we derive

$$q(\alpha) = \sup_{n\in\mathbb{N}} \left(\frac{1}{n}(1+\delta_n)\log A_n(\alpha) - (1+|\alpha|)\delta_n\right).$$

It follows that $q(\alpha) > -\infty$ for all $\alpha \in \mathbb{R}$, and since the right-hand-side is a supremum over a family of continuous functions (with respect to α), we also derive that q is lower semicontinuous. Finally, it follows from Hölder's inequality that the functions $\alpha \mapsto \frac{1}{t}\log A_t(\alpha)$ are convex and, hence, so is the limit q. □

4 Level-1 LDP

In this section we assume again **(SLD)**. Thus, Part 1 of Proposition 3.1 holds, and again $N = N(t,n)$ and $t' = t'(t,n)$ are as in (3.2).

Lemma 4.1. *Let $f \in C(\Omega_r, \mathbb{R}^d)$ for some $r \in \mathbb{N}$ and set $z_t = S_{t-r+1}f$. Then, $(z_t)_{t\in\mathbb{N}}$ is ψ-compatible. (We take the convention that $S_j f = 0$ if $j \leq 0$).*

Proof. Clearly z_t is $\mathcal{F}_t$-measurable. Recall that by its definition (3.12), $\psi_{n,t}$ is expressed as

$$\psi_{n,t}(w) = b w_{[1,n]}\xi^1 w_{[n+1,2n]}\xi^2 \dots \xi^{N-1} w_{[(N-1)n+1,Nn]}, \qquad w \in \Omega_{t'}.$$

For $n \geq r$ and t large enough, we have

$$\begin{aligned} h(n,t) &= \frac{1}{t}\sup_{w\in\Lambda_{t'}} \left| z_t(\psi_{n,t}(w)) - \sum_{k=0}^{N-1} z_n(w_{[kn+1,(k+1)n]})\right| \\ &= \frac{1}{t}\sup_{w\in\Lambda_{t'}} \left| \sum_{s=0}^{t-r} f\big(\varphi^s(\psi_{n,t}(w)\big) - \sum_{k=0}^{N-1}\sum_{s=0}^{n-r} f\big(\varphi^s(w_{[kn+1,(k+1)n]})\big)\right| \\ &\leq \frac{\|f\|}{t}(t-r+1-N(n-r+1)) \leq \|f\|\frac{(t-t'+Nr)}{t} \leq \|f\|\left(1-\frac{t'}{t}+\frac{r}{n}\right), \end{aligned}$$

where the first inequality follows from the observation that all the terms of the iterated sum are also present in the first sum. By (3.3), it follows that (3.20) holds, and so $(z_t)_{t\in\mathbb{N}}$ is ψ-compatible. □

Proposition 4.2. *For all $f \in C(\Omega, \mathbb{R}^d)$ and all $\alpha \in \mathbb{R}^d$, the limit*

$$q_f(\alpha) := \lim_{t\to\infty} \frac{1}{t}\log\left\langle \mathrm{e}^{(\alpha, S_t f)}, \mathbb{P}\right\rangle$$

exists and is finite. Moreover, the map $(f,\alpha) \mapsto q_f(\alpha)$ is convex in both arguments, $|\alpha|$-Lipschitz with respect to f, and $\|f\|$-Lipschitz with respect to α.

Proof. For each t, the function

$$(\alpha, f) \mapsto \frac{1}{t} \log \left\langle \mathrm{e}^{(\alpha, S_t f)}, \mathbb{P} \right\rangle$$

has the convexity and Lipschitz properties stated in the proposition (convexity follows again from Hölder's inequality). By Lemmas 3.12 and 4.1, for every $r \in \mathbb{N}$ and $f \in C(\Omega_r, \mathbb{R}^d)$, the limit

$$\lim_{t\to\infty} \frac{1}{t} \log \left\langle \mathrm{e}^{(\alpha, S_t f)}, \mathbb{P} \right\rangle = \lim_{t\to\infty} \frac{1}{t} \log \left\langle \mathrm{e}^{(\alpha, S_{t-r+1} f)}, \mathbb{P} \right\rangle$$

exists and is finite for all $\alpha \in \mathbb{R}^d$. Thus, $q_f(\alpha)$ exists for all $f \in C_{\mathrm{fin}}(\Omega, \mathbb{R}^d)$ and $\alpha \in \mathbb{R}^d$. Since $C_{\mathrm{fin}}(\Omega, \mathbb{R}^d)$ is dense in $C(\Omega, \mathbb{R}^d)$, the $|\alpha|$-Lipschitz continuity in f implies that the limit also exists for all $f \in C(\Omega, \mathbb{R}^d)$. The convexity and Lipschitz properties are preserved in the limit. □

Proposition 4.3. *Let $f \in C(\Omega, \mathbb{R}^d)$ and set $z_t = S_t f$. Then, $(z_t)_{t\in\mathbb{N}}$ is admissible and the laws of $(\frac{1}{t} z_t)_{t\in\mathbb{N}}$ are exponentially tight. Thus $(\frac{1}{t} z_t)_{t\in\mathbb{N}}$ satisfies the LDP (see (2.5) and (2.6)) with good convex rate function I_f, where I_f is the Fenchel–Legendre transform of q_f.*

Proof. We first prove the admissibility claim, with $X = \mathbb{R}^d$. Let $x = \frac{1}{2}x_1 + \frac{1}{2}x_2 \in \mathbb{R}^d$ and let $\varepsilon > 0$. Since f is continuous, for $\delta = \varepsilon/6$ there exists an integer $r \geq 1$ and an $\mathcal{F}_r$-measurable function $\widetilde{f}$ such that $\|f - \widetilde{f}\| \leq \delta$. Define now $\widetilde{z}_t = S_{t-r+1}\widetilde{f}$, which is $\mathcal{F}_t$-measurable. We have

$$\left\| \frac{1}{t} z_t - \frac{1}{t} \widetilde{z}_t \right\| = \left\| \frac{1}{t} S_t f - \frac{1}{t} S_{t-r+1} \widetilde{f} \right\| \leq \delta + \frac{r-1}{t} \|f\|. \tag{4.1}$$

By Lemma 4.1, $(\widetilde{z}_t)_{t\in\mathbb{N}}$ is ψ-compatible. Denote by $\overline{s}$ and $\underline{s}$ the functions defined in (3.15), and let $\tilde{\overline{s}}$ and $\tilde{\underline{s}}$ be the corresponding functions defined for $\tilde{z}_t$. As a consequence of Proposition 3.11, we have

$$\tilde{\overline{s}}(B(x, \varepsilon - 2\delta)) \geq \frac{\tilde{\underline{s}}(B(x_1, \varepsilon - 3\delta))}{2} + \frac{\tilde{\underline{s}}(B(x_2, \varepsilon - 3\delta))}{2}. \tag{4.2}$$

Using (4.1), we obtain that for t large enough,

$$\mathbb{P}\left(\frac{1}{t} z_t \in B(x, \varepsilon)\right) \geq \mathbb{P}\left(\frac{1}{t} \widetilde{z}_t \in B(x, \varepsilon - 2\delta)\right),$$

$$\mathbb{P}\left(\frac{1}{t} \widetilde{z}_t \in B(x_i, \varepsilon - 3\delta)\right) \geq \mathbb{P}\left(\frac{1}{t} z_t \in B(x_i, \varepsilon - 5\delta)\right).$$

By combining this with (4.2), we obtain that

$$s(B(x, \varepsilon)) \geq \frac{\underline{s}(B(x_1, \varepsilon - 5\delta))}{2} + \frac{\underline{s}(B(x_2, \varepsilon - 5\delta))}{2}.$$

Hence, $(z_t)_{t\in\mathbb{N}}$ is admissible. Since $\|\frac{1}{t} z_t\| \leq \|f\|$ for all t, the laws of $(\frac{1}{t} z_t)_{t\in\mathbb{N}}$ are exponentially tight. Thus, by Proposition 3.9, the LDP holds with good convex rate function I_f (which is equal to $-s$ in the notation of Definition 3.7). We now denote by I_f^* the Fenchel–Legendre transform of I_f, and by q_f^* the Fenchel–Legendre transform of q_f.

In order to identify I_f and q_f^* (see [9, Theorem 4.5.10] for a similar argument), we use Varadhan's integral theorem and the convexity of I_f. For $\alpha, u \in \mathbb{R}^d$, let $\phi_\alpha(u) = (\alpha, u)$ (which is continuous as a function of u for fixed α), and let P_t be the distribution of $\frac{1}{t} z_t$. We have

$$q_f(\alpha) = \lim_{t\to\infty} \frac{1}{t} \log \int_{\mathbb{R}^d} \mathrm{e}^{t\phi_\alpha(u)} \mathrm{d}P_t(u).$$

Since for any $\gamma > 1$ we have $q_f(\gamma\alpha) < \infty$ by Proposition 4.2, the conditions of Varadhan's theorem (see [9, Theorem 4.3.1] or [10, Theorem 2.1.10]) are met, and we obtain $q_f(\alpha) = I_f^*(\alpha)$. Since this is true for all $\alpha \in \mathbb{R}^d$ and since I_f is convex and continuous (in particular, lower semicontinuous), we find $I_f = q_f^*$, which completes the proof. □

5 LDP for entropy production

In this section we assume **(SSD)**, so that $\psi_{n,t}$ is well defined and Parts 1 and 2 of Proposition 3.1 hold. In particular, (3.7) shows that $(\frac{1}{t}\sigma_t)_{t\in\mathbb{N}}$ is ψ-compatible and hence admissible by Proposition 3.11.

Proposition 5.1. *The limit*

$$q(\alpha) := \lim_{t\to\infty} \frac{1}{t} \log \langle \mathrm{e}^{\alpha\sigma_t}, \mathbb{P}\rangle$$

exists for all $\alpha \in \mathbb{R}$ and takes value in $(-\infty, \infty]$. The function q is lower semicontinuous and convex. We have $q(0) = 0$ and $q(-1) \le 0$, so that q is non-positive (and hence finite) on $[-1, 0]$.

Proof. Since $(\frac{1}{t}\sigma_t)_{t\in\mathbb{N}}$ is admissible, we find by Lemma 3.12 (with $d = 1$) that q exists, takes value in $(-\infty, \infty]$, and is lower semicontinuous and convex. We have obviously $q(0) = 0$. Moreover,

$$\frac{1}{t} \log \left\langle \mathrm{e}^{-\sigma_t}, \mathbb{P}\right\rangle = \frac{1}{t} \log \sum_{w\in\Omega_t^+} \frac{\widehat{\mathbb{P}}_t(w)}{\mathbb{P}_t(w)} \mathbb{P}_t(w) = \frac{1}{t} \log \widehat{\mathbb{P}}_t(\Omega_t^+) \le 0,$$

and so $q(-1) \le 0$. By convexity, q is non-positive (and hence finite) on $[-1, 0]$. This completes the proof. □

In the sequel, we denote by I^* and q^* the Fenchel–Legendre transforms of I and q.

Proposition 5.2. *The sequence $(\frac{1}{t}\sigma_t)_{t\in\mathbb{N}}$ satisfies the LDP (see (2.9) and (2.10)) with a convex rate function I given by $I(s) = q^*(s)$ for all $s \in \mathbb{R}$. Moreover, if $q(\alpha) < \infty$ for all α in a neighborhood of 0, then I is a good rate function.*

Proof. Since $(\frac{1}{t}\sigma_t)_{t\in\mathbb{N}}$ is admissible, it satisfies by Proposition 3.9 the weak LDP with convex rate function I. To strengthen the result to the LDP (*i.e.*, to show that (2.10) is true also for unbounded Γ), we separate the following two cases (recall that q is finite and non-positive on $[-1, 0]$, and that $q(-1) \le 0 = q(0)$).

- If $q(\alpha) < \infty$ in a neighborhood of the origin, a standard application of Chebychev's inequality shows that the laws of $(\frac{1}{t}\sigma_t)_{t\in\mathbb{N}}$ are exponentially tight, so that the weak LDP is in fact the LDP, and I is a good rate function.
- If $q(\alpha) = \infty$ for all $\alpha > 0$, then we have $\lim_{x\to+\infty} q^*(x) = 0$. The identification $I = q^*$, which we prove below, implies that $\lim_{x\to+\infty} I(x) = 0$ (in particular I is not a good rate function). We now show that the LDP still holds. If Γ is a closed set such that $\sup\Gamma = +\infty$, then $\inf_{x\in\Gamma} I(x) = 0$, and hence (2.10) is trivial. Assume on the contrary that Γ is a closed set such that $\sup\Gamma < \infty$ (but possibly $\inf\Gamma = -\infty$). Then, since $q(-1) < \infty$, Chebychev's inequality provides the necessary exponential tightness on the negative half-line in order to show (2.10) (by a minor and standard adaptation of the argument in the proof of Proposition 3.9).

We now turn to the comparison of I and q^*. If $q(\alpha) < \infty$ for all $\alpha \in \mathbb{R}$, we can proceed exactly as in Proposition 4.3, by using Varadhan's theorem to obtain that $q = I^*$, and then the convexity of I to obtain that $I = q^*$. However, in the general case, more specific estimates are required in order to show that $q = I^*$. We split the proof of this identity into three steps. Steps 1 and 3 are almost identical to the proof of Varadhan's theorem (see [9, Theorem 4.3.1] or [10, Theorem 2.1.10]), although our assumptions are slightly different. Step 2, however, is quite specific to our setup (see Remark 5.3 below).

