A Γ-convergence approach to large deviations

MAURO MARIANI

Abstract. A rigorous connection between large deviations theory and Γ-convergence is established. Applications include representations formulas for rate functions, a contraction principle for measurable maps, a large deviations principle for coupled systems and a second order Sanov theorem.

Mathematics Subject Classification (2010): 60F10 (primary); 49N99 (secondary).

1. Introduction

Let X be a Polish space, that is a completely metrizable, separable topological space. The space $\mathcal{P}(X)$ of Borel probability measures on X is a Polish space as well, if equipped with the so-called *narrow* (otherwise called *weak*) topology. Such a topology enjoys several characterizations, see [12, Theorem 3.1.5]. A sequence (μ_n) in $\mathcal{P}(X)$ converges narrowly to μ iff $\overline{\lim}_n \mu_n(C) \leq \mu(C)$ or $\underline{\lim}_n \mu_n(O) \geq \mu(O)$ for all $C \subset X$ closed and $O \subset X$ open, or equivalently iff the integrals of bounded continuous functions converge.

A *Large Deviations principle* (LDP) for (μ_n) on X is then classically defined as an exponential version of the inequalities on closed and open sets stated above for the narrow convergence; and the Brycs-Varadhan theorem [5, Chapter 4.4] can be regarded as a Large Deviations' (LD) analog of the characterization of narrow convergence by the convergence of integrals of continuous bounded functions.

In this paper, we further extend the analogies between narrow convergence and LD to other characterizations. At least when X is compact, it is easy to see that the narrow convergence $\mu_n \to \mu$ is equivalent to the Γ-convergence of the relative entropy functional $H(\cdot|\mu_n)$ to $H(\cdot|\mu)$ and also to the Γ-convergence of the maps $K \mapsto -\log \mu_n(K)$ to $K \mapsto -\log \mu(K)$, where the compact subsets K of X are equipped with the Hausdorff topology (which indeed coincides with the Kuratowski topology on compact sets). In Section 3 we provide the LD analogs of these state-

ments, proving in particular that LD is also a notion of convergence in a metric space $\mathcal{W}(X)$, containing both probability measures and functionals. Thus convergence of measures to measures in $\mathcal{W}(X)$ is equivalent to the narrow convergence, convergence of functionals to functionals is equivalent to Γ-convergence, and convergence of measures to functionals is indeed LD, see Theorem 5.1 for a precise statement. It is worth to remark that various approaches to LD are possible by the means of variational analysis of the relative entropy functional [3], the one in this paper being indeed inspired by the techniques in [6,10].

In Section 4 we apply the results in Section 3 to get some general properties of LDP. Proposition 4.1 gives some explicit representations of the LD rate functionals, that generalize the so-called Laplace-Varadhan method for proving LDPs. In Proposition 4.4 a version of a so-called contraction principle is provided for measurable (not just continuous) contraction maps. In Theorem 4.6 we give sufficient conditions to recover a LDP for a coupled system of metric random variables, from the LDPs for the (independent) components of an associated system with frozen variables. In Theorem 6.1, we apply the results in Section 3-4 to provide a second order version of the Sanov theorem for triangular arrays of i.i.d. random variables whose law also satisfies a LDP (see the discussion in Section 6 for applications).

2. Preliminaries

In this section we recall the basic notions concerning Γ-convergence and LD. Hereafter $\mathfrak{B}(X)$ denotes be the Borel σ-algebra on the Polish space X and $\mathcal{P}(X)$ the set of Borel probability measures on X. For $\mu \in \mathcal{P}(X)$ and f a μ-integrable function on X, $\mu(f)$ will denote the integral of f with respect to μ. $\mathcal{P}(X)$ is hereafter equipped with the narrow topology, namely the weakest topology such that the maps $\mathcal{P}(X) \ni \mu \mapsto \mu(f) \in \mathbb{R}$ are continuous, for all $f \in C_{\mathrm{b}}(X)$.

We also let $\mathfrak{K}(X)$ be the collection of compact subsets of X, equipped with the Hausdorff topology[1].

---

[1] If X is compact, the Hausdorff topology on $\mathfrak{K}(X)$ coincides with the Kuratowski topology, see [11, Chapter 4]. The latter is often considered in the theory of Γ-convergence, due its equivalence with the Kuratowski convergence of epigraphs, see, *e.g.*, [2, Theorem 4.16]. Here we use a slightly different construction, as we lift functions and measures on X to functionals on $\mathfrak{K}(X)$. The Hausdorff topology comes more handy, since $\mathfrak{K}(X)$ is itself Polish. The price to pay is that possibly some of the statements in the paper would extend to closed sets, not just compact sets, if the Kuratowski topology would be used. However, as long as one sticks with exponentially tight families of probabilities (or uniformly coercive functions) the two notions of convergence are equivalent, so that the price to pay for working with the Hausdorff topology is negligible for all the applications discussed here.

Namely, fixed a compatible distance d on X, define $d_H \colon \mathfrak{K}(X) \times \mathfrak{K}(X) \to [0, +\infty[$ as

$$d_H(K, K') := \inf\{\varepsilon > 0, \ K \subset K'^\varepsilon, \ K' \subset K^\varepsilon\}$$

where, for $A \in \mathfrak{B}(X)$, A^ε denotes the ε-enlargement of A with respect to the distance d. As well known, d_H defines a distance on $\mathfrak{K}(X)$, and the associated topology τ_H does not depend on the choice of the compatible distance d. Moreover, $(\mathfrak{K}(X), \tau_H)$ is a Polish space, see [11, Chapter 4], and it is understood that $\mathfrak{K}(X)$ is equipped with such a topology in the following.

2.1. Γ-convergence

Γ-convergence is the relevant notion of convergence for functionals, whenever problems related to minima and minimizers are investigated.

Definition 2.1. A functional $I \colon X \to [0, +\infty]$ is said *lower semicontinuous* if for each $\ell \geq 0$ the set $\{x \in X \ : \ I(x) \leq \ell\}$ is closed. I is said *coercive* if for each $\ell \geq 0$ the set $\{x \in X \ : \ I(x) \leq \ell\}$ is precompact.

Let (I_n) be a sequence of functionals $I_n \colon X \to [0, +\infty]$. (I_n) is said *equicoercive* on X if for each $\ell > 0$, $\cup_n \{x \in X \ : \ I_n(x) \leq \ell\}$ is precompact.