Step 1: $q \geq I^*$. We denote by P_t the law of $t^{-1}\sigma_t$. For any $x, \alpha \in \mathbb{R}$ and $\varepsilon > 0$, we have

$$\begin{aligned} q(\alpha) &\geq \liminf_{t\to\infty} \frac{1}{t} \log \int_{|x-y|<\varepsilon} \mathrm{e}^{t\alpha y} P_t(dy) \\ &\geq (\alpha x - |\alpha|\varepsilon) + \liminf_{t\to\infty} \frac{1}{t} \log P_t((x-\varepsilon, x+\varepsilon)) \\ &\geq (\alpha x - |\alpha|\varepsilon) - \inf_{|x-y|<\varepsilon} I(y) \geq \alpha x - |\alpha|\varepsilon - I(x). \end{aligned}$$

Letting $\varepsilon \to 0$, we get $q(\alpha) \geq \alpha x - I(x)$. Since α and x are arbitrary, we obtain $q \geq I^*$.

Step 2: Tail estimates. Let

$$\alpha_\pm = \lim_{x\to\pm\infty} \frac{I(x)}{x}.$$

By convexity these limits exist, and since I is non-negative we have $\pm\alpha_\pm \in [0, \infty]$. Moreover, since $0 \geq q(-1) \geq I^*(-1) = \sup_{x\in\mathbb{R}}(-x - I(x))$, we actually have $\alpha_- \in [-\infty, -1]$, and in particular $\alpha_- < \alpha_+$. Let $\alpha \in (\alpha_-, \alpha_+)$, and set $\delta = \frac{1}{2}\min(1, |\alpha - \alpha_-|, |\alpha - \alpha_+|)$. Then there exists $c > 0$ such that

$$I(x) \geq \alpha x + \delta|x| - c \quad \text{for all } x \in \mathbb{R}.$$

Using this and (3.21) we find

$$\begin{aligned} \frac{1}{t} \log P_t\big((k-1, k+1)\big) &\leq \gamma_t - \inf_{y\in B(k, 1+(1+|k|)\gamma_t)} I(y) \\ &\leq -\alpha k - \delta|k| + c' + c''|k|\gamma_t, \end{aligned}$$

where $\gamma_t \to 0$, and the constants c', c'' are independent of t and k. It follows that, for all t large enough,

$$P_t\big((k-1, k+1)\big) \leq \exp\Big(\big(-\alpha k - \tfrac{\delta}{2}|k| + c'\big)t\Big),$$

whence there exists $C > 0$, depending only on α, such that for all $K > 0$,

$$R_K := \limsup_{t\to\infty} \frac{1}{t} \log \int_{|x|>K} \mathrm{e}^{\alpha t x} P_t(dx) \leq -K\frac{\delta}{2} + C. \tag{5.1}$$

Step 3: $q \leq I^*$. If $\alpha \notin [\alpha_-, \alpha_+]$, we clearly have $I^*(\alpha) = +\infty \geq q(\alpha)$. It therefore remains to show that $q(\alpha) \leq I^*(\alpha)$ for all $\alpha \in [\alpha_-, \alpha_+]$. Since both I^* and q are convex, lower semicontinuous functions, it is enough to consider $\alpha \in (\alpha_-, \alpha_+)$. We now fix $\alpha \in (\alpha_-, \alpha_+)$, $\varepsilon > 0$ and $K > 0$. For all $x \in [-K, K]$, there exists an open neighborhood G_x such that

$$\inf_{y\in\overline{G}_x} I(y) \geq (I(x) - \varepsilon) \wedge \varepsilon^{-1}, \qquad \sup_{y\in G_x} \alpha y \leq \alpha x + \varepsilon.$$

We extract a finite subcover $\{G_{x_1}, \ldots, G_{x_n}\}$ of $[-K, K]$ and write

$$\begin{aligned} \limsup_{t\to\infty} \frac{1}{t} \log \int_{G_{x_i}} \mathrm{e}^{\alpha t y} P_t(dy) &\leq \limsup_{t\to\infty} \frac{1}{t} \log \left(\mathrm{e}^{\alpha t x_i + \varepsilon t} P_t(G_{x_i})\right) \\ &\leq \alpha x_i + \varepsilon - (I(x_i) - \varepsilon) \wedge \varepsilon^{-1} \\ &= \max\left(\alpha x_i - I(x_i) + 2\varepsilon, \alpha x_i + \varepsilon - \varepsilon^{-1}\right) \\ &\leq \max\left(I^*(\alpha) + 2\varepsilon, |\alpha|K + \varepsilon - \varepsilon^{-1}\right). \end{aligned}$$

It follows that

$$\begin{aligned} q(\alpha) &\leq \limsup_{t\to\infty} \frac{1}{t} \log \left(\int_{|x|>K} \mathrm{e}^{\alpha t x} P_t(dx) + \sum_{i=1}^{n} \int_{G_{x_i}} \mathrm{e}^{\alpha t x} P_t(dx) \right) \\ &\leq \max\left\{ R_K, I^*(\alpha) + 2\varepsilon, |\alpha| K + \varepsilon - \varepsilon^{-1} \right\}. \end{aligned}$$

Sending $\varepsilon \to 0$ shows that $q(\alpha) \leq \max(R_K, I^*(\alpha))$. Finally, sending $K \to \infty$ and using (5.1) yields $q(\alpha) \leq I^*(\alpha)$, which completes the proof. □

Remark 5.3. The tail estimates in Step 2 above are equivalent to the statement that $q(\alpha) < \infty$ for all $\alpha \in (\alpha_-, \alpha_+)$, which is obviously a necessary condition in order to have $q = I^*$. The uniform bound (3.21) is crucial in Step 2. An instructive example of what can go wrong without it is given by the family of distributions

$$\frac{dP_t}{dx} = (1 - \mathrm{e}^{-t^2}) \sqrt{t/\pi}\, \mathrm{e}^{-tx^2} + \frac{1}{2} \mathrm{e}^{-t^2} (\delta_t(x) + \delta_{-t}(x)), \qquad t \in \mathbb{N},$$

which satisfies the LDP with rate function $I(x) = x^2$. Here $\alpha_\pm = \pm\infty$, while[34]

$$q(\alpha) = \lim_{t\to\infty} \frac{1}{t} \log \int \mathrm{e}^{t\alpha x} P_t(dx) = \alpha^2/4 + \infty \mathbb{1}_{|\alpha|>1}.$$

One now easily checks that q and I^* coincide only on $[-1, 1]$, and that I and q^* coincide only on $[-1/2, 1/2]$.

Lemma 5.4. *If $\widehat{\mathbb{P}} = \Theta\mathbb{P}$ with Θ as in Definition 2.5, then q satisfies*

$$q(-\alpha) = q(\alpha - 1), \qquad \alpha \in \mathbb{R}, \tag{5.2}$$

and I satisfies the Gallavotti–Cohen symmetry

$$I(-s) = I(s) + s, \qquad s \in \mathbb{R}. \tag{5.3}$$

Proof. Recalling that $\theta_t = \theta_t^{-1}$ and that θ_t leaves Ω_t^+ invariant (see Remark 2.6), we find

$$\begin{aligned} \left\langle \mathrm{e}^{-\alpha\sigma_t}, \mathbb{P} \right\rangle &= \sum_{w\in\Omega_t^+} \mathbb{P}_t^{1-\alpha}(w) \widehat{\mathbb{P}}_t^{\alpha}(w) = \sum_{w\in\Omega_t^+} \mathbb{P}_t^{1-\alpha}(\theta_t(w)) \widehat{\mathbb{P}}_t^{\alpha}(\theta_t(w)) \\ &= \sum_{w\in\Omega_t^+} \widehat{\mathbb{P}}_t^{1-\alpha}(w) \mathbb{P}_t^{\alpha}(w) = \left\langle \mathrm{e}^{(\alpha-1)\sigma_t}, \mathbb{P} \right\rangle, \end{aligned}$$

which yields (5.2). Although one can derive (5.3) from (5.2) and the identity $I = q^*$, we provide here a direct derivation based on the LDP and the following *transient fluctuation relation* (see [8] and references therein): using that $\sigma_t \circ \theta_t = -\sigma_t$, we find

$$\begin{aligned} \mathbb{P}_t \left(\frac{1}{t} \sigma_t = s \right) &= \sum_{w\in\Omega_t^+ : \sigma_t(w) = ts} \mathbb{P}_t(w) = \sum_{w\in\Omega_t^+ : \sigma_t(w) = ts} \mathrm{e}^{ts} \widehat{\mathbb{P}}_t(w) \\ &= \sum_{w\in\Omega_t^+ : \sigma_t(\theta_t(w)) = ts} \mathrm{e}^{ts} \widehat{\mathbb{P}}_t(\theta_t(w)) = \sum_{w\in\Omega_t^+ : \sigma_t(w) = -ts} \mathrm{e}^{ts} \mathbb{P}_t(w) \\ &= \mathrm{e}^{ts} \mathbb{P}_t \left(\frac{1}{t} \sigma_t = -s \right). \end{aligned}$$

[34]The quantity $\mathbb{1}_{|\alpha|>1}$ is equal to 1 if $|\alpha| > 1$ and to 0 otherwise.

From this we obtain that, for all $\varepsilon > 0$ and $s \in \mathbb{R}$,

$$\left|\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{P}_t\left(\frac{1}{t}\sigma_t \in B(s,\varepsilon)\right) - \liminf_{t\to\infty}\frac{1}{t}\log\mathbb{P}_t\left(\frac{1}{t}\sigma_t \in B(-s,\varepsilon)\right) - s\right| \leq \varepsilon.$$

By the construction of the rate function I, sending $\varepsilon \to 0$ gives $|-I(s) + I(-s) - s| = 0$, which is (5.3). □

6 Level-3 LDP

6.1 Main result

In this section, we assume **(SLD)** and prove Theorem 2.13. For technical reasons (in fact, in order to invert a Fenchel–Legendre transform in the proof of Proposition 6.1 below), we consider a slightly more general situation, viewing $\mu_t(\omega) := \frac{1}{t}\sum_{s=0}^{t-1}\delta_{\varphi^s(\omega)}$ as an element of the space $X = \mathcal{M}(\Omega)$ of signed Borel measures on Ω. We endow $\mathcal{M}(\Omega)$ with the weak-$\star$ topology with respect to the natural pairing[35]

$$\langle f, \nu\rangle = \int f\mathrm{d}\nu, \qquad \nu \in \mathcal{M}(\Omega), \quad f \in C(\Omega).$$

Recall that $C(\Omega)$ is endowed with the topology of uniform convergence. With these topologies, the spaces $\mathcal{M}(\Omega)$ and $C(\Omega)$ are the continuous dual of each other (with the natural identification). The induced topology on $\mathcal{P}(\Omega)$ is the weak topology that we have considered so far.

We shall show that for every open set $O \subset \mathcal{M}(\Omega)$ and every closed set $\Gamma \subset \mathcal{M}(\Omega)$,

$$\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(\mu_t \in O\right) \geq -\inf_{\nu\in O}\mathbb{I}(\nu), \tag{6.1}$$

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(\mu_t \in \Gamma\right) \leq -\inf_{\nu\in \Gamma}\mathbb{I}(\nu), \tag{6.2}$$

where $\mathbb{I}$ is given by (2.17) on $\mathcal{P}(\Omega)$, and where $\mathbb{I}(\nu) = +\infty$ on $\mathcal{M}(\Omega)\setminus\mathcal{P}(\Omega)$. Since $\mu_t(\omega) \in \mathcal{P}(\Omega)$ for all ω, this will immediately imply the LDP on $\mathcal{P}(\Omega)$ in Theorem 2.13.

For $f \in C(\Omega)$, let

$$Q(f) = \lim_{t\to\infty}\frac{1}{t}\log\left\langle \mathrm{e}^{S_t f}, \mathbb{P}\right\rangle. \tag{6.3}$$

By Proposition 4.2 (in the special case $d = 1, \alpha = 1$), the limit (6.3) exists and is finite, and the function Q is convex and 1-Lipschitz .

Proposition 6.1. *The sequence $(\mu_t)_{t\in\mathbb{N}}$ satisfies the LDP with respect to the weak-$\star$ topology on $\mathcal{M}(\Omega)$ for some good rate function $\mathbb{I}$ (see (6.1) and (6.2)). Moreover, $\mathbb{I}$ is the Fenchel–Legendre transform of Q, i.e., for all $\nu \in \mathcal{M}(\Omega)$,*

$$\mathbb{I}(\nu) = \sup_{f\in C(\Omega,\mathbb{R})}\big(\langle f,\nu\rangle - Q(f)\big). \tag{6.4}$$

Finally, $\mathbb{I}(\nu) = +\infty$ for all $\nu \in \mathcal{M}(\Omega)\setminus\mathcal{P}_\varphi(\Omega)$.

Proof. We set $z_t = t\mu_t$, and we define $\overline{s}$ and $\underline{s}$ as in (3.15). We first show that the sequence $(z_t)_{t\in\mathbb{N}}$ is admissible. A neighborhood basis of $\nu \in \mathcal{M}(\Omega)$ is given by

$$\mathcal{N}_\nu = \left\{G(\nu, f,\varepsilon) := \{\mu \in \mathcal{M}(\Omega) : |\langle f, \mu - \nu\rangle| < \varepsilon\} : \varepsilon > 0, f \in C(\Omega,\mathbb{R}^d), d \geq 1\right\}.$$

[35]We shall reserve the symbols $\mathbb{P}, \mathbb{Q}$ for the elements of $\mathcal{P}(\Omega)$ and denote by μ, ν the elements of $\mathcal{M}(\Omega)$.

We immediately have

$$\frac{1}{t} z_t(\omega) = \mu_t(\omega) \in G(\nu, f, \varepsilon) \iff \frac{1}{t} S_t f(\omega) \in B(\langle f, \nu \rangle, \varepsilon).$$

Fix now $\nu = \frac{1}{2}\nu_1 + \frac{1}{2}\nu_2 \in \mathcal{M}(\Omega)$, and consider a neighborhood $G(\nu, f, \varepsilon)$ of ν. Let $x = \frac{1}{2}x_1 + \frac{1}{2}x_2$ with $x_i = \langle f, \nu_i \rangle$. Since $(S_t f)_{t \in \mathbb{N}}$ is admissible by Proposition 4.3, there exists $\varepsilon' > 0$ such that

$$\begin{aligned}
\underline{s}(G(\nu, f, \varepsilon)) &= \liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{1}{t} S_t f \in B(x, \varepsilon)\right) \\
&\geq \limsup_{t\to\infty} \frac{1}{2t} \log \mathbb{P}\left(\frac{1}{t} S_t f \in B(x_1, \varepsilon')\right) + \limsup_{t\to\infty} \frac{1}{2t} \log \mathbb{P}\left(\frac{1}{t} S_t f \in B(x_2, \varepsilon')\right) \\
&= \frac{1}{2}\overline{s}(G(\nu_1, f, \varepsilon')) + \frac{1}{2}\overline{s}(G(\nu_2, f, \varepsilon')).
\end{aligned}$$

This implies that $(z_t)_{t\in\mathbb{N}}$ is admissible. Moreover, since μ_t belongs to the compact subset $\mathcal{P}(\Omega)$ for all t, the laws of $(\frac{1}{t} z_t)_{t\in\mathbb{N}}$ trivially form an exponentially tight family, so that by Proposition 3.9, μ_t satisfies the LDP with good convex rate function $\mathbb{I}$ defined by

$$\mathbb{I}(\nu) = -\inf_{G\in\mathcal{N}_\nu} \underline{s}(G) = -\inf_{G\in\mathcal{N}_\nu} \overline{s}(G).$$

We now show that $\mathbb{I}(\nu) = +\infty$ when $\nu \notin \mathcal{P}_\varphi(\Omega)$. Since $\mu_t \in \mathcal{P}(\Omega)$, and since $\mathcal{P}(\Omega)$ is closed, one immediately obtains $\mathbb{I}(\nu) = +\infty$ if $\nu \notin \mathcal{P}(\Omega)$. Now, if $\nu \in \mathcal{P}(\Omega) \setminus \mathcal{P}_\varphi(\Omega)$, one can find a function $g \in C(\Omega)$ such that $f := g - g \circ \varphi$ satisfies $\langle f, \nu \rangle = 1$. Then, for all $\mu \in G(\nu, f, 1/2)$, we have $\langle f, \mu \rangle > 1/2$. However, by construction,

$$\langle f, \mu_t(\omega) \rangle = \frac{1}{t}(g(\omega) - g \circ \varphi^t(\omega)) \leq \frac{2}{t}\|g\|,$$

which is eventually $< 1/2$. Thus, $\mathbb{P}(\mu_t \in G(\nu, f, 1/2)) = 0$ for t large enough, and

$$\mathbb{I}(\nu) \geq -\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}(\mu_t \in G(\nu, f, 1/2)) = +\infty.$$

Following the same ideas as in Proposition 4.3 (see also [9, Theorem 4.5.10] and [34]), we now identify $\mathbb{I}$ and Q^* using Varadhan's integral theorem and the convexity of $\mathbb{I}$. For fixed $f \in C(\Omega)$, let $\phi_f = \langle f, \cdot \rangle$, which is a continuous function on $\mathcal{M}(\Omega)$. Denoting by P_t be the law of μ_t, we have

$$Q(f) = \lim_{t\to\infty} \frac{1}{t} \log \int_{\mathcal{M}(\Omega)} \mathrm{e}^{t\phi_f(\nu)} \mathrm{d}P_t(\nu).$$

Since for any $\gamma > 1$ we have $Q(\gamma f) < \infty$ (or more simply, using that P_t is supported on the compact set $\mathcal{P}(\Omega)$), we can apply Varadhan's theorem, and obtain[36]

$$Q(f) = \mathbb{I}^*(f) = \sup_{\nu\in\mathcal{M}(\Omega)} (\langle f, \nu \rangle - \mathbb{I}(\nu)).$$

Since this is true for all $f \in C(\Omega)$, and since $\mathbb{I}$ is convex and lower semicontinuous, we find that $\mathbb{I} = Q^*$, which is (6.4) (see [4, Theorem 3.10] or [9, Lemma 4.5.8] for variants of the duality principle between convex conjugate functions that apply in the present setup). $\square$

[36] Recall that $C(\Omega)$ is the dual of $\mathcal{M}(\Omega)$ with the weak-$\star$ topology, so $\mathbb{I}^*$ is naturally defined on $C(\Omega)$.

6.2 Alternative expression for the rate function

Assuming also **(UD)**, we now derive an alternative expression for the rate function $\mathbb{I}$ of Proposition 6.1. This new expression will imply, in particular, that $\mathbb{I}$ is affine on $\mathcal{P}_\varphi(\Omega)$.

Given $\mathbb{Q} \in \mathcal{P}(\Omega)$ and $t \in \mathbb{N}$, consider the relative entropy (recall (2.15))

$$\operatorname{Ent}(\mathbb{Q}_t|\mathbb{P}_t) = \varsigma_t(\mathbb{Q}) - h_t(\mathbb{Q}),$$

where we set, with the usual convention that $0 \log 0 = 0$,

$$\varsigma_t(\mathbb{Q}) = \begin{cases} -\sum_{w\in\Omega_t} \mathbb{Q}_t(w)\log \mathbb{P}_t(w) & \text{if } \mathbb{Q}_t \ll \mathbb{P}_t, \\ +\infty & \text{otherwise,} \end{cases}$$
$$h_t(\mathbb{Q}) = -\sum_{w\in\Omega_t} \mathbb{Q}_t(w)\log \mathbb{Q}_t(w).$$

For $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$, we have

$$\lim_{t\to\infty} \frac{1}{t} h_t(\mathbb{Q}) = h(\mathbb{Q}),$$

where $h(\mathbb{Q})$ is the Kolmogorov–Sinai entropy of $\mathbb{Q}$ with respect to φ. The limit exists, is finite, and the mapping $h : \mathcal{P}_\varphi(\Omega) \to [0,\infty)$ is upper semicontinuous and affine.[37] For completeness, a proof of these elementary properties of the Kolmogorov–Sinai entropy is provided in Lemma A.4.

We first need a technical lemma.

Lemma 6.2. *Assume* **(UD)**. *Let f, g be two non-negative random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that f is $\mathcal{F}_n$-measurable and g is $\mathcal{F}_r$-measurable with $n, r \in \mathbb{N}$. Then*

$$\left\langle f\,(g\circ\varphi^{n+\tau_n}), \mathbb{P}\right\rangle \le \mathrm{e}^{c_n+\tau_n\log|\mathcal{A}|}\left\langle f,\mathbb{P}\right\rangle \left\langle g,\mathbb{P}\right\rangle.$$

Proof. Recalling the conventions in the beginning of Section 3, we obtain by **(UD)**

$$\begin{aligned}\left\langle f\,(g\circ\varphi^{n+\tau_n}), \mathbb{P}\right\rangle &= \sum_{u\xi v\in\Omega^+_{n+\tau_n+r}} f(u)g(v)\mathbb{P}(u\xi v)\\ &\le \sum_{u\xi v\in\Omega^+_{n+\tau_n+r}} f(u)g(v)\mathrm{e}^{c_n}\mathbb{P}(u)\mathbb{P}(v)\\ &\le \sum_{\xi\in\Omega_{\tau_n}} \mathrm{e}^{c_n}\left\langle f,\mathbb{P}\right\rangle \left\langle g,\mathbb{P}\right\rangle,\end{aligned}$$

which implies the claim, since $|\Omega_{\tau_n}| = |\mathcal{A}|^{\tau_n}$. (The factor involving $\mathcal{A}$ would not be needed with the alternative **(UD)** assumption mentioned in Remark 2.2.) □

Proposition 6.3. *Suppose that* **(UD)** *is satisfied.*[38] *Then the limit*

$$\varsigma(\mathbb{Q}) := \lim_{t\to\infty} \frac{1}{t}\varsigma_t(\mathbb{Q}) \tag{6.5}$$

[37] See Corollary 4.3.14, Corollary 4.3.17 and the remark following it in the book [21].

[38] Although **(SLD)** is a standing assumption in this section, observe that this proposition does not rely on it provided that we *define* $\mathbb{I}$ by (6.4). This proposition does not rely on the validity of the LDP for μ_t either.

exists for any measure $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$, and the mapping $\varsigma : \mathcal{P}_\varphi(\Omega) \to [0,\infty]$ is lower semicontinuous. Moreover, for any $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$, we have

$$\mathbb{I}(\mathbb{Q}) = \varsigma(\mathbb{Q}) - h(\mathbb{Q}) = \lim_{t\to\infty} \frac{1}{t}\mathrm{Ent}(\mathbb{Q}_t|\,\mathbb{P}_t), \tag{6.6}$$

and $\mathbb{I}$ is an affine function of $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$.

Proof. We first prove the existence and lower semicontinuity of the limit (6.5) by using a classical subadditivity argument. For each pair of integers t, n with $t \geq n + \tau_n$, we let $M = \lfloor t/(n+\tau_n)\rfloor$. By using **(UD)** $M-1$ times, we find for all $w \in \Omega_t^+$

$$\log \mathbb{P}_t(w) \leq \log \mathbb{P}_{(n+\tau_n)M}(w_{[1,(n+\tau_n)M]}) \leq \sum_{k=0}^{M-1} \log \mathbb{P}_n(w_{[k(n+\tau_n)+1,k(n+\tau_n)+n]}) + (M-1)c_n,$$

where both sides may be $-\infty$. Integrating this inequality with respect to $-t^{-1}\mathbb{Q}$ and using the translation invariance of $\mathbb{Q}$ yields

$$\frac{1}{t}\varsigma_t(\mathbb{Q}) \geq \frac{M}{t}\varsigma_n(\mathbb{Q}) - \frac{(M-1)c_n}{t} \geq \left(\frac{1}{n+\tau_n} - \frac{1}{t}\right)\varsigma_n(\mathbb{Q}) - \frac{c_n}{n+\tau_n},$$

where we have also used that $\varsigma_n(\mathbb{Q}) \geq 0$. Sending now $t \to \infty$ yields

$$\liminf_{t\to\infty} \frac{1}{t}\varsigma_t(\mathbb{Q}) \geq \frac{1}{1+\tau_n/n}\left(\frac{1}{n}\varsigma_n(\mathbb{Q}) - \frac{c_n}{n}\right).$$

Taking the $\limsup$ as $n \to \infty$ shows that the limit $\varsigma(\mathbb{Q})$ exists (it can be infinite). We then find that

$$\varsigma(\mathbb{Q}) = \sup_{n\in\mathbb{N}} \frac{1}{1+\tau_n/n}\left(\frac{1}{n}\varsigma_n(\mathbb{Q}) - \frac{c_n}{n}\right),$$

and thus, since $\mathbb{Q} \mapsto \varsigma_n(\mathbb{Q})$ is continuous for all n, we obtain that ς is lower semicontinuous.

We now fix $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$ and establish (6.6). The second equality in (6.6) follows from the definitions, and we need to show that $\mathbb{I}(\mathbb{Q}) = \mathbb{L}(\mathbb{Q}) := \lim_{t\to\infty} \frac{1}{t}\mathrm{Ent}(\mathbb{Q}_t|\,\mathbb{P}_t)$.

We first deal with the special case where there exists $t_0 \in \mathbb{N}$ such that $\mathbb{Q}_{t_0}$ is not absolutely continuous with respect to $\mathbb{P}_{t_0}$. Then $\mathrm{Ent}(\mathbb{Q}_t|\,\mathbb{P}_t) = \infty$ for all $t \geq t_0$, and thus $\mathbb{L}(\mathbb{Q}) = \infty$. Let us choose $w \in \Omega_{t_0}$ such that $\mathbb{Q}(w) > 0 = \mathbb{P}(w)$. For all $n \in \mathbb{N}$, let $f_n(\omega) = n\mathbb{1}_{\omega_{[1,t_0]}=w}$. Observe that $\langle f_n, \mathbb{Q}\rangle = n\mathbb{Q}(w)$ and that $Q(f_n) = 0$, since $S_t f_n$ vanishes on the support of $\mathbb{P}$. Thus, using (6.4), we see that

$$\mathbb{I}(\mathbb{Q}) \geq \langle f_n, \mathbb{Q}\rangle - Q(f_n) = n\mathbb{Q}(w) \quad \text{for all } n \in \mathbb{N},$$

so that $\mathbb{I}(\mathbb{Q}) = \infty$.