Definition 2.2. The Γ-liminf (also denoted Γ-$\underline{\lim}$) and Γ-limsup (also denoted Γ-$\overline{\lim}$) of a sequence (I_n) of functionals $I_n \colon X \to [0, +\infty]$ are two functionals on X defined as follows. For $x \in X$

$$\left(\Gamma\text{-}\underline{\lim_n} I_n\right)(x) := \inf\left\{\underline{\lim_n} I_n(x_n), \ (x_n) \text{ sequence in } X \text{ such that } x_n \to x\right\}$$

$$\left(\Gamma\text{-}\overline{\lim_n} I_n\right)(x) := \inf\left\{\overline{\lim_n} I_n(x_n), \ (x_n) \text{ sequence in } X \text{ such that } x_n \to x\right\}.$$

Whenever Γ-$\underline{\lim} I_n = $ Γ-$\overline{\lim} I_n = I$, (I_n) is said to Γ-converge to I in X, and I is called the Γ-limit (also denoted Γ-lim) of (I_n).

2.2. Large deviations

Hereafter (μ_n) is a sequence in $\mathcal{P}(X)$ and (a_n) is a sequence of positive reals such that $\lim_n a_n = +\infty$.

Definition 2.3. The sequence (μ_n) is *exponentially tight* with speed (a_n) if

$$\inf_{K \subset X, \text{ compact}} \overline{\lim_n} \frac{1}{a_n} \log \mu_n(K^c) = -\infty.$$

Definition 2.4. Let $I \colon X \to [0, +\infty]$ be a lower semicontinuous functional. Then (μ_n) satisfies

- A *LD lower bound* with speed (a_n) and rate I, if for each open set $O \subset X$

$$\underline{\lim_n} \frac{1}{a_n} \log \mu_n(O) \geq - \inf_{x \in O} I(x);$$

- A *LD weak upper bound* with speed (a_n) and rate I, if for each compact set $K \subset X$

$$\varlimsup_n \frac{1}{a_n} \log \mu_n(K) \leq - \inf_{x \in K} I(x);$$

- A *LD upper bound* with speed (a_n) and rate I, if for each closed set $C \subset X$

$$\varlimsup_n \frac{1}{a_n} \log \mu_n(C) \leq - \inf_{x \in C} I(x).$$

(μ_n) satisfies a *(weak) LDP* if both the lower and (weak) upper bounds hold with same rate and speed.

It is immediate to check that if (μ_n) is exponentially tight and satisfies a weak LD upper bound, then it satisfies a LD upper bound.

2.3. Relative entropy

Given μ, $\nu \in \mathcal{P}(X)$ and $\mathfrak{F} \subset \mathfrak{B}(X)$ a σ-algebra, the relative entropy of ν with respect to μ on $\mathfrak{F}$ is defined as

$$H_{\mathfrak{F}}(\nu|\mu) := \sup_{\varphi} \big\{ \nu(\varphi) - \log \mu(e^{\varphi}) \big\}, \tag{2.1}$$

where the supremum runs over the bounded $\mathfrak{F}$-measurable functions φ on X. For a fixed μ, $H_{\mathfrak{F}}(\cdot|\mu)$ is a positive, convex functional on $\mathcal{P}(X)$. If $\mathfrak{F} = \mathfrak{B}(X)$, the subindex $\mathfrak{F}$ will be dropped hereafter. In such a case, $H(\cdot|\mu)$ is also lower semicontinuous and coercive on $\mathcal{P}(X)$.

2.4. Regular set-maps

If $a > 0$, $\mu \in \mathcal{P}(X)$ and $I \colon X \to [0, +\infty]$ a lower semicontinuous functional, define the set-maps $l_{a,\mu}$, $l_I \colon \mathfrak{K}(X) \to [0, +\infty]$ as

$$l_{a,\mu}(K) = -\frac{1}{a} \log \mu(K) \tag{2.2}$$
$$l_I(K) = \inf_{x \in K} I(x). \tag{2.3}$$

Since probability measures are regular on Polish spaces, it is easy to check that $l_{a,\mu}$ is lower semicontinuous on $\mathfrak{K}(X)$, while the lower semicontinuity of I implies that l_I is lower semicontinuous as well.

3. Large deviations and Γ-convergence

The equivalence of probabilistic statements concerning LD (labeled P), Γ-convergence statements concerning relative entropies (labeled H) and set-maps (labeled L) is established in this section.

An equivalent formulation of narrow convergence of probability measures is first introduced in Subsection 3.1. Although only needed in proofs to appear later in the paper, it gives an easy example of the ideas concerning the analogous LD statements in Subsection 3.2. Proofs are provided in Subsection 3.3.

3.1. Weak convergence and relative entropy

Let (μ_n) be a sequence in $\mathcal{P}(X)$, and define $H_n \colon \mathcal{P}(X) \to [0, +\infty]$ as

$$H_n(\nu) := H(\nu|\mu_n).$$

The parameter $a > 0$ has no special role in the next two propositions, one could fix $a = 1$. Yet, it will become relevant when LD are considered.

Proposition 3.1. *The following are equivalent:*

(P) (μ_n) *is tight in* $\mathcal{P}(X)$;
(H) (H_n) *is equicoercive on* $\mathcal{P}(X)$;
(L) *For* $a > 0$, (l_{a,μ_n}) *is equicoercive on* $\mathfrak{K}(X)$.

Proposition 3.2. *The following are equivalent:*

(P1) $\mu_n \to \mu$ *in* $\mathcal{P}(X)$;
(P2) *For each sequence* (φ_n) *of Borel measurable functions* $\varphi_n \colon X \to \bar{\mathbb{R}}$ *bounded from below*
$$\varliminf_n \mu_n(\varphi_n) \geq \mu(\Gamma\text{-}\varliminf_n \varphi_n);$$

(H) $(\Gamma\text{-}\lim_n H_n)(\nu) = H(\nu|\mu)$;
(L1) *For* $a > 0$, $\Gamma\text{-}\lim_n l_{a,\mu_n} = l_{a,\mu}$;
(L2) *For* $a > 0$, $\Gamma\text{-}\varlimsup_n l_{a,\mu_n} \leq l_{a,\mu}$;
(L3) *For* $a > 0$, $\Gamma\text{-}\varliminf_n l_{a,\mu_n} \geq l_{a,\mu}$.