Suppose now that $\mathbb{Q}_t \ll \mathbb{P}_t$ for all $t \in \mathbb{N}$. We shall first prove that $\mathbb{L}(\mathbb{Q}) \geq \mathbb{I}(\mathbb{Q})$. Let $f \in C_{\mathrm{fin}}(\Omega)$ be $\mathcal{F}_r$-measurable for some $r \in \mathbb{N}$, and let

$$A_t = \left\langle \mathrm{e}^{S_t f}, \mathbb{P}\right\rangle.$$

By Jensen's inequality and the invariance of $\mathbb{Q}$, we have

$$\begin{aligned}
\log A_t = \log\left\langle \mathrm{e}^{S_t f}, \mathbb{P}\right\rangle &= \log \int_{\Sigma_{t+r-1}} \mathrm{e}^{S_t f}\, \frac{\mathrm{d}\mathbb{P}_{t+r-1}}{\mathrm{d}\mathbb{Q}_{t+r-1}}\,\mathrm{d}\mathbb{Q}_{t+r-1} \\
&\geq \int_{\Sigma_{t+r-1}} \left(S_t f - \log\frac{\mathrm{d}\mathbb{Q}_{t+r-1}}{\mathrm{d}\mathbb{P}_{t+r-1}}\right)\mathrm{d}\mathbb{Q}_{t+r-1} \\
&= t\langle f, \mathbb{Q}\rangle - \mathrm{Ent}(\mathbb{Q}_{t+r-1}|\mathbb{P}_{t+r-1}),
\end{aligned}$$

where Σ_t is the support of $\mathbb{Q}_t$. Dividing by t and sending $t \to \infty$ shows that $\mathbb{L}(\mathbb{Q}) \geq \langle f, \mathbb{Q}\rangle - Q(f)$. Since $f \in C_{\text{fin}}(\Omega)$ is arbitrary, $C_{\text{fin}}(\Omega)$ is dense in $C(\Omega)$, and Q is Lipschitz, we find

$$\mathbb{L}(\mathbb{Q}) \geq \sup_{f \in C_{\text{fin}}(\Omega)} (\langle f, \mathbb{Q}\rangle - Q(f)) = \sup_{f \in C(\Omega)} (\langle f, \mathbb{Q}\rangle - Q(f)) = \mathbb{I}(\mathbb{Q}).$$

It remains to prove that $\mathbb{L}(\mathbb{Q}) \leq \mathbb{I}(\mathbb{Q})$. Fix two integers $n, M \geq 1$ and let $t = n'M$ where $n' = \tau_n + n$. Consider the $\mathcal{F}_n$-measurable function $f = \frac{1}{n'} \log \frac{\mathbb{Q}_n}{\mathbb{P}_n}$. This function is well defined on the support of $\mathbb{Q}_n$ (and, hence, on the support of $\mathbb{P}_n$), and we define it by $-\infty$ on the complement. Note that

$$\operatorname{Ent}(\mathbb{Q}_n | \mathbb{P}_n) = n'\langle f, \mathbb{Q}\rangle. \tag{6.7}$$

We have (see Figure 1) the decomposition

$$S_t f = \sum_{s=0}^{n'-1} f_s^{(M)}, \quad f_s^{(M)}(\omega) = \sum_{k=0}^{M-1} f(\omega_{[kn'+s+1, kn'+s+n]}). \tag{6.8}$$

Using Hölder's inequality and translation invariance leads to

$$\langle \mathrm{e}^{S_t f}, \mathbb{P}\rangle \leq \prod_{s=0}^{n'-1} \langle \mathrm{e}^{n' f_s^{(M)}}, \mathbb{P}\rangle^{1/n'} = \langle \mathrm{e}^{n' f_0^{(M)}}, \mathbb{P}\rangle.$$

Using then Lemma 6.2 recursively $M-1$ times and translation invariance, we obtain

$$\langle \mathrm{e}^{S_t f}, \mathbb{P}\rangle \leq \mathrm{e}^{(M-1)d_n} \langle \mathrm{e}^{n' f}, \mathbb{P}\rangle^M = \mathrm{e}^{(M-1)d_n} \left\langle \frac{\mathbb{Q}_n}{\mathbb{P}_n}, \mathbb{P}\right\rangle^M = \mathrm{e}^{(M-1)d_n} \left(\mathbb{Q}_n(\Omega_n^+)\right)^M \leq \mathrm{e}^{(M-1)d_n},$$

where $d_n = c_n + \tau_n \log |\mathcal{A}| = o(n)$. Thus,

$$\frac{1}{t} \log A_t \leq \frac{(M-1)d_n}{t} \leq \frac{d_n}{n'},$$

whence $Q(f) \leq \frac{d_n}{n'}$. Combining this with (6.7), we derive

$$\frac{1}{n} \operatorname{Ent}(\mathbb{Q}_n | \mathbb{P}_n) = \frac{n'}{n} \langle f, \mathbb{Q}\rangle \leq \frac{n'}{n} (\langle f, \mathbb{Q}\rangle - Q(f)) + \frac{d_n}{n} \leq \frac{n'}{n} \mathbb{I}(\mathbb{Q}) + \frac{d_n}{n}.$$

Passing to the limit as $n \to \infty$ shows that $\mathbb{L}(\mathbb{Q}) \leq \mathbb{I}(\mathbb{Q})$, and (6.6) follows. Finally, since both $\mathbb{Q} \mapsto h(\mathbb{Q})$ and $\mathbb{Q} \mapsto \varsigma(\mathbb{Q})$ are affine, we obtain from (6.6) that so is $\mathbb{I}$. □

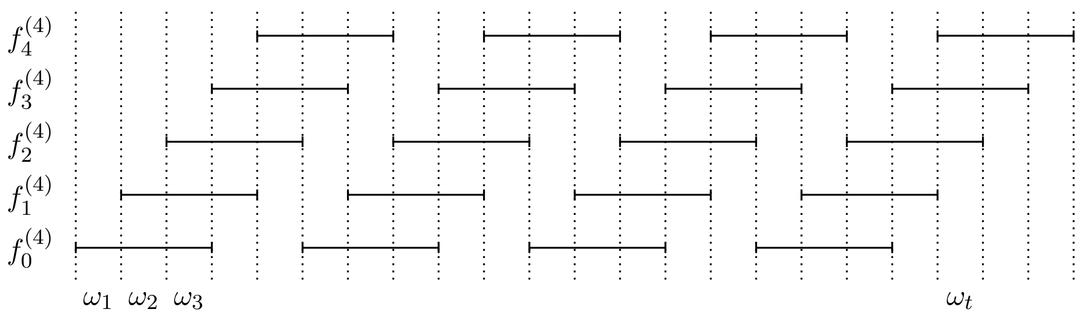


Figure 1: Illustration of (6.8) in the case $n = 3$, $\tau_n = 2$, $M = 4$

6.3 Level-3 fluctuation relation

Proposition 6.4. *Assume **(UD)**[39] and that $\widehat{\mathbb{P}} = \Theta\mathbb{P}$ with Θ as in Definition 2.5. Then, for any $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$ such that $(\Theta\mathbb{Q})_t$ and $\mathbb{Q}_t$ are equivalent for all t, $\mathbb{I}(\mathbb{Q}) < +\infty$, and $\mathbb{I}(\Theta\mathbb{Q}) < +\infty$, we have*

$$\mathbb{I}(\Theta\mathbb{Q}) = \mathbb{I}(\mathbb{Q}) + \mathrm{ep}(\mathbb{Q}), \tag{6.9}$$

where we set $\mathrm{ep}(\mathbb{Q}) = \varsigma(\Theta\mathbb{Q}) - \varsigma(\mathbb{Q})$. *Moreover,*

$$\mathrm{ep}(\mathbb{Q}) = \lim_{t\to\infty} \frac{1}{t}\langle \sigma_t, \mathbb{Q}\rangle. \tag{6.10}$$

Proof. **(UD)** and Proposition 6.3 imply

$$\mathbb{I}(\mathbb{Q}) = \varsigma(\mathbb{Q}) - h(\mathbb{Q}), \quad \mathbb{I}(\Theta\mathbb{Q}) = \varsigma(\Theta\mathbb{Q}) - h(\Theta\mathbb{Q}). \tag{6.11}$$

Since θ_t is a bijection, we see that

$$h_t(\mathbb{Q}) = -\sum_{w\in\Omega_t} \mathbb{Q}_t(w)\log\mathbb{Q}_t(w) = -\sum_{w\in\Omega_t} \mathbb{Q}_t\big(\theta_t(w)\big)\log\mathbb{Q}_t\big(\theta_t(w)\big) = h_t(\Theta\mathbb{Q}).$$

It follows that $h(\Theta\mathbb{Q}) = h(\mathbb{Q})$. Combining this with (6.11), we arrive at (6.9). We now prove (6.10). As was already observed in the proof of Proposition 6.3, the conditions $\mathbb{I}(\mathbb{Q}) < +\infty$, and $\mathbb{I}(\Theta\mathbb{Q}) < +\infty$ imply that $\mathbb{Q}_t \ll \mathbb{P}_t$ and $(\Theta\mathbb{Q})_t \ll \mathbb{P}_t$ for all t. We remark that

$$\begin{aligned}\langle \sigma_t, \mathbb{Q}\rangle &= -\left\langle \log\frac{\mathbb{Q}_t}{\mathbb{P}_t}, \mathbb{Q}_t\right\rangle + \left\langle \log\frac{\mathbb{Q}_t}{\widehat{\mathbb{P}}_t}, \mathbb{Q}_t\right\rangle = -\left\langle \log\frac{\mathbb{Q}_t}{\mathbb{P}_t}, \mathbb{Q}_t\right\rangle + \left\langle \log\frac{(\Theta\mathbb{Q})_t}{\mathbb{P}_t}, (\Theta\mathbb{Q})_t\right\rangle \\ &= -\mathrm{Ent}(\mathbb{Q}_t|\mathbb{P}_t) + \mathrm{Ent}((\Theta\mathbb{Q})_t|\mathbb{P}_t).\end{aligned}$$

Dividing this relation by t, passing to the limit as $t \to \infty$, and using (6.6), we obtain

$$\lim_{t\to\infty}\frac{1}{t}\langle\sigma_t,\mathbb{Q}\rangle = \mathbb{I}(\Theta\mathbb{Q}) - \mathbb{I}(\mathbb{Q}).$$

Comparing this with (6.9), we arrive at the required relation (6.10). □

A Appendix

A.1 Technical results

We first prove two lemmas justifying Remarks 2.3 and 2.4.

Lemma A.1. *Assume that **(UD)** holds for both $\mathbb{P}$ and $\widehat{\mathbb{P}}$, and that **(SLD)** holds for both $\mathbb{P}$ and $\widehat{\mathbb{P}}$ with the same ξ, in the sense that for all $t \in \mathbb{N}$, all $u \in \Omega_t$ and all $v \in \Omega_{\mathrm{fin}}$, $|v| \geq 1$, there exists $|\xi| \leq \tau_t$ such that for both $\mathbb{P}^\sharp = \mathbb{P}$ and $\mathbb{P}^\sharp = \widehat{\mathbb{P}}$,*

$$\mathrm{e}^{-c_t}\mathbb{P}^\sharp(u)\mathbb{P}^\sharp(v) \leq \mathbb{P}^\sharp(u\xi v).$$

*Then **(SSD)** holds (for some larger τ_t and c_t).*

[39] The same remark as in Proposition 6.3 applies: although it is a standing assumption in this section, **(SLD)** is not necessary in this proposition if we simply define $\mathbb{I}$ by (6.4).

Proof. Let $u \in \Omega_t$ and $v \in \Omega_{\text{fin}}$. Then, by Lemma 3.5, there exists $b \in \Omega_{\tau_t}$ such that $\mathbb{P}^\sharp(v) \geq \mathbb{P}^\sharp(bv) \geq \mathbb{P}^\sharp(v)\mathrm{e}^{-C\tau_t}$. By assumption there is $|\xi| \leq \tau_t$ such that

$$\mathbb{P}^\sharp(u\xi bv) \geq \mathrm{e}^{-c_t}\mathbb{P}^\sharp(u)\mathbb{P}^\sharp(bv) \geq \mathrm{e}^{-c_t - C\tau_t}\mathbb{P}^\sharp(u)\mathbb{P}^\sharp(v).$$

Let then $\xi' = \xi b$. By **(UD)**, we have

$$\mathbb{P}^\sharp(u\xi' v) \leq \mathrm{e}^{c_t}\mathbb{P}^\sharp(u)\mathbb{P}^\sharp(\xi'_{[\tau_t+1,|\xi'|]}v) \leq \mathrm{e}^{c_t}\mathbb{P}^\sharp(u)\mathbb{P}^\sharp(v).$$

(Note that $\xi'_{[\tau_t+1,|\xi'|]}$ may be the empty word.) Thus,

$$\mathrm{e}^{-c_t - C\tau_t}\mathbb{P}^\sharp(u)\mathbb{P}^\sharp(v) \leq \mathbb{P}^\sharp(u\xi' v) \leq \mathrm{e}^{c_t}\mathbb{P}^\sharp(u)\mathbb{P}^\sharp(v).$$

Since $|\xi'| \leq 2\tau_t$, **(SSD)** holds with τ_t and c_t replaced with the sequences $2\tau_t$ and $c_t + C\tau_t$, which are also $o(t)$. □

Turning to Remark 2.4, we now give a sufficient condition for $\mathbb{P}$ to be ergodic (which is fulfilled, in particular, if **(SLD)** holds with $\sup_t \tau_t < \infty$ and $\sup_t c_t < \infty$).

Lemma A.2. *Assume the following form of lower decoupling: there exist $c > 0$ and $k \in \mathbb{N}_0$ such that for all $t \in \mathbb{N}$, all $u \in \Omega_t$ and all $v \in \Omega_{\text{fin}}$, $|v| \geq 1$,*

$$\sum_{\substack{\xi \in \Omega_{\text{fin}} \\ \tau_t - k \leq |\xi| \leq \tau_t}} \mathbb{P}(u\xi v) \geq \mathrm{e}^{-c}\mathbb{P}(u)\mathbb{P}(v).$$

Then $\mathbb{P}$ is ergodic.