3.2. Large deviations and relative entropy

In this section the LD analogs of Proposition 3.1 and Proposition 3.2 are stated. Hereafter $\mathbf{a} = (a_n)$ is a sequence of strictly positive real numbers such that $\lim_n a_n = +\infty$, (μ_n) is a sequence in $\mathcal{P}(X)$ and $I \colon X \to [0, +\infty]$ a measurable function. Define $H_n^{\mathbf{a}} \colon \mathcal{P}(X) \to [0, +\infty]$ as

$$H_n^{\mathbf{a}}(\nu) := \tfrac{1}{a_n} H(\nu|\mu_n)$$

and, recalling (2.2), $l_n^{\mathbf{a}} \colon \mathfrak{K}(X) \to [0, +\infty]$ as

$$l_n^{\mathbf{a}} = l_{a_n, \mu_n}.$$

Theorem 3.3. *The following are equivalent:*

(P) (μ_n) *is exponentially tight with speed* (a_n);
(H) $(H_n^{\mathbf{a}})$ *is equicoercive;*
(L) $(l_n^{\mathbf{a}})$ *is equicoercive.*

Theorem 3.4. *The following are equivalent:*

(P1) *(μ_n) satisfies a LD lower bound with speed (a_n) and rate I;*

(P2) *For each sequence (φ_n) of measurable maps $\varphi_n \colon X \to \overline{\mathbb{R}}$*

$$\varliminf_n \frac{1}{a_n} \log \mu_n\big(\exp(a_n\varphi_n)\big) \geq \sup_{x \in X} \left\{ \big(\Gamma\text{-}\varliminf_n \varphi_n\big)(x) - I(x) \right\}$$

where one understands $\big(\Gamma\text{-}\varliminf_n \varphi_n\big)(x) - I(x) = -\infty$ whenever $I(x) = +\infty$;

(H1) *For each $x \in X$, $\big(\Gamma\text{-}\varlimsup_n H_n^{\mathbf{a}}\big)(\delta_x) \leq I(x)$, where $\delta_x \in \mathcal{P}(X)$ is the Dirac mass concentrated at x;*

(H2) *For each $v \in \mathcal{P}(X)$, $\big(\Gamma\text{-}\varlimsup_n H_n^{\mathbf{a}}\big)(v) \leq v(I)$.*

If I is lower semicontinuous, the above statements are also equivalent to

(L) *$\Gamma\text{-}\varlimsup_n l_n^{\mathbf{a}} \leq l_I$.*

Theorem 3.5. *Assume that I is lower semicontinuous. Then the following are equivalent:*

(P1) *(μ_n) satisfies a LD weak upper bound with speed (a_n) and rate I;*

(P2) *For each sequence (φ_n) of measurable maps $\varphi_n \colon X \to \overline{\mathbb{R}}$ bounded from below and such that*

$$\sup_{K \subset X \text{ compact}} \varlimsup_n \frac{\mu_n\big(\mathbf{1}_{K^c} \exp(-a_n\varphi_n)\big)}{\mu_n\big(\exp(-a_n\varphi_n)\big)} = 0 \tag{3.1}$$

the following inequality holds

$$\varlimsup_n \frac{1}{a_n} \log \mu_n\big(\exp(-a_n\varphi_n)\big) \leq \sup_{x \in X} \left\{ - \big(\Gamma\text{-}\varliminf_n \varphi_n\big)(x) - I(x) \right\}$$

where $-\big(\Gamma\text{-}\varliminf_n \varphi_n\big)(x) - I(x) := -\infty$ whenever $I(x) = +\infty$;

(H1) *For each $x \in X$, $\big(\Gamma\text{-}\varliminf_n H_n^{\mathbf{a}}\big)(\delta_x) \geq I(x)$;*

(H2) *For each $v \in \mathcal{P}(X)$, $\big(\Gamma\text{-}\varliminf_n H_n^{\mathbf{a}}\big)(v) \geq v(I)$;*

(L) *$\Gamma\text{-}\varliminf_n l_n^{\mathbf{a}} \geq l_I$.*

Assume furthermore that (μ_n) satisfies the equivalent conditions of Theorem 3.3. Then the above statements are also equivalent to:

(P3) *μ_n satisfies a LD upper bound with speed (a_n) and rate I;*

(P4) *For each sequence (φ_n) of measurable maps $\varphi_n \colon X \to \overline{\mathbb{R}}$, bounded from below it holds*

$$\varlimsup_n \frac{1}{a_n} \log \mu_n\big(\exp(-a_n\varphi_n)\big) \leq \sup_{x \in X} \left\{ - \big(\Gamma\text{-}\varliminf_n \varphi_n\big)(x) - I(x) \right\}$$

where $-\big(\Gamma\text{-}\varliminf_n \varphi_n\big)(x) - I(x) := -\infty$ whenever $I(x) = +\infty$.

3.3. Proofs for Section 3

We start by recalling some basic facts concerning Γ-convergence theory and relative entropies. The claims in the following three remarks are easy to prove.

Remark 3.6. The Γ-liminf and Γ-limsup of (I_n) are lower semicontinuous functionals, coercive if (I_n) is equicoercive.

Let $J \colon X \to [0, +\infty]$. Then:

(i) If for each sequence $x_n \to x$, $\underline{\lim}_n I_n(x_n) \geq J(x)$, then $J \leq \Gamma\text{-}\underline{\lim}_n I_n$;
(ii) If there exists a sequence $x_n \to x$ such that $\overline{\lim}_n I_n(x_n) \leq J(x)$, then $J \geq \Gamma\text{-}\overline{\lim}_n I_n$.

The $\Gamma\text{-}\underline{\lim}_n I_n$ and $\Gamma\text{-}\overline{\lim}_n I_n$ are, respectively, the smallest and the largest lower semicontinuous functionals on X satisfying conditions (i) and (ii) above.

Moreover for each $x \in X$, open set $O \subset X$ and compact set $K \subset X$ the following hold:

(a) There exists a sequence $x_n \to x$ such that

$$\overline{\lim_n} I_n(x_n) \leq \left(\Gamma\text{-}\overline{\lim_n} I_n \right)(x),$$

and

$$\overline{\lim_n} \inf_{y \in O} I_n(y) \leq \inf_{y \in O} \left(\Gamma\text{-}\underline{\lim_n} I_n \right)(x);$$

(b) For each sequence $x_n \to x$

$$\underline{\lim_n} I_n(x_n) \geq \left(\Gamma\text{-}\underline{\lim_n} I_n \right)(x)$$

and

$$\underline{\lim_n} \inf_{y \in K} I_n(y) \leq \inf_{y \in K} \left(\Gamma\text{-}\underline{\lim_n} I_n \right)(y).$$

Additionally, if I_n is equicoercive then for each closed set $C \subset X$

$$\underline{\lim_n} \inf_{y \in C} I_n(y) \leq \inf_{y \in C} \left(\Gamma\text{-}\underline{\lim_n} I_n \right)(x).$$

Hereafter for $\mu \in \mathcal{P}(X)$ and A a Borel subset of X such that $\mu(A) > 0$, $\mu^A \in \mathcal{P}(X)$ denotes the probability measure obtained by conditioning μ on A.

Remark 3.7. If $\mathfrak{G} \subset \mathfrak{F}$ then

$$H_{\mathfrak{G}}(\nu|\mu) \leq H_{\mathfrak{F}}(\nu|\mu). \tag{3.2}$$

If $\mathfrak{F}$ is the Borel σ-algebra of X, the supremum over φ in (2.1) can equivalently run over the test functions $\varphi \in L_1(X, d\nu)$, or equivalently over $\varphi \in C_c(X)$, or

equivalently over the set of measurable functions φ taking only a finite number of values. Moreover

$$H(\nu|\mu) = \begin{cases} \int_X \mu(dx)\, \frac{d\nu}{d\mu}(x) \log \frac{d\nu}{d\mu}(x) & \text{if } \nu \ll \mu \\ +\infty & \text{otherwise.} \end{cases} \tag{3.3}$$

In particular, for $\nu(A) > 0$, $\frac{d\nu^A}{d\mu} = \frac{\mathbf{1}_A}{\nu(A)} \frac{d\nu}{d\mu}$, so that

$$
\begin{aligned}
H(\nu^A|\mu) &= -\log \nu(A) + \tfrac{1}{\nu(A)} \int_A d\nu(x)\, \log \frac{d\nu}{d\mu}(x) \\
&\le -\log \nu(A) + \frac{1}{\nu(A)} \left(H(\nu|\mu) + (1 - \nu(A)) \log \left(\tfrac{1-\nu(A)}{1-\mu(A)} \right) \right) \quad (3.4) \\
&\le -\log \nu(A) + \frac{1}{\nu(A)} H(\nu|\mu) + 1 - \tfrac{\mu(A)}{\nu(A)}
\end{aligned}
$$

where the first inequality follows by taking φ constant on A^c in the definition (2.1).

Let Y be also a Polish space, $\lambda \in \mathcal{P}(Y)$, $\theta \colon X \to Y$ measurable and $\mathfrak{F}_\theta$ the associated σ-algebra. If $\lambda \in \mathcal{P}(Y)$ then

$$H\left(\lambda|\mu \circ \theta^{-1}\right) = \inf_{\nu:\, \nu \circ \theta^{-1} = \lambda} H(\nu|\mu) = H_{\mathfrak{F}_\theta}(\nu|\mu) \quad \text{for all } \nu \,:\, \nu \circ \theta^{-1} = \lambda. \tag{3.5}$$

If $H(\lambda|\mu \circ \theta^{-1}) < +\infty$ then the infimum in (3.5) is attained, namely:

$$
\begin{aligned}
H\left(\lambda|\mu \circ \theta^{-1}\right) &= H(\bar\nu|\mu) \\
\bar\nu(dx) &= \mu(dx)\, \mathbb{E}^\nu \left(\frac{d\nu}{d\mu} \Big| \mathfrak{F}_\theta \right)(x) \quad \text{for all } \nu \,:\, \nu \circ \theta^{-1} = \lambda.
\end{aligned} \tag{3.6}
$$

Remark 3.8. If $\mathfrak{G} = \sigma\left((E^i)_{i=0}^N\right)$ is a σ-algebra generated by a finite partition of X, then

$$H_{\mathfrak{G}}(\nu|\mu) = \sum_{i=0}^N \nu(E^i) \log \frac{\nu(E^i)}{\mu(E^i)},$$

where we understand $\nu(A) \log \frac{\nu(A)}{\mu(A)} = 0$ whenever $\nu(A) = 0$ and $\nu(A) \log \frac{\nu(A)}{\mu(A)} = +\infty$ if $\mu(A) = 0$ but $\nu(A) > 0$.

Moreover taking $\varphi = \log(1 + \mu(A))\mathbf{1}_A$ in (2.1) one obtains

$$\nu(A) \le \frac{\log 2 + H(\nu|\mu)}{\log \left(1 + \frac{1}{\mu(A)}\right)} \tag{3.7}$$

whenever $H(\nu|\mu) < +\infty$ and $\mu(A) > 0$.

Remark 3.9. Let μ, $\nu \in \mathcal{P}(X)$, and $(K_\ell)_{\ell \in \mathbb{N}}$ a sequence of compact subsets of X such that $\lim_\ell \mu(K_\ell) = 1$. Then for each $\delta > 0$, $\ell \in \mathbb{N}$ there exists a finite family $(E^i_{\delta,\ell})^{N_{\delta,\ell}}_{i=1}$ of Borel subsets of X such that:

(i) $\cup_i E^i_{\delta,\ell} \supset K_\ell$ and $E^i_{\delta,\ell} \cap E^{i'}_{\delta,\ell} = \emptyset$ if $i \neq i'$;

(ii) diameter$(E^i_{\delta,\ell}) \leq \delta$, for $i = 1, \ldots, N_{\delta,\ell}$;

(iii) $\mu(\partial E^i_{\delta,\ell}) = \nu(\partial E^i_{\delta,\ell}) = 0$, for $i = 1, \ldots, N_{\delta,\ell}$;

(iv) Each $E^i_{\delta,\ell}$ has nonempty interior.

Set $E^0_{\delta,\ell} = X \setminus \cup_{i \geq 1} E^i_{\delta,\ell}$. One can assume the following with no loss of generality:

(v) The partition $(E^i_{\delta,\ell})^{N_{\delta,\ell}}_{i=0}$ is finer than $(E^i_{\delta',\ell'})^{N_{\delta',\ell'}}_{i=0}$ if $\delta \leq \delta'$ and $\ell \geq \ell'$.

Moreover, if $\mathfrak{G}_{\delta,\ell}$ is the σ-algebra generated by $(E^i_{\delta,\ell})$

$$\lim_\ell \lim_\delta H_{\mathfrak{G}_{\delta,\ell}}(\nu|\mu) = H(\nu|\mu).$$

Proof. Fix δ, ℓ and take a finite cover of K_ℓ with open balls $B_{\delta/2}(x_i)$ of radius $\delta/2$ and centered at $x_i \in K_\ell$, $i = 1, \ldots, N_{\ell,\delta}$. Take $r > 0$ such that $r \leq \delta/2$ and $r \leq$ distance(x_i, x_j) for all $i \neq j$. By σ-additivity of μ and ν, there exists $\delta' \in]\delta/2, \delta/2 + r[$ such that $\mu(\partial B_{\delta'}(x_i)) = \nu(\partial B_{\delta'}(x_i)) = 0$ for all i. Then take

$$\begin{cases} E^1_{\delta,\ell} = B_{\delta'}(x_1) \\ E^i_{\delta,\ell} = B_{\delta'}(x_i) \setminus \cup_{j<i} E^j_{\delta,\ell} & \text{for } i > 1. \end{cases}$$

It is immediate to check that $(E^i_{\delta,\ell})$ satisfies (i)-(iv); and by a refining procedure one gets the $E^i_{\delta,\ell}$ to satisfy (v) as well. The convergence of the relative entropies is a consequence of (i)-(v). ☐

Next we turn to the proofs of the statements in Subsections 3.1 and 3.2.

Proof of Proposition 3.1. (H)⇒(P). μ_n is in the 0 sublevel set of H_n, and thus (μ_n) is precompact (and tight) by the definition of equicoercivity.