Proof. Consider first two cylinder sets $\mathcal{C}_1$ and $\mathcal{C}_2$ given by $\mathcal{C}_i = \{\omega \in \Omega \mid \omega_{[1,r]} \in C_i\}$ for some $r \in \mathbb{N}$ and sets $C_i \subset \Omega_r$, $i = 1, 2$. Observe that, by assumption,

$$\begin{aligned}\sum_{j=\tau_t-k}^{\tau_t} \mathbb{P}(\mathcal{C}_1 \cap \varphi^{-r-j}\mathcal{C}_2) &= \sum_{j=\tau_t-k}^{\tau_t} \sum_{u \in C_1} \sum_{v \in C_2} \sum_{\xi \in \Omega_j} \mathbb{P}(u\xi v) \\ &\geq \mathrm{e}^{-c} \sum_{u \in C_1} \sum_{v \in C_2} \mathbb{P}(u)\mathbb{P}(v) = \mathrm{e}^{-c}\,\mathbb{P}(\mathcal{C}_1)\,\mathbb{P}(\mathcal{C}_2).\end{aligned}$$

Thus, there exists $t \in [r + \tau_t - k, r + \tau_t]$ such that

$$\mathbb{P}(\mathcal{C}_1 \cap \varphi^{-t}\mathcal{C}_2) \geq C\,\mathbb{P}(\mathcal{C}_1)\,\mathbb{P}(\mathcal{C}_2),$$

where $C = \mathrm{e}^{-c}/(k+1) > 0$. Since any Borel set in Ω can be approximated by cylinder sets (and the constant C is independent of the choice of $\mathcal{C}_i$), it follows that $\mathbb{P}$ is ergodic. The details are as follows. Assume $B \subset \Omega$ is an invariant Borel set (*i.e.*, $\mathbb{P}(B \triangle \varphi^{-1}B) = 0$),[40] and let $\varepsilon > 0$. We can find two cylinder sets $\mathcal{C}_1, \mathcal{C}_2$ that approximate B and B^c, in the sense that $\mathbb{P}(B^c \triangle \mathcal{C}_1) + \mathbb{P}(B \triangle \mathcal{C}_2) \leq \varepsilon$. Then,

$$\begin{aligned}0 = \sup_{t \in \mathbb{N}_0} \mathbb{P}(B^c \cap \varphi^{-t}B) &\geq \sup_{t \in \mathbb{N}_0} \mathbb{P}(\mathcal{C}_1 \cap \varphi^{-t}\mathcal{C}_2) - \varepsilon \\ &\geq C\,\mathbb{P}(\mathcal{C}_1)\,\mathbb{P}(\mathcal{C}_2) - \varepsilon \geq C(\mathbb{P}(B^c) - \varepsilon)(\mathbb{P}(B) - \varepsilon) - \varepsilon \\ &\geq C\,\mathbb{P}(B^c)\mathbb{P}(B) - (2C + 1)\varepsilon.\end{aligned}$$

Since ε was arbitrary, we have $\mathbb{P}(B^c)\mathbb{P}(B) = 0$, so that $\mathbb{P}(B) \in \{0, 1\}$. This completes the proof. □

The next lemma proves the properties of irreducible Markov processes mentioned in Example 2.20.

[40] The symmetric difference $A \triangle B$ of two sets A and B is defined as $(A \setminus B) \cup (B \setminus A)$.

Lemma A.3. *Let $\mathbb{P} \in \mathcal{P}_\varphi(\Omega)$ be a Markov process. Then **(UD)** holds. Assume furthermore that it is irreducible (i.e., that for all $a, b \in \mathcal{A}$, there exists $\xi^{(a,b)} \in \Omega_{\text{fin}}$ such that $\mathbb{P}(a\xi^{(a,b)}b) > 0$). Then **(SLD)** holds. If, in addition, $\widehat{\mathbb{P}} \in \mathcal{P}_\varphi(\Omega)$ is another Markov process such that $\mathbb{P}_2 \ll \widehat{\mathbb{P}}_2$, then **(SSD)** holds.*

Proof. Since $\mathbb{P}$ is Markov and shift-invariant, we have

$$\mathbb{P}(w) = \mathbb{P}_1(w_1)P(w_1; w_2)P(w_2; w_3)\cdots P(w_{t-1}; w_t), \qquad w \in \Omega_t,$$

for some transition matrix $(P(a; b))_{a,b\in\mathcal{A}}$.

Upper Decoupling. We show that **(UD)** holds with $\tau_t \equiv 0$ and

$$c_t \equiv -\min_{a\in\mathcal{A}:\mathbb{P}_1(a)>0} \log \mathbb{P}_1(a).$$

Indeed, given $u \in \Omega_t$ and $v \in \Omega_{\text{fin}}$ such that $\mathbb{P}_1(v_1) > 0$, we have

$$\mathbb{P}(uv) = \frac{P(u_t; v_1)}{\mathbb{P}_1(v_1)}\mathbb{P}(u)\mathbb{P}(v) \leq \mathrm{e}^{c_t}\mathbb{P}(u)\mathbb{P}(v).$$

If $\mathbb{P}_1(v_1) = 0$, then by invariance $\mathbb{P}(uv) = 0 \leq \mathrm{e}^{c_t}\mathbb{P}(u)\mathbb{P}(v)$, so that **(UD)** is proved.

Selective Lower Decoupling. Assume now that the process is irreducible. This implies that $\mathbb{P}_1(a) > 0$ for all $a \in \mathcal{A}$. Let $\tau = \max_{a,b\in\mathcal{A}} |\xi^{(a,b)}|$. Given two words $u \in \Omega_t$ and $v \in \Omega_{\text{fin}}$, $|v| \geq 1$, let $\xi = \xi^{(u_t,v_1)}$. Then either $|\xi| = 0$, in which case

$$\mathbb{P}(uv) = \frac{P(u_t; v_1)}{\mathbb{P}_1(v_1)}\mathbb{P}(u)\mathbb{P}(v), \tag{A.1}$$

or $k := |\xi^{(u_t,v_1)}| \geq 1$, in which case

$$\mathbb{P}(u\xi v) = \frac{P(u_t; \xi_1)\mathbb{P}(\xi)P(\xi_k; v_1)}{\mathbb{P}_1(\xi_1)\mathbb{P}_1(v_1)}\mathbb{P}(u)\mathbb{P}(v). \tag{A.2}$$

The factors in front of $\mathbb{P}(u)\mathbb{P}(v)$ on the right-hand sides of (A.1) and (A.2) are positive and depend only on u_t and v_1. We obtain a lower bound by taking the minimum over all possible values of u_t and v_1. This implies that **(SLD)** holds with $\tau_t \equiv \tau$ and some c_t independent of t.

Selective Symmetric Decoupling. Assume finally that $\widehat{\mathbb{P}}$ is another Markov process such that $\mathbb{P}_2 \ll \widehat{\mathbb{P}}_2$. Then by the Markov property we have that $\mathbb{P}_t \ll \widehat{\mathbb{P}}_t$ for all t, so $\widehat{\mathbb{P}}$ is irreducible, and one can choose the same $\xi^{(a,b)}$ as for $\mathbb{P}$ (*i.e.*, we have both $\mathbb{P}(a\xi^{(a,b)}b) > 0$ and $\widehat{\mathbb{P}}(a\xi^{(a,b)}b) > 0$). Considerations similar to the above imply that **(SSD)** holds. □

Finally, for the reader's convenience, we prove some well-known properties of the Kolmogorov–Sinai entropy that are used in the proof of Proposition 6.3 (see for example Section 4.3 of [21]).

Lemma A.4. *For all $\mathbb{Q} \in \mathcal{P}_\varphi(\Omega)$, the limit*

$$h(\mathbb{Q}) = \lim_{t\to\infty} \frac{1}{t} h_t(\mathbb{Q}) \tag{A.3}$$

exists, is finite, and the mapping $h : \mathcal{P}_\varphi(\Omega) \to [0, \infty)$ is upper semicontinuous and affine.

Proof. First, it follows from φ-invariance and the inequality $\log x \leq x - 1$ that

$$h_{t+t'}(\mathbb{Q}) - h_t(\mathbb{Q}) - h_{t'}(\mathbb{Q}) = \sum_{w\in\Omega_t} \sum_{w'\in\Omega_{t'}} \mathbb{Q}(ww') \log \frac{\mathbb{Q}(w)\mathbb{Q}(w')}{\mathbb{Q}(ww')} \leq 0.$$

By subadditivity, the limit (A.3) exists, is finite, and $h(\mathbb{Q}) = \inf_{t\in\mathbb{N}} \frac{1}{t} h_t(\mathbb{Q})$. Moreover, as an infimum over a family of continuous functions, h is upper semicontinuous. That h is affine is an immediate consequence of the following relation: for all $\mathbb{Q}^{(1)}, \mathbb{Q}^{(2)} \in \mathcal{P}_\varphi(\Omega)$, and all $p_1 \in (0,1)$, $p_2 = 1 - p_1$, we have

$$\sum_{i=1,2} p_i h_t(\mathbb{Q}^{(i)}) \leq h_t\left(\sum_{i=1,2} p_i \mathbb{Q}^{(i)}\right) \leq \sum_{i=1,2} p_i h_t(\mathbb{Q}^{(i)}) - \sum_{i=1,2} p_i \log p_i. \tag{A.4}$$

To complete the proof, we now establish (A.4). The first inequality follows from the concavity of $x \mapsto f(x) := -x \log x$. Indeed, we have

$$\sum_{i=1,2} p_i h_t(\mathbb{Q}^{(i)}) = \sum_{w\in\Omega_t} \sum_{i=1,2} p_i f\left(\mathbb{Q}^{(i)}(w)\right) \leq \sum_{w\in\Omega_t} f\left(\sum_{i=1,2} p_i \mathbb{Q}^{(i)}(w)\right) = h_t\left(\sum_{i=1,2} p_i \mathbb{Q}^{(i)}\right).$$

For the second inequality, we observe that

$$\begin{aligned} h_t\left(\sum_{i=1,2} p_i \mathbb{Q}^{(i)}\right) &= -\sum_{w\in\Omega_t} \sum_{i=1,2} p_i \mathbb{Q}^{(i)}(w) \log\left(\sum_{j=1,2} p_j \mathbb{Q}^{(j)}(w)\right) \\ &\leq -\sum_{w\in\Omega_t} \sum_{i=1,2} p_i \mathbb{Q}^{(i)}(w) \log\left(p_i \mathbb{Q}^{(i)}(w)\right) \\ &= -\sum_{i=1,2} p_i \sum_{w\in\Omega_t} \mathbb{Q}^{(i)}(w) \log\left(\mathbb{Q}^{(i)}(w)\right) - \sum_{i=1,2} p_i \log p_i \sum_{w\in\Omega_t} \mathbb{Q}^{(i)}(w) \\ &= \sum_{i=1,2} p_i h_t(\mathbb{Q}^{(i)}) - \sum_{i=1,2} p_i \log p_i. \end{aligned}$$

The proof is complete. □

A.2 Hidden Markov chain example

In the section we discuss the hidden Markov chain of Example 2.24. For reasons of space, we only outline the main steps of the analysis; the details are easy to fill.

Let $(\gamma(n))_{n\in\mathbb{N}_0}$ be a sequence of non-negative numbers such that $\gamma(0) = 0$ and $\gamma(n+1) \geq \gamma(n) + \varepsilon$ for all n and some $\varepsilon > 0$. We consider a countable Markov chain with states $0, 1, 2, \ldots$ such that from each state $n \in \mathbb{N}_0$, we move either to $n+1$ with probability $g(n+1) := \mathrm{e}^{\gamma(n)-\gamma(n+1)}$ or we move to 0 with probability $1 - g(n+1)$ (see Figure 2).

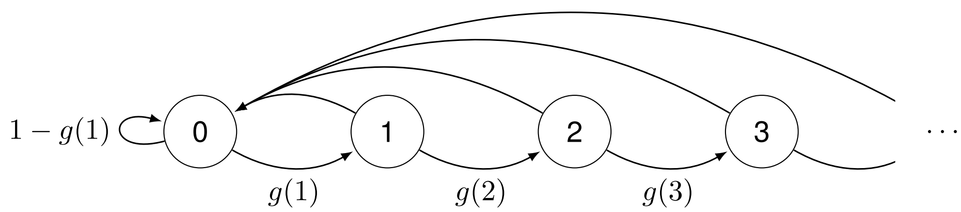


Figure 2: Illustration of the Markov chain.

This Markov chain admits a unique invariant measure, and we denote by $\mathbb{Q}$ the corresponding Markov process on $\mathbb{N}_0^{\mathbb{N}}$. Note that

$$\mathbb{Q}_1(n) = Z^{-1}\mathrm{e}^{-\gamma(n)}, \quad Z = \sum_{n=0}^{\infty} \mathrm{e}^{-\gamma(n)}.$$

Let $\mathcal{A} = \{\mathrm{a}, \mathrm{b}\}$ and let $\Psi : \mathbb{N}_0^{\mathbb{N}} \to \mathcal{A}^{\mathbb{N}}$ be defined by $\Psi(\omega_1, \omega_2, \dots) = (\psi(\omega_1), \psi(\omega_2), \dots)$, where $\psi(0) = \mathrm{a}$ and $\psi(n) = \mathrm{b}$ for all $n \geq 1$. Our main object of interest is the invariant probability measure on $\Omega = \mathcal{A}^{\mathbb{N}}$ defined by $\mathbb{P} = \mathbb{Q} \circ \Psi^{-1}$. The following holds:

- The measure $\mathbb{P}$ has full support, and for any words u and v we have

  $$\mathbb{P}(u\mathrm{a}v) = \mathbb{P}(u\mathrm{a})\mathbb{P}(\mathrm{a}v)/\mathbb{P}_1(\mathrm{a}).$$

  Moreover, $\mathbb{P}$ is reversible in the sense that $\mathbb{P}_t(w_1, \dots, w_t) = \mathbb{P}_t(w_t, \dots, w_1)$ for all $t \in \mathbb{N}$ and $w \in \Omega_t$.