(P) ⇒ (H). Let (ν_n) be a sequence in $\mathcal{P}(X)$ such that $\overline{\lim}_n H(\nu_n|\mu_n) < +\infty$. Since μ_n is tight, there exists an increasing sequence (K_ℓ) of compacts such that $\lim_\ell \overline{\lim}_n \mu_n(K^c_\ell) = 0$. Since $H(\nu_n|\mu_n)$ is uniformly bounded, the application of (3.7) with $A = K^c_\ell$ yields $\lim_\ell \overline{\lim}_n \nu_n(K^c_\ell) = 0$. Namely (ν_n) is tight.

(P) ⇔ (L). It is trivial. ☐

Proof of Proposition 3.2. (P1)⇒(P2). Assume $\mu_n \to \mu$ and let $(E^i_{\delta,\ell})^{N_{\delta,\ell}}_{i=0}$ be as in Remark 3.9 with $\nu = \mu$. Let (φ_n) be as in the statement (P2), and define

$\varphi_{n,\delta,\ell}$, $\varphi_{\delta,\ell} \colon X \to \bar{\mathbb{R}}$ by

$$\varphi_{n,\delta,\ell}(x) = \inf_{y \in E^i_{\delta,\ell}} \varphi_n(y) \qquad \text{if } x \in E^i_{\delta,\ell}$$

$$\varphi_{\delta,\ell}(x) = \varliminf_n \varphi_{n,\delta,\ell}(x).$$

Note that by Remark 3.9-(iii) and (v) one has:

$$\mu\left(\cup_{\ell,\delta>0} \cup_{i=1}^{N_{\delta,\ell}} \partial E^i_{\delta,\ell}\right) = \lim_\ell \lim_\delta \sum_{i=0}^{N_{\delta,\ell}} \mu(E^i_{\delta,\ell}) = 0.$$

On the other hand, if $x \notin \cup_{\ell,\delta>0} \cup_{i=1}^{N_{\delta,\ell}} \partial E^i_{\delta,\ell}$ it is easy to check that

$$\lim_\ell \lim_\delta \varphi_{\delta,\ell}(x) = (\Gamma\text{-}\varliminf_n \varphi_n)(x),$$

the limit being monotone increasing by Remark 3.9-(v). Thus $\lim_\ell \lim_\delta \varphi_{\delta,\ell} = (\Gamma\text{-}\varliminf_n \varphi_n)$ μ-a.e., and by monotone convergence

$$\mu(\Gamma\text{-}\varliminf_n \varphi_n) = \lim_\ell \lim_\delta \mu(\varphi_{\delta,\ell}) = \lim_\ell \lim_\delta \sum_{i=0}^{N_{\delta,\ell}} \left[\mu(E^i_{\delta,\ell}) \varliminf_n \inf_{y \in E^i_{\delta,\ell}} \varphi_{n,\delta,\ell}(y)\right]$$

$$= \lim_\ell \lim_\delta \varliminf_n \sum_{i=0}^{N_{\delta,\ell}} \left[\mu_n(E^i_{\delta,\ell}) \inf_{y \in E^i_{\delta,\ell}} \varphi_{n,\delta,\ell}(y)\right] \le \varliminf_n \mu_n(\varphi_n),$$

where last equality follows from Remark 3.9-(iii).

(P2) $\Rightarrow$ (H). By (P2), $\mu_n(\varphi) \to \mu(\varphi)$ for each $\varphi \in C_b(X)$. Let now $\nu \in \mathcal{P}(X)$ and let (ν_n) be an arbitrary sequence in $\mathcal{P}(X)$ such that $\nu_n \to \nu$. Then

$$\varliminf_n H(\nu_n|\mu_n) = \varliminf_n \sup_{\varphi \in C_b(X)} \left\{\nu_n(\varphi) - \log\mu_n(e^\varphi)\right\}$$

$$\ge \sup_{\varphi \in C_b(X)} \varliminf_n \left\{\nu_n(\varphi) - \log\mu_n(e^\varphi)\right\}$$

$$= \sup_{\varphi \in C_b(X)} \left\{\nu(\varphi) - \log\mu(e^\varphi)\right\} = H(\nu|\mu).$$

Namely $\varliminf_n H_n(\nu_n) \ge H(\nu)$, and thus the Γ-liminf inequality holds. It is enough to prove the Γ-limsup inequality for ν such that $H(\nu|\mu) < +\infty$. In particular ν is absolutely continuous with respect to μ. Let $(E^i_{\delta,\ell})_{i=0}^{N_{\delta,\ell}}$ be as in Remark 3.9. Fix δ, $\ell > 0$, and for n large enough define the probability $\nu_{n,\delta,\ell} \in \mathcal{P}(X)$ as

$$\nu_{n,\delta,\ell}(A) = \sum_{i=0}^{N_{\delta,\ell}} \nu(E^i_{\delta,\ell}) \frac{\mu_n(A \cap E^i_{\delta,\ell})}{\mu_n(E^i_{\delta,\ell})};$$

$\nu_{n,\delta,\ell}$ is well defined since $\nu(E^i_{\delta,\ell}) = 0$ whenever $\mu_n(E^i_{\delta,\ell}) = 0$ for n large enough.

Then

$$\lim_{\ell} \lim_{\delta} \lim_{n} \nu_{n,\delta,\ell} = \nu.$$

On the other hand, by explicit calculation, $H(\nu_{n,\delta,\ell}|\mu_n) = H_{\mathfrak{G}_{\delta,\ell}}(\nu|\mu_n)$, and recalling that the sets $E^i_{\delta,\sigma}$ are μ- and ν-regular

$$\overline{\lim_{n}} H(\nu_{n,\delta,\ell}|\mu_n) = \overline{\lim_{n}} H_{\mathfrak{G}_{\delta,\ell}}(\nu|\mu_n) = \overline{\lim_{n}} \sum_{i=0}^{N_{\delta,\ell}} \nu(E^i_{\delta,\ell}) \log \frac{\nu(E^i_{\delta,\ell})}{\mu_n(E^i_{\delta,\ell})}$$

$$= \sum_{i=0}^{N_{\delta,\ell}} \nu(E^i_{\delta,\ell}) \log \frac{\nu(E^i_{\delta,\ell})}{\mu(E^i_{\delta,\ell})} = H_{\mathfrak{G}_{\delta,\ell}}(\nu|\mu) \leq H(\nu|\mu).$$

Thus there exist sequences $(\delta_n), (\ell_n)$ such that $\nu_n := \nu_{n,\delta_n,\ell_n} \to \nu$ and $\overline{\lim}_n H(\nu_n|\mu_n) \leq H(\nu|\mu)$.

(H) $\Rightarrow$ (P1). μ_n is the unique minimizer of H_n, and μ is the unique minimizer of H. Since converging sequences of minimizers converge to minimizers of the Γ-limit (see [2, Proposition 7.18]), one is left to show that (μ_n) is precompact in $\mathcal{P}(X)$, namely that

$$\sup_{K \in \mathcal{K}(X)} \underline{\lim_{n}} \mu_n(K) = 1. \tag{3.8}$$

(H) implies that there exists a sequence (ν_n) converging to μ such that $\overline{\lim}_n H(\nu_n|\mu_n) \leq H(\mu|\mu)$; so that, in view of the tightness of (ν_n)

$$\lim_{n} H(\nu_n|\mu_n) = 0, \qquad \sup_{K \in \mathcal{K}(X)} \underline{\lim_{n}} \nu_n(K) = 1. \tag{3.9}$$

Reversing the inequality in (3.7) (with $A = K$), one gets for each $K \in \mathcal{K}(X)$

$$\mu_n(K) \geq \frac{1}{2^{\frac{1}{\nu_n(K)}} \exp[H(\nu_n|\mu_n)] - 1}.$$

Taking the liminf in n and the supremum over $K \in \mathcal{K}(X)$, one gets (3.8) by (3.9).

(P1) $\Rightarrow$ (L1). Fix $\varepsilon > 0$, $K \in \mathfrak{K}(X)$, and let K^ε be the open ε-enlargement of K with respect to any compatible metric on X. For each n, take a compact set $K_{n,\varepsilon}$ such that $K \subset K_{n,\varepsilon} \subset K^\varepsilon$ and $\mu_n(K_{n,\varepsilon}) \geq \mu_n(K^\varepsilon) - \varepsilon$. Then by (P1)

$$\underline{\lim_{\varepsilon \downarrow 0}} \underline{\lim_{n}} \mu_n(K_{n,\varepsilon}) \geq \underline{\lim_{\varepsilon \downarrow 0}} \left(\underline{\lim_{n}} \mu_n(K^\varepsilon) - \varepsilon \right) \geq \underline{\lim_{\varepsilon \downarrow 0}} \mu(K^\varepsilon) - \varepsilon = \mu(K).$$

Thus there exists $\varepsilon_n \downarrow 0$ such that $K_n := K_{n,\varepsilon_n} \to K$ in $\mathfrak{K}(X)$ and $\overline{\lim}_n l_{a,\mu_n}(K_n) \leq l_{a,\mu}(K)$. Namely the Γ-limsup inequality holds.

Fix now $K \in \mathfrak{K}(X)$ and let (K_n) be a sequence converging to K in $\mathfrak{K}(X)$. Define $Q_m := K \bigcup \cup_{n \geq m} K_n$. Then Q_m is compact for all m, and by (P1)

$$\varlimsup_n \mu_n(K_n) \leq \varlimsup_m \varlimsup_n \mu_n(Q_m) \leq \varlimsup_m \mu(Q_m) = \mu(K),$$

which is the Γ-liminf inequality for (l_{a,μ_n}).

(L2) $\Rightarrow$ (P1). Let $\varepsilon > 0$ and $O \subset X$ be open. By the regularity of μ on X, there exists a compact set $K \subset O$ such that $\mu(K) \geq \mu(O) - \varepsilon$. By the Γ-limsup inequality for l_{a,μ_n} and Remark 3.6-(a), there exists a sequence (K_n) in $\mathfrak{K}(X)$ such that $\underline{\lim}_n \mu_n(K_n) \geq \mu(K)$. Since K is compact $O \supset K^\delta$ for some $\delta > 0$, so that, for n large enough, $K_n \subset O$. Thus

$$\varliminf_n \mu_n(O) \geq \varliminf_n \mu_n(K_n) \geq \mu(K) \geq \mu(O) - \varepsilon$$

and we conclude since $\varepsilon > 0$ was arbitrary.

(L3) $\Rightarrow$ (P1). By sequential compactness of Γ-convergence [2, Chapter 10], from any subsequence $(\mu_{n'})$ of (μ_n) one can extract a further subsequence $\mu_{n''}$ such that $l_{a,\mu_{n''}}$ Γ-converges to $l_{a,\mu}$, thus $\mu_{n''} \to \mu$ by the statement (L2)$\Rightarrow$(P1) proved above. Since $\mathcal{P}(X)$ is Polish, the Urysohn property holds, and $\mu_n \to \mu$.

(L1) $\Rightarrow$ (L2) and (L1) $\Rightarrow$ (L3) are trivial. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem 3.3. (P) $\Rightarrow$ (H). By (P), for each $\ell > 0$, there exists a compact set $K_\ell \subset X$ such that $\mu_n(K_\ell^c) \leq e^{-\ell a_n}$. By (3.7), for each $\nu \in \mathcal{P}(X)$, one has

$$\nu(K_\ell^c) \leq \frac{\frac{1}{a_n} H(\nu|\mu_n) + \frac{\log 2}{a_n}}{\frac{1}{a_n} \log\left(1 + \frac{1}{\mu_n(K_\ell^c)}\right)} \leq \frac{H_n^{\mathbf{a}}(\nu) + \frac{\log 2}{a_n}}{\ell}.$$

Let n_0 be such that $a_n \geq 1$ for $n \geq n_0$. Then for $M > 0$

$$\cup_{n \geq n_0}\left\{\nu \in \mathcal{P}(X) \,:\, H_n^{\mathbf{a}}(\nu) \leq M\right\} \subset \left\{\nu \in \mathcal{P}(X) \,:\, \forall \ell > 0, \, \nu(K_\ell^c) \leq \frac{M + \log 2}{\ell}\right\},$$

which is a tight set, and thus precompact in $\mathcal{P}(X)$. Since $\cup_{n < n_0}\{\nu \in \mathcal{P}(X) \,: H_n^{\mathbf{a}}(\nu) \leq M\}$ is precompact, we conclude.

(H) $\Rightarrow$ (P). Note that, by (3.4), for each $\ell > 0$ and integer $n_0 \geq 1$

$$\mathcal{P}_{n_0,\ell} := \cup_{n \geq n_0} \left\{\mu_n^{K^c}, \, K \subset X \text{ is compact and } \mu_n(K^c) \geq e^{-\ell a_n}\right\}$$

$$\subset \cup_{n \geq n_0} \left\{\nu \in \mathcal{P}(X) \,:\, H_n^{\mathbf{a}}(\nu) \leq \ell\right\}.$$

Therefore by (H), for each $\ell > 0$ there exists $n_0(\ell)$ such that $\mathcal{P}_{n_0(\ell),\ell}$ is precompact in $\mathcal{P}(X)$, and thus tight. In particular, for each $\ell > 0$ there exists a compact set

$K_\ell \subset X$ such that $\mu_n^{K^c}(K_\ell^c) \leq 1/2$, for each $n \geq n_0(\ell)$ and each compact set K such that $\mu_n(K^c) \geq \exp(-\ell \, a_n)$. But $\mu_n^{K^c}(K^c) = 1$ for each compact set K with $\mu_n(K^c) > 0$. Thus $K_\ell \neq K$ for each compact set K such that $\mu_n(K^c) \geq \exp(-\ell \, a_n)$ for some $n \geq n_0(\ell)$. Namely $\mu_n(K_\ell^c) \leq \exp(-\ell \, a_n)$ for each $\ell > 0$ and $n \geq n_0(\ell)$.