- Let us set

  $$p_t(w) = \frac{\mathbb{P}_{t+2}(\mathrm{a}w\mathrm{a})}{\mathbb{P}_1(\mathrm{a})}, \qquad w \in \Omega_t,$$

  with the convention $p_0(\kappa) = \mathbb{P}_2(\mathrm{aa})/\mathbb{P}_1(\mathrm{a}) = 1 - g(1)$, where κ is the empty word. Then

  $$p_{|u|+|v|+1}(u\mathrm{a}v) = p_{|u|}(u)p_{|v|}(v) \quad \text{for all } u, v \in \Omega_{\mathrm{fin}}.$$

- The quantities $\mathbb{P}_{t+1}(\mathrm{b}^t\mathrm{a})$, $\mathbb{P}_{t+1}(\mathrm{ab}^t)$ and $\mathbb{P}_t(\mathrm{b}^t)$ are bounded above and below by some constant (independent of t) times $\mathrm{e}^{-\gamma(t)}$. More generally, the quantities $\mathbb{P}_{t+1}(\mathrm{a}w)$, $\mathbb{P}_{t+1}(w\mathrm{a})$ and $p_t(w)$ are bounded above and below by a constant (independent of t and $w \in \Omega_t$) times $\mathbb{P}_t(w)$.
- $\mathbb{P}$ satisfies **(SLD)** with $\tau_t \equiv 1$ and $\sup_t c_t < \infty$ (by taking $\xi = \mathrm{a}$ in (2.2)).
- $\mathbb{P}$ satisfies **(UD)** with $\tau_t \equiv 0$ and $c_t = c + \sup_{n \in \mathbb{N}_0}[\gamma(n) + \gamma(t) - \gamma(n+t)]$ for some $c > 0$, provided that $(\gamma(n))_{n\in\mathbb{N}_0}$ is such that $c_t = o(t)$.

We assume now that $\widehat{\mathbb{P}}$ is constructed in the same way, with $(\widehat{\gamma}(n))_{n\in\mathbb{N}_0}$ satisfying the same conditions as $(\gamma(n))_{n\in\mathbb{N}_0}$. We define $\widehat{p}_t$ in the same way as p_t. Then the following holds:

- The pair $(\mathbb{P}, \widehat{\mathbb{P}})$ satisfies **(SSD)** with $\tau_t \equiv 1$ and $\sup_t c_t < \infty$ (by taking $\xi = \mathrm{a}$ in (2.4)).
- The function q defined in (2.8) can be written as

  $$q(\alpha) = \lim_{t\to\infty} \frac{1}{t} \log \sum_{w\in\Omega_t} \zeta_{t,\alpha}(w), \qquad \alpha \in \mathbb{R},$$

  where the quantity $\zeta_{t,\alpha}(w) := \mathrm{e}^{(\alpha+1)\log p_t(w) - \alpha \log \widehat{p}_t(w)}$ defined for $t \in \mathbb{N}_0$ and $w \in \Omega_t$ satisfies the relation

  $$\zeta_{|u|+|v|+1,\alpha}(u\mathrm{a}v) = \zeta_{|u|,\alpha}(u)\zeta_{|v|,\alpha}(v), \qquad u, v \in \Omega_{\mathrm{fin}}.$$

- We have $q(\alpha) = -\log \rho(\alpha)$, where $\rho(\alpha)$ is the radius of convergence of the power series

  $$R_\alpha(x) = \sum_{t\in\mathbb{N}_0} r_t(\alpha)x^t, \qquad r_t(\alpha) = \sum_{w\in\Omega_t} \zeta_{t,\alpha}(w).$$

- Let us set

  $$U_\alpha(x) = \sum_{t\in\mathbb{N}_0} u_t(\alpha)x^t, \qquad u_t(\alpha) = \zeta_{t,\alpha}(\mathrm{b}^t),$$

and observe that

$$u_t(\alpha) = (1 - g(t+1))^{\alpha+1}(1 - \widehat{g}(t+1))^{-\alpha}\mathrm{e}^{-(\alpha+1)\gamma(t)+\alpha\widehat{\gamma}(t)},$$

where $1 - g(t+1)$ and $1 - \widehat{g}(t+1)$ are bounded from above by one and from below by a constant $c > 0$ uniformly in t. The radius of convergence of U_α is

$$\kappa(\alpha) = \liminf_{t\to\infty} \exp((\alpha+1)t^{-1}\gamma(t) - \alpha t^{-1}\widehat{\gamma}(t)).$$

- By sorting the words $w \in \Omega_t$ according to the position of the first occurrence of the symbol a (if there is one), we get the renewal equation

$$r_t(\alpha) = \sum_{k=0}^{t-1} u_k(\alpha) r_{t-1-k}(\alpha) + u_t(\alpha).$$

This relation translates into the algebraic equation $R_\alpha(x) = xR_\alpha(x)U_\alpha(x) + U_\alpha(x)$, so that

$$R_\alpha(x) = \frac{U_\alpha(x)}{1 - xU_\alpha(x)}.$$

The power series defined above have strictly positive coefficients for all α. They are strictly increasing functions of $x \geq 0$, and jointly lower semicontinuous in $\alpha \in \mathbb{R}$ and $x \geq 0$. As already discussed, the quantity of interest is the radius of convergence $\rho(\alpha)$ of R_α. For each fixed α, we are in one of the following two cases:

(a) There exists $x > 0$ such that $xU_\alpha(x) = 1$, and in this case $\rho(\alpha) = x$.

(b) $xU_\alpha(x) < 1$ for all $0 \leq x \leq \kappa(\alpha)$, in which case $\rho(\alpha) = \kappa(\alpha)$.

In both cases, we have $\rho(\alpha) = \sup\{x \geq 0 : xU_\alpha(x) \leq 1\}$. In case (a), $\rho'(\alpha)$ can be obtained by differentiating the relation $\rho(\alpha)U_\alpha(\rho(\alpha)) \equiv 1$, and we obtain

$$\rho'(\alpha) = -\frac{\sum_{n\geq 0} \partial_\alpha u_n(\alpha)\rho^{n+1}(\alpha)}{\sum_{n\geq 0} u_n(\alpha)(n+1)\rho^n(\alpha)}. \tag{A.5}$$

In case (b), we simply have $\rho'(\alpha) = \kappa'(\alpha)$. (Of course these relations hold only if the corresponding quantities are well defined, and if α is not at a transition point between the two cases.)

Different situations can occur depending on the concrete choice of $(\gamma(n))_{n\in\mathbb{N}_0}$ and $(\widehat{\gamma}(n))_{n\in\mathbb{N}_0}$. We now briefly discuss six interesting cases. We do not give any proofs, as these examples are easily understood by substituting the relevant values in the formulas for $\kappa(\alpha)$ and $\rho(\alpha)$ (and their derivative). The interested reader may wish to investigate the matter further by trying other expressions for $(\gamma(n))_{n\in\mathbb{N}_0}$ and $(\widehat{\gamma}(n))_{n\in\mathbb{N}_0}$.

Example 1. Let $\gamma(n) = n$ and $\widehat{\gamma}(n) = n^2$. We have $\kappa(\alpha) = +\infty$ for $\alpha < 0$, $\kappa(0) = \mathrm{e}$, and $\kappa(\alpha) = 0$ for $\alpha > 0$. For $\alpha \leq 0$ we are in case (a), and it follows from the identity $\rho(\alpha)U_\alpha(\rho_\alpha) \equiv 1$ that ρ, and hence q, are analytic on $(-\infty, 0)$. We know already that $q(0) = 0$. When $\alpha > 0$, we have $\rho(\alpha) = \kappa(\alpha) = 0$, and hence we are in case (b) with $q(\alpha) = +\infty$.[41] Evaluating the quantity (A.5) in the limit $\alpha \uparrow 0$ and $\rho(\alpha) \downarrow \rho(0) = 1$ shows that $0 > \rho'(0^-) > -\infty$. Since $q'(\alpha) = -\frac{d}{d\alpha}\log\rho(\alpha) = -\frac{\rho'(\alpha)}{\rho(\alpha)}$, we conclude that $q'(0^-)$ is finite (numerical evaluation gives $q'(0^-) = 0.3294...$). See Figure 3a.[42]

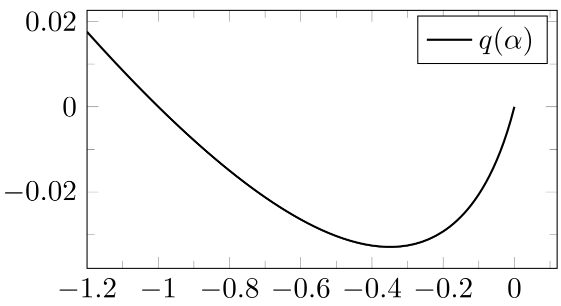
[41] The fact that $q(\alpha)$ is infinite when $\alpha > 0$ also follows from the observation that in the sum $\sum_{w\in\Omega_t} \mathrm{e}^{\alpha\sigma_t(w)}\mathbb{P}_t(w)$, the contribution of $w = \mathrm{b}^t$ grows like $\mathrm{e}^{-(\alpha+1)t+\alpha t^2}$.

[42] A bisection method was used to find $\rho(\alpha) = \sup\{x \geq 0 : xU_\alpha(x) \leq 1\}$. After that, $q = -\log\rho$ was obtained by direct computation.

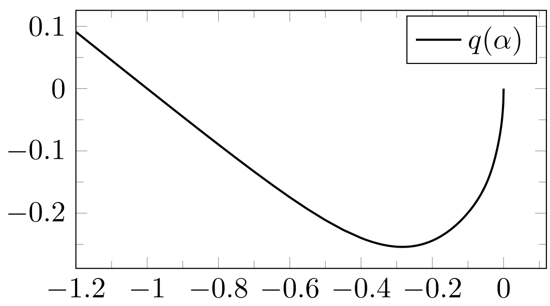
Example 2. By replacing n^2 with $\exp(2n)$ in Example 1, we obtain the same results except that now $q'(0^-) = +\infty$. See Figure 3b.

Example 3. Let $\gamma(n) = n + cn^2$ and $\widehat{\gamma}(n) = n^2$ with $c \in (0, 1)$. An analysis similar to that of Example 1 shows that $q(\alpha)$ is infinite for $\alpha > \alpha_* := c/(1-c)$, and finite for $\alpha \le \alpha_*$. The case $c = 1/2$, $\alpha_* = 1$ is represented in Figure 3c.

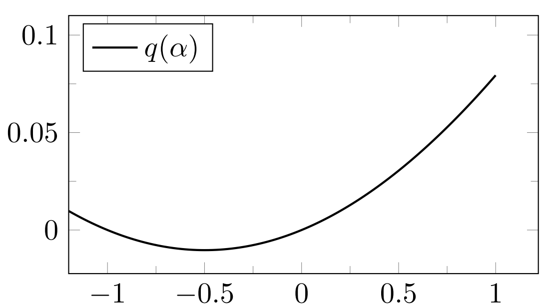
Example 4. Let $\gamma(n) = cn^2$ and $\widehat{\gamma}(n) = n^2 + n^{3/2}$ with $c \in (0, 1)$. For $\alpha < \alpha_* := c/(1-c)$ we are in case (a), while for $\alpha \ge \alpha_*$ we are in case (b) with $\kappa(\alpha) = 0$ and $q(\alpha) = +\infty$. The function q increases continuously to $+\infty$ when $\alpha \uparrow \alpha_*$. The case $c = 1/2$ is plotted in Figure 3d.



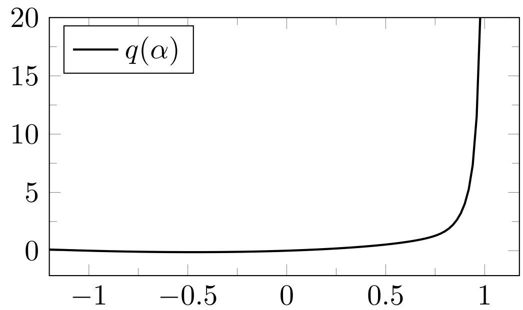
(a) Example 1



(b) Example 2



(c) Example 3



(d) Example 4

Figure 3: Graph of q for Examples 1–4. The horizontal axis represents α. In Examples 1–3, the line is interrupted where q jumps to $+\infty$.

Example 5. Let $\gamma(n) = n$ and $\widehat{\gamma}(n) = 2n - 2\log(1 + n/2)$. In this case, $\kappa(\alpha) = \exp(1 - \alpha)$. One shows that there exists $\alpha_* > 1$ (numerical evaluation gives $\alpha_* = 1.1305...$) such that we are in case (a) if $\alpha \le \alpha_*$, and in case (b) if $\alpha > \alpha_*$. By evaluating (A.5) as $\alpha \uparrow \alpha_*$, $\rho(\alpha) \downarrow \rho(\alpha_*) = \kappa(\alpha_*)$ and comparing with $\kappa'(\alpha_*)$, one shows that q is not differentiable at α_*. See Figure 4.