(P) ⇔ (L). It is trivial.   □

Proof of Theorem 3.4. (P1) ⇒ (H1). For $x \in X$ and $\delta > 0$ let $B_\delta(x)$ the open ball of radius δ centered at x. Fix n and define $\nu_{n,\delta} \in \mathcal{P}(X)$ by

$$\nu_{n,\delta} := \begin{cases} \mu_n^{B_\delta(x)} & \text{if } \mu_n(B_\delta(x)) > 0 \\ \delta_x & \text{otherwise} \end{cases}$$

and note that $H(\nu_{n,\delta}|\mu_n) = -\log \mu_n(B_\delta(x))$, where we understand $-\log(0) = +\infty$. By (P1), for each $\delta > 0$

$$\overline{\lim_n} H_n^{\mathbf{a}}(\nu_{n,\delta}) = -\underline{\lim_n} \frac{1}{a_n} \log \mu_n(B_\delta(x)) \leq \inf_{y \in B_\delta(x)} I(y) \leq I(x).$$

On the other hand $\lim_\delta \lim_n \nu_{n,\delta} = \delta_x$ in $\mathcal{P}(X)$. In particular, by a diagonal argument, there exists a sequence (δ_n) converging to 0 (slowly enough) such that $\lim_n \nu_{n,\delta_n} = \delta_x$ and $\overline{\lim_n} H_n^{\mathbf{a}}(\nu_{n,\delta_n}) \leq I(x)$. (H1) follows by Remark 3.6-(ii).

(H1) ⇒ (P2). Let $Y := \left\{ x \in X : (\Gamma\text{-}\underline{\lim}_n \varphi_n)(x) > -\infty \right\}$. By the definition of the Γ-liminf, for each $x \in Y$ there exist $\delta(x) > 0$ and $n_0(x) \in \mathbb{N}$ such that

$$\inf_{y \in B_{\delta(x)}} \inf_{n \geq n_0(x)} \varphi_n(y) > -\infty. \tag{3.10}$$

For $x \in Y$, let $(\nu_{n,x})$ be a sequence converging to δ_x in $\mathcal{P}(X)$ and such that $\overline{\lim} H_n^{\mathbf{a}}(\nu_{n,x}) \leq I(x)$. Such a sequence exists by (H1). By (3.4), it is easily seen that $(\nu_{n,x})$ can be assumed to be concentrated on $B_{\delta(x)}(x)$. By (2.1)

$$\log \mu_n(e^\varphi) \geq -H(\nu_{n,x}|\mu_n) + \nu_{n,x}(\varphi) \tag{3.11}$$

for each measurable $\varphi \colon X \to [-\infty, +\infty]$, provided we read the right hand side as $-\infty$ whenever $H(\nu_{n,x}|\mu_n) = +\infty$ or $\nu_{n,x}(\varphi^-) = +\infty$. Evaluating (3.11) for $\varphi = a_n \varphi_n$, taking the liminf in n and next optimizing on $x \in Y$

$$\underline{\lim_n} \frac{1}{a_n} \log \mu_n\big(\exp(a_n \varphi_n) \big) \geq \sup_{x \in Y} \left\{ -\overline{\lim_n} H_n^{\mathbf{a}}(\nu_{n,x}) + \underline{\lim_n} \nu_{n,x}(\varphi_n) \right\}.$$

Since $\nu_{n,x}$ is concentrated on $B_{\delta(x)}(x)$, and by (3.10) φ_n is bounded from below on $B_{\delta(x)}(x)$ for $n \geq n_0(x)$, (H1) and Proposition 3.2 yield

$$\underline{\lim_n} \frac{1}{a_n} \log \mu_n\big(\exp(a_n \varphi_n) \big) \geq \sup_{x \in Y} \left\{ -I(x) + (\Gamma\text{-}\underline{\lim_n} \varphi_n)(x) \right\}$$
$$= \sup_{x \in X} \left\{ -I(x) + (\Gamma\text{-}\underline{\lim_n} \varphi_n)(x) \right\}.$$

(P2) $\Rightarrow$ (P1). Fix an open set $O \subset X$ and $M > 0$. Then $\varphi_n \equiv M\mathbf{1}_O$ is lower semicontinuous, and thus coincides with its Γ-limit. It follows that

$$\frac{1}{a_n} \log \mu_n(\exp(a_n \varphi_n)) = \frac{1}{a_n} \log(1 + \mu_n(O) \exp(a_n M)))$$
$$\leq \frac{\log 2}{a_n} + \max(0, M + \frac{1}{a_n} \log(\mu_n(O)). \tag{3.12}$$

By (P2) applied to such a sequence φ_n, by taking the limit in (3.12) one gathers

$$\max\left(-M, \varlimsup_n \frac{1}{a_n} \log(\mu_n(O))\right) \geq -M + \sup_{x \in X}\left(M\mathbf{1}_O - I(x)\right) \geq -\inf_{x \in O} I(x).$$

This implies (P1) by taking $M \to \infty$.

(H2) $\Rightarrow$ (H1). Take $\nu = \delta_x$.

(H1) $\Rightarrow$ (H2). Since $H_n^{\mathbf{a}}$ is a convex functional, $\Gamma\text{-}\varlimsup_n H_n^{\mathbf{a}}$ is also convex. For an arbitrary $\nu \in \mathcal{P}(X)$, by Jensen inequality and (H1)

$$(\Gamma\text{-}\varlimsup_n H_n^{\mathbf{a}})(\nu) = (\Gamma\text{-}\varlimsup_n H_n^{\mathbf{a}})\left(\int_{\mathcal{P}(X)} \nu(dx)\,\delta_x\right)$$
$$\leq \int_{\mathcal{P}(X)} \nu(dx)\,(\Gamma\text{-}\varlimsup_n H_n^{\mathbf{a}})(\delta_x) \leq \int_{\mathcal{P}(X)} \nu(dx)\,I(x).$$

(P1) $\Rightarrow$ (L). Fix $\varepsilon > 0$, $K \in \mathfrak{K}(X)$, and let K^ε be the open ε-enlargement of K with respect any fixed compatible metric on X. Then, by the regularity of μ_n on X, for each n there exists $K_{n,\varepsilon} \subset K^\varepsilon$ compact such that $\mu_n(K_{n,\varepsilon}) \geq \exp(-\varepsilon a_n)\,\mu_n(K^\varepsilon)$. By (P1)

$$\varlimsup_{\varepsilon \downarrow 0} \varlimsup_n l_n^{\mathbf{a}}(K_{n,\varepsilon}) \leq \varlimsup_{\varepsilon \downarrow 0} \varlimsup_n -\frac{1}{a_n} \log \mu_n(K^\varepsilon) - \varepsilon \leq \varlimsup_{\varepsilon \downarrow 0} \inf_{x \in K^\varepsilon} I(x) - \varepsilon = \inf_{x \in K} I(x).$$

Thus there exists $\varepsilon_n \downarrow 0$ such that $K_n := K_{n,\varepsilon_n} \cup K$ converges to K in $\mathfrak{K}(X)$ and $\lim_n l_n^{\mathbf{a}}(K_n) \leq l_I(K)$. Namely the Γ-limsup inequality holds by Remark 3.6-(ii).