Example 6. Take now $\gamma(0) = 0$, $\gamma(1) = 0.01$, $\gamma(n) = n + 5\log(1 + n/5)$ for $n \ge 2$, and $\widehat{\gamma}(n) = 10 + 2n + 5\log(1 + n/5)$ for all $n \ge 0$. Here again, $\kappa(\alpha) = \exp(1 - \alpha)$. Explicit computations show that $\kappa(\alpha)U_\alpha(\kappa(\alpha)) < 1$ on some interval $I = (-0.6418\ldots, -0.2042\ldots)$ and $\kappa(\alpha)U_\alpha(\kappa(\alpha)) \in (1, +\infty]$ outside of the closure of I (the numerical values in the definition of $\gamma(n)$ and $\widehat{\gamma}(n)$ were chosen to ensure the existence of such an interval). Then for $\alpha \in I$ we are in case (b), and for $\alpha \notin I$ we are in case (a). It follows that q is analytic everywhere except at the boundaries of I. Explicit computations using (A.5) show that q is not differentiable at those boundaries. See Figure 5.

We note that in all the examples except the last, both $\mathbb{P}$ and $\widehat{\mathbb{P}}$ satisfy **(UD)** with $\tau_t \equiv 0$ and

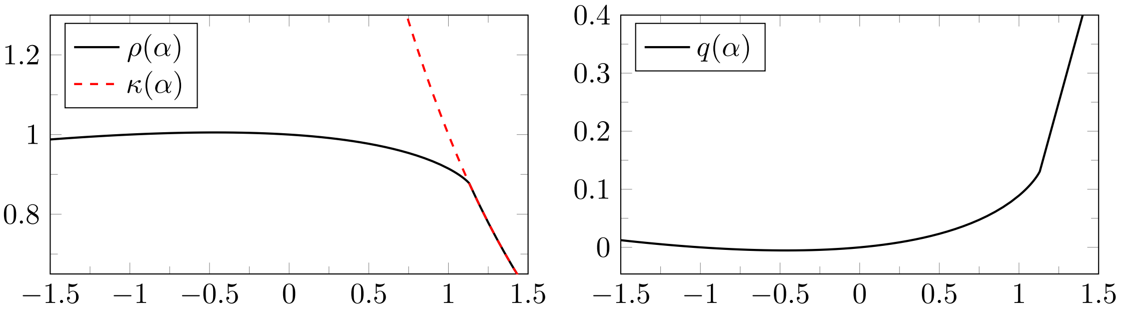


Figure 4: Numerical evaluation of ρ, κ and q for Example 5.

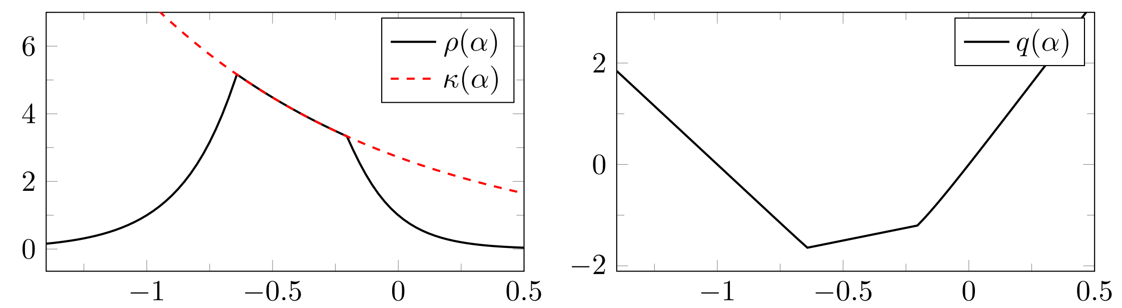


Figure 5: Numerical evaluation of ρ, κ and q for Example 6.

$\sup_t c_t < \infty$. In the last example, both measures satisfy **(UD)** with $\sup_t c_t = \infty$ and $c_t = o(t)$. In order to make direct comparison with the results of [3] and [8], we introduce the following construction.

Remark A.5. Given the pair $(\mathbb{P}, \widehat{\mathbb{P}})$ constructed in this section, one can define a new pair of measures $(\mathscr{P}, \widehat{\mathscr{P}})$ on $\mathscr{A}^{\mathbb{N}}$, with the product alphabet $\mathscr{A} = \mathcal{A} \times \mathcal{A}$, such that $(\mathscr{P}, \widehat{\mathscr{P}})$ are related by an involution as in Definition 2.5. For any word $(u, v) \in \mathscr{A}^t = \mathcal{A}^t \times \mathcal{A}^t$, define $\mathscr{P}(u, v) = \mathbb{P}(u)\widehat{\mathbb{P}}(v)$, and set $\widehat{\mathscr{P}}(u, v) = \mathscr{P}(\theta_t(u, v))$ where[43] $\theta_t(u, v) = (v, u)$. It is easy to show that the pair $(\mathscr{P}, \widehat{\mathscr{P}})$ satisfies **(SSD)** (with $\xi = (\mathrm{a}, \mathrm{a}) \in \mathscr{A}$) and that the entropy production Σ_t of the pair $(\mathscr{P}, \widehat{\mathscr{P}})$ can be expressed in terms of the entropy production σ_t of $(\mathbb{P}, \widehat{\mathbb{P}})$ by

$$\Sigma_t(u, v) := \log \frac{\mathscr{P}_t(u, v)}{\widehat{\mathscr{P}}_t(u, v)} = \sigma_t(u) - \sigma_t(v), \qquad u, v \in \Omega_t.$$

As a consequence, we find that

$$\mathscr{Q}(\alpha) := \lim_{t\to\infty} \frac{1}{t} \log \langle \mathrm{e}^{\alpha\Sigma_t}, \mathscr{P} \rangle = q(\alpha) + q(-\alpha - 1), \qquad \alpha \in \mathbb{R}. \tag{A.6}$$

Note that $\mathscr{Q}$ satisfies the symmetry (2.12), *i.e.*, $\mathscr{Q}(-\alpha) = \mathscr{Q}(\alpha - 1)$ for all $\alpha \in \mathbb{R}$.

We finish with a brief comment on how, for the pairs of measures $(\mathscr{P}, \widehat{\mathscr{P}})$ constructed from the pairs $(\mathbb{P}, \widehat{\mathbb{P}})$ in the above six examples, the results of [8] and [3] fail to apply or to provide the global LDP in Theorem 2.8. We remark that in the first five cases the measures $\mathscr{P}$ and $\widehat{\mathscr{P}}$ satisfy **(UD)** with $\tau_t \equiv 0$ and $\sup_t c_t < \infty$, while in the last case, both measures satisfy **(UD)** with $\sup_t c_t = \infty$ and $c_t = o(t)$.

[43] Since $\mathbb{P}$ and $\widehat{\mathbb{P}}$ are reversible, one could as well define $\theta_t(u, v) = ((v_t, \ldots, v_1), (u_t, \ldots, u_1))$. This means that both cases of Definition 2.5 are actually covered.

- The results of [3] apply to the pairs $(\mathscr{P}, \widehat{\mathscr{P}})$ of Examples 1–5 above, and give a local LDP for in the interval $(\mathscr{Q}'(-1^+), \mathscr{Q}'(0^-))$. Only in Example 2 is this interval equal to $\mathbb{R}$ (see (A.6)), and hence in this case [3] gives the full LDP. Example 6 cannot be handled by the method of [3], because it does not satisfy **(UD)** with $\sup_t c_t < \infty$. Note also that in Example 6, $\mathscr{Q}$ is not differentiable in $(-1, 0)$, unlike in the situation of [3].
- As mentioned in [8, Example 0.16], if the sequence $(G_t)_{t\in\mathbb{N}} \subset C(\mathscr{A}^{\mathbb{N}})$ defined by $G_t = \log \mathscr{P}_t$[44] is *asymptotically additive* (see the definition and characterizations in [8]), then [8, Theorem 4.5] applies, and provides the global LDP for $t^{-1}\sigma_t$. This is the case of Examples 5 and 6 above. Examples 1–4 clearly cannot be handled by [8], as under the assumptions therein the entropic pressure $e(\alpha) = q(-\alpha)$ is finite for all $\alpha \in \mathbb{R}$.

A.3 Weak Gibbs measures

LDPs for weak Gibbs measures (on shift spaces and more general dynamical systems) have been abundantly studied; see for example [5, 42, 43, 35, 8] and the references therein. The weak Gibbs condition and our decoupling assumptions are essentially incomparable (see below). We show here that given a weak Gibbs measure supported on a subshift satisfying a suitable *specification property*, one can still construct a map $\psi_{n,t}$ satisfying the conclusions of Proposition 3.1. As our results use **(SLD)** and **(SSD)** only through the conclusions of Proposition 3.1, they remain valid for weak Gibbs measures.

We first introduce the notion of weak Gibbs measure on a subshift. The measure $\mathbb{P}$ can be viewed as an invariant measure for the dynamical system (Ω^+, φ). Recall that $\Omega^+ = \operatorname{supp} \mathbb{P}$ was defined in (2.1). In this setup the following two conditions are natural. We again assume that $\tau_t = o(t)$.

Weak specification property (WSP). *For all $t \in \mathbb{N}$, all $u \in \Omega_t^+$, and all $v \in \Omega_{\rm fin}^+$, there exists $\xi \in \Omega_{\rm fin}^+$ satisfying $|\xi| \le \tau_t$ such that $u\xi v \in \Omega_{\rm fin}^+$.*[45]

Weak Gibbs condition (WGC). *The measure $\mathbb{P}$ is* weak Gibbs[46] *with respect to some potential $f \in C(\Omega^+)$; i.e., there exists a real number p (called* pressure*) and a real sequence $(d_t)_{t\in\mathbb{N}}$ such that $d_t = o(t)$, and for all $\omega \in \Omega^+$ and all $t \in \mathbb{N}$,*

$$e^{-d_t + S_t f(\omega) - tp} \le \mathbb{P}_t(\omega_1, \dots, \omega_t) \le e^{d_t + S_t f(\omega) - tp}.$$

Measures satisfying **(WGC)** with $\sup_t d_t < \infty$ are called *Gibbs measures* (see Example 2.21). Without loss of generality, we shall assume that $p = 0$ (by replacing f with $f - p$ if necessary).

Note that **(WSP)** is a condition on the structure of the set Ω^+, while **(WGC)** is a condition[47] on $\mathbb{P}$. Once the set Ω^+ is fixed, **(WGC)** implies a strong lower bound on the probability of the "allowed" words:

$$\mathbb{P}_t(w) \ge e^{-Ct}, \qquad t \in \mathbb{N}, \quad w \in \Omega_t^+, \tag{A.7}$$

where $C = \|f\| + \sup_t d_t/t$. Our decoupling assumptions are different in philosophy, as they are formulated at the level of measures only. They compare to **(WSP)** and **(WGC)** as follows.

- As mentioned in Example 2.21, if **(WSP)** holds and $\mathbb{P}$ is a Gibbs measure (*i.e.*, **(WGC)** holds with $\sup_t d_t < \infty$), then the **(UD)** and **(SLD)** assumptions are satisfied. On the contrary, if

[44] $\mathscr{P}_t$ is defined as the marginal of $\mathscr{P}$ on the first t coordinates of $\mathscr{A}^{\mathbb{N}}$.

[45] For similar and weaker forms of specification properties and related results, see for example [36, 37, 40, 23]

[46] All the considerations in this section can be adapted with minor technical changes to the case where the potential is *asymptotically additive*. For definitions, see for example [8] and references therein.

[47] From the point of view of dynamical systems, one is first given a subshift Ω^+ satisfying **(WSP)**, and then one introduces (weak) Gibbs measures on it.

$\sup_t d_t = \infty$, then **(WSP)** and **(WGC)** do not imply any of our decoupling assumptions in general.

- **(SLD)** implies **(WSP)**.
- Even put together, **(SLD)** and **(UD)** do not imply **(WGC)** in general, as (A.7) may fail. Indeed, **(SLD)** ensures that there is one $\xi \in \Omega_{\mathrm{fin}}$, $|\xi| \le \tau_t$ such that $\mathbb{P}(u\xi v) \ge \mathrm{e}^{-c_t}\mathbb{P}(u)\mathbb{P}(v)$, and says nothing about $\mathbb{P}(u\xi' v)$ for $\xi' \ne \xi$. See Example 2.24 and Appendix A.2.

We now establish an analogue of Proposition 3.1 for **(WSP)** and **(WGC)**. Here, n, t, t', N, $\mathbb{P}^{(n)}$ and $\Lambda_{t'}$ are as in Section 3.1.

Proposition A.6. *Assume **(WSP)** and **(WGC)**. Then there exists a map $\psi_{n,t} : \Omega_{t'} \to \Omega_t$ such that the following holds.*

1. *We have*

$$\mathbb{P}_{t'}^{(n)} \circ \psi_{n,t}^{-1} \le \mathrm{e}^{g(n,t)}\, \mathbb{P}_t, \quad \text{with } \lim_{n\to\infty} \limsup_{t\to\infty} \frac{1}{t} g(n,t) = 0. \tag{A.8}$$

2. *Let $\widehat{\mathbb{P}} \in \mathcal{P}_\varphi(\Omega)$, and assume one of the following: (a) $\widehat{\mathbb{P}} = \Theta\mathbb{P}$ with Θ as in Definition 2.5 and $\theta_t(\Omega_t^+) = \Omega_t^+$; or (b) there exists $\widehat{d}_t = o(t)$ and $\widehat{f} \in C(\Omega)$ such that for all $\omega \in \Omega^+$,*[48]

$$\mathrm{e}^{-\widehat{d}_t + S_t \widehat{f}(\omega)} \le \widehat{\mathbb{P}}_t(\omega_1, \dots, \omega_t) \le \mathrm{e}^{\widehat{d}_t + S_t \widehat{f}(\omega)}. \tag{A.9}$$

Then $\psi_{n,t}$ can be chosen so that, in addition to (A.8),

$$\lim_{n\to\infty} \limsup_{t\to\infty} \frac{1}{t} \sup_{w\in\Lambda_{t'}} \left| \sigma_t(\psi_{n,t}(w)) - \sum_{k=0}^{N-1} \sigma_n(w_{[kn+1,(k+1)n]}) \right| = 0, \tag{A.10}$$

and there exists $c > 0$ such that

$$|\sigma_t(w)| \le ct, \qquad t \in \mathbb{N}, \quad w \in \Omega_t^+. \tag{A.11}$$

Proof. We first prove 1. For each $w \in \Omega_{t'}$, we write $w = w^1 w^2 \dots w^N$ with $w^i \in \Omega_n$. Setting

$$\psi_{n,t}(w) = bw^1\xi^1 w^2 \xi^2 \dots w^{N-1}\xi^{N-1} w^N \in \Omega_t \tag{A.12}$$

for some suitable $\xi^i \in \Omega_n$ satisfying $|\xi^i| \le \tau_n$ for all i, we shall prove that

$$\mathbb{P}(\psi_{n,t}(w)) \ge \mathrm{e}^{-g_1(n,t)}\mathbb{P}^{(n)}(w), \quad \text{with } \lim_{n\to\infty} \limsup_{t\to\infty} \frac{1}{t} g_1(n,t) = 0. \tag{A.13}$$

Then, the conclusion of Part 1 follows from the same combinatorial argument as in Proposition 3.1 (see the discussion after (3.14)). In order to prove (A.13), we assume that $w^i \in \Omega_n^+$ for all i (equivalently, that $w \in \Lambda_{t'}$), as the result is trivial otherwise. By following the same strategy as in Proposition 3.1, using **(WSP)** instead of **(SLD)**, we can choose $\xi_1, \dots, \xi^{N-1}$ and b such that $\psi_{n,t}(w) \in \Omega_t^+$. Next, let $\omega \in \Omega^+$ be such that $\omega_{[1,t]} = \psi_{n,t}(w)$. By **(WGC)**,

$$\mathbb{P}_t(\psi_{n,t}(w)) \ge \mathrm{e}^{S_t f(\omega) - d_t} \ge \mathrm{e}^{-g_1(n,t)} \prod_{i=1}^N \mathbb{P}_n(w^i), \tag{A.14}$$

where $g_1(n,t) = d_t + Nd_n + (t-t')\|f\|$. The relations $N \le t/n$ and (3.3) imply (A.13), which completes the proof of Part 1.

[48] Note that no requirement is made for $\omega \notin \Omega^+$.

We now prove Part 2. By combining (A.14) with the corresponding upper bound, we obtain

$$\lim_{n\to\infty}\limsup_{t\to\infty}\frac{1}{t}\sup_{w\in\Lambda_{t'}}\left|\log\mathbb{P}_t(\psi_{n,t}(w))-\log\mathbb{P}_{t'}^{(n)}(w)\right|=0. \tag{A.15}$$

Assume first that $\widehat{\mathbb{P}}$ satisfies (b). By (A.9), the relation (A.15) also holds with $\mathbb{P}$ replaced with $\widehat{\mathbb{P}}$, and (A.10) immediately follows. Moreover, by **(WGC)** and (A.9), we obtain (A.11) with $c=\|f-\widehat{f}\|+\sup_t(d_t+\widehat{d}_t)/t$.

Assume now that $\widehat{\mathbb{P}}$ satisfies (a) and let $w\in\Lambda_{t'}$. Then $\psi_{n,t}(w)\in\Omega_t^+$ and $\theta_t(\psi_{n,t}(w))\in\Omega_t^+$. Hence there exists $\widehat{\omega}\in\Omega^+$ such that $\widehat{\omega}_{[1,t]}=\theta_t(\psi_{n,t}(w))$. Since $\widehat{\mathbb{P}}_t(\psi_{n,t}(w))=\mathbb{P}_t(\theta_t(\psi_{n,t}(w)))$, we obtain by **(WGC)** that

$$\mathrm{e}^{-d_t+S_tf(\widehat{\omega})}\le\widehat{\mathbb{P}}_t(\omega_1,\dots,\omega_t)\le\mathrm{e}^{d_t+S_tf(\widehat{\omega})}.$$

This and **(WGC)** imply (A.11) with $c=2\|f\|+2\sup_t d_t/t$. Using the notation (A.12), and introducing $\widehat{b}=\theta_{|b|}(b)$, $\widehat{\xi^i}=\theta_{|\xi^i|}(\xi^i)$, $\widehat{w}^i=\theta_n(w^i)$, we have either $\theta_t(\psi_{n,t}(w))=\widehat{b}\widehat{w}^1\widehat{\xi}^1\dots\widehat{\xi}^{N-1}\widehat{w}^N$ or $\theta_t(\psi_{n,t}(w))=\widehat{w}^N\widehat{\xi}^{N-1}\dots\widehat{\xi}^1\widehat{w}^1\widehat{b}$ (see Definition 2.5). Computations similar to (A.14) show that (A.15) also holds with $\mathbb{P}$ replaced with $\widehat{\mathbb{P}}$, so that (A.10) again follows. □

By using Proposition A.6 instead Proposition 3.1, our results apply as follows.

- All the conclusions of Theorem 2.7 are valid under **(WSP)** and **(WGC)**.
- All the conclusions of Theorem 2.8 are valid under **(WSP)**, **(WGC)**, and the assumptions in Part 2 of Proposition A.6. The finiteness of $q(\alpha)$ for all $\alpha\in\mathbb{R}$ follows from (A.11).
- All the conclusions of Theorem 2.13 are valid under **(WSP)** and **(WGC)**. The estimates requiring **(UD)** in Proposition 6.3 can easily be adapted by using the following consequence of **(WGC)**: for all $w^1,w^2,\dots,w^N\in\Omega_n$ and all $\xi^1,\xi^2,\dots,\xi^{N-1}\in\Omega_{\tau_n}$, we have

$$\mathbb{P}_{Nn+(N-1)\tau_n}(w^1\xi^1w^2\xi^2\cdots w^{N-1}\xi^{N-1}w^n)\le\mathrm{e}^{h(n,t)}\prod_{i=1}^N\mathbb{P}_n(w^i),$$

  where $h(n,t):=d_{Nn+(N-1)\tau_n}+Nd_n+(N-1)\tau_n\|f\|$ satisfies

$$\lim_{n\to\infty}\limsup_{t\to\infty}t^{-1}h(n,t)=0.$$

References

[1] R. R. Bahadur and S. L. Zabell. Large deviations of the sample mean in general vector spaces. *Ann. Probab.*, 7(4):587–621, 1979.

[2] T. Benoist, N. Cuneo, V. Jakšić, and C.-A. Pillet. On entropy production of repeated quantum measurements II. In preparation.

[3] T. Benoist, V. Jakšić, C.-A. Pillet, and Y. Pautrat. On entropy production of repeated quantum measurements I. General theory. *Commun. Math. Phys.*, 357(1):77–123, Jan 2018.

[4] A. Brøndsted. Conjugate convex functions in topological vector spaces. *Mat.-Fys. Medd. Danske Vid. Selsk.*, 34(2):1–26, 1964.

[5] H. Comman. Strengthened large deviations for rational map and full-shifts, with unified proof. *Nonlinearity*, 22(6):1413–1429, 2009.

[6] N. Cuneo, V. Jakšić, V. Nersesyan, C.-A. Pillet, M. Porta, and A. Shirikyan. In preparation.

[7] N. Cuneo, V. Jakšić, C.-A. Pillet, and A. Shirikyan. What is the fluctuation theorem? In preparation.

[8] N. Cuneo, V. Jakšić, C.-A. Pillet, and A. Shirikyan. Fluctuation theorem and thermodynamic formalism. Preprint, 2017.

[9] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer–Verlag, Berlin, 2000.

[10] J.-D. Deuschel and D. W. Stroock. *Large Deviations*. Academic Press, Boston, 1989.

[11] A. Eizenberg, Y. Kifer, and B. Weiss. Large deviations for $\mathbb{Z}^d$-actions. *Commun. Math. Phys.*, 164(3):433–454, 1994.

[12] D. E. Evans and R. Høegh-Krohn. Spectral properties of positive maps on C^*-algebras. *J. London Math. Soc.*, 17(2):345–355, 1978.

[13] D. J. Evans, E. G. D. Cohen, and G. P. Morriss. Probability of second law violations in shearing steady states. *Phys. Rev. Lett.*, 71:2401–2404, 1993.

[14] D. J. Evans and D. J. Searles. Equilibrium microstates which generate second law violating steady states. *Phys. Rev. E*, 50:1645–1648, 1994.

[15] D.-J. Feng. Lyapunov exponents for products of matrices and multifractal analysis. II. General matrices. *Israel J. Math.*, 170:355–394, 2009.

[16] H. Föllmer and S. Orey. Large deviations for the empirical field of a Gibbs measure. *Ann. Probab.*, 16(3):961–977, 1988.

[17] G. Gallavotti. Reversible Anosov diffeomorphisms and large deviations. *Math. Phys. Electron. J.*, 1:Paper 1, 1–12, 1995.

[18] G. Gallavotti and E. G. D. Cohen. Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.*, 74:2694–2697, 1995.

[19] G. Gallavotti and E. G. D. Cohen. Dynamical ensembles in stationary states. *J. Stat. Phys.*, 80(5-6):931–970, 1995.

[20] V. Jakšić, Y. Ogata, C.-A. Pillet, and R. Seiringer. Quantum hypothesis testing and non-equilibrium statistical mechanics. *Rev. Math. Phys.*, 24(6):1230002, 2012.

[21] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press, Cambridge, 1995.

[22] Y. Kifer. Large deviations in dynamical systems and stochastic processes. *Trans. Amer. Math. Soc.*, 321(2):505–524, 1990.

[23] D. Kwietniak, M. Lacka, and P. Oprocha. A panorama of specification-like properties and their consequences. In *Dynamics and numbers*, volume 669 of *Contemp. Math.*, pages 155–186. Amer. Math. Soc., Providence, RI, 2016.

[24] O. Lanford. Entropy and equilibrium states in classical statistical mechanics. In *Lecture Notes in Physics*, volume 20, pages 1–113. Springer, New York, 1973.

[25] J. L. Lebowitz and H. Spohn. A Gallavotti-Cohen-type symmetry in the large deviation functional for stochastic dynamics. *J. Stat. Phys.*, 95(1-2):333–365, 1999.

[26] J. T. Lewis and C.-E. Pfister. Thermodynamic probability theory: some aspects of large deviations. *Russian Math. Surveys*, 50(2):279–317, 1995.

[27] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan. Large deviations and the thermodynamic formalism: a new proof of the equivalence of ensembles. In M. Fannes, C. Maes, and A. Verbeure, editors, *On Three Levels. Micro-, Meso-, and Macro-Approaches in Physics*, pages 183–192. Springer, New York, 1994.

[28] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan. Entropy, concentration of probability and conditional limit theorems. *Markov Process. Related Fields*, 1(3):319–386, 1995.

[29] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.

[30] Y. Ogata and L. Rey-Bellet. Ruelle-Lanford functions and large deviations for asymptotically decoupled quantum systems. *Rev. Math. Phys.*, 23(2):211–232, 2011.

[31] E. Olivier and A. Thomas. Weak gibbs property and system of numeration. Preprint, 2006.

[32] E. Olivier and A. Thomas. Infinite convolution of Bernoulli measures, PV numbers and related problems in the dynamics of fractal geometry. In Y. Lacroix, P. Liardet, and J.-P. Thouvenot, editors, *École de Théorie Ergodique*, volume 20 of *Sémin. Congr.*, pages 163–192. Soc. Math. France, Paris, 2010.

[33] S. Orey and S. Pelikan. Large deviation principles for stationary processes. *Ann. Probab.*, 16(4):1481–1495, 1988.

[34] C.-E. Pfister. Thermodynamical aspects of classical lattice systems. In *In and out of equilibrium (Mambucaba, 2000)*, volume 51 of *Progr. Probab.*, pages 393–472. Birkhäuser, Boston, MA, 2002.

[35] C.-E. Pfister and W. Sullivan. Weak Gibbs measures and large deviations. *Nonlinearity*, 1(31):49–53, 2018.

[36] C.-E. Pfister and W. G. Sullivan. Billingsley dimension on shift spaces. *Nonlinearity*, 16(2):661–682, 2003.

[37] C.-E. Pfister and W. G. Sullivan. Large deviations estimates for dynamical systems without the specification property. Applications to the β-shifts. *Nonlinearity*, 18(1):237–261, 2005.

[38] D. Ruelle. Correlation functionals. *J. Math. Phys.*, 6:201–220, 1965.

[39] D. Ruelle. *Thermodynamic Formalism.* Cambridge University Press, Cambridge, 2004.

[40] D. J. Thompson. Irregular sets, the β-transformation and the almost specification property. *Trans. Amer. Math. Soc.*, 364(10):5395–5414, 2012.

[41] A. C. D. van Enter, R. Fernández, and A. D. Sokal. 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.

[42] P. Varandas. Non-uniform specification and large deviations for weak Gibbs measures. *J. Stat. Phys.*, 146(2):330–358, 2012.

[43] P. Varandas and Y. Zhao. Weak specification properties and large deviations for non-additive potentials. *Ergod. Theory Dyn. Syst.*, 35(03):968–993, 2015.

[44] L.-S. Young. Large deviations in dynamical systems. *Trans. Amer. Math. Soc.*, 318(2):525–543, 1990.