(L) $\Rightarrow$ (P1). Let $\varepsilon > 0$ and $O \subset X$ open. Since I is lower semicontinuous, there exists $K \subset O$ compact such that $\inf_{x \in K} I(x) \leq \inf_{x \in O} I(x) + \varepsilon$. By (L) there exists a sequence (K_n) converging to K in $\mathfrak{K}(X)$ such that $\varlimsup_n -\frac{1}{a_n} \log \mu_n(K_n) \leq \inf_{x \in K} I(x)$. Since for n large enough $K_n \subset O$

$$\varlimsup_n -\frac{1}{a_n} \log \mu_n(O) \leq \varlimsup_n -\frac{1}{a_n} \log \mu_n(K_n) \leq \inf_{x \in K} I(x) \leq \inf_{x \in K} I(x) + \varepsilon$$

and (P1) follows since $\varepsilon > 0$ is arbitrary. $\qquad\square$

Proof of Theorem 3.5. (P1) $\Rightarrow$ (H1). Let $x \in X$ and $(\nu_n) \subset \mathcal{P}(X)$ be such that $\lim_n \nu_n = \delta_x$ in $\mathcal{P}(X)$. In view of Remark 3.6-(i), it is enough to show $\varliminf_n H_n^{\mathbf{a}}(\nu_n) \geq$

$I(x)$. Fix $\varepsilon > 0$; since (ν_n) is tight, there exists $K \subset X$ compact such that $\nu_n(K) \geq 1 - \varepsilon$ for all n. By (3.7), for each Borel set $A \subset X$

$$H_n^{\mathbf{a}}(\nu_n) \geq \frac{\nu_n(A)}{a_n} \log\left(1 + \frac{1}{\mu_n(A)}\right) - \frac{\log 2}{a_n} \geq -\frac{\nu_n(A)}{a_n} \log \mu_n(A) - \frac{\log 2}{a_n}.$$

Take now $A = K \cap \overline{B}_\varepsilon(x)$, where $\overline{B}_\varepsilon(x)$ is the closed ball of radius ε centered at x. Note that A is compact and $\underline{\lim}_n \nu_n(A) \geq 1 - \varepsilon$, thus by (P1)

$$\varliminf_n H_n^{\mathbf{a}}(\nu_n) \geq -(1-\varepsilon)\varlimsup_n \frac{1}{a_n} \log \mu_n(A) \geq (1-\varepsilon) \inf_{y \in A} I(y) \geq (1-\varepsilon) \inf_{y \in \overline{B}_\varepsilon(x)} I(y).$$

Since $\varepsilon > 0$ was arbitrary, one can take the limit $\varepsilon \downarrow 0$ in the above formula, and since I is lower semicontinuous the right hand side in the above formula converges to $I(x)$.

(H2) $\Rightarrow$ (H1). Take $\nu = \delta_x$.

(H1) $\Rightarrow$ (H2). Assume (H1). Let $\nu \in \mathcal{P}(X)$ and (ν_n) be a sequence converging to ν in $\mathcal{P}(X)$. One needs to show that $\underline{\lim}_n H_n^{\mathbf{a}}(\nu_n) \geq \nu(I)$.

For $\delta, \ell > 0$ let $(E_{\delta,\ell}^i)_{i=0}^{N_{\delta,\ell}}$ be as in Remark 3.9 (with $\mu = \nu$). For $i \in \{0, \ldots, N_{\delta,\ell}\}$ such that $\nu_n(E_{\delta,\ell}^i) > 0$ define the probability measures $\nu_{n,\delta,\ell}^i := \nu_n^{E_{\delta,\ell}^i} \in \mathcal{P}(X)$. Then by (3.3), for each n, $\ell > 0$,

$$\begin{aligned}
H(\nu_n|\mu_n) &= \sum_{i=0}^{N_{\delta,\ell}} \nu_n(E_{\delta,\ell}^i) H(\nu_{n,\delta,\ell}^i|\mu_n) + \nu_n(E_{\delta,\ell}^i) \log \nu_n(E_{\delta,\ell}^i) \\
&\geq \sum_{i=0}^{N_{\delta,\ell}} \nu_n(E_{\delta,\ell}^i) H(\nu_{n,\delta,\ell}^i|\mu_n) - \log N_{\delta,\ell},
\end{aligned} \tag{3.13}$$

where the terms in the above sums are understood to vanish for all i such that $\nu_n(E_{\delta,\ell}^i) = 0$. Dividing (3.13) by a_n, taking the liminf and recalling that the sets $E_{\delta,\ell}^i$ are ν-regular

$$\varliminf_n H_n^{\mathbf{a}}(\nu_n) \geq \sum_{i=0}^{N_{\delta,\ell}} \varliminf_n \nu_n(E_{\delta,\ell}^i) H_n^{\mathbf{a}}(\nu_{n,\delta,\ell}^i) = \sum_{i=0}^{N_{n,\ell}} \nu(E_{\delta,\ell}^i) \varliminf_n H_n^{\mathbf{a}}(\nu_{n,\delta,\ell}^i) = \int I_{\delta,\ell}(x) d\nu(x),$$

where $I_{\delta,\ell}$ is defined by

$$I_{\delta,\ell}(x) := \varliminf_n H_n^{\mathbf{a}}(\nu_{n,\delta,\ell}^i) \qquad \text{if } x \in E_{\delta,\ell}^i.$$

Note that $I_{\delta,\ell}$ is monotone both in δ and ℓ, the partitions $\{E_{\delta,\ell}^i\}$ are increasing as $\delta \downarrow 0$ and $\ell \uparrow +\infty$, see (2.1). By monotone convergence

$$\varliminf_n H_n^{\mathbf{a}}(\nu_n) \geq \int \left(\lim_\ell \lim_\delta I_{\delta,\ell}(x)\right) d\nu(x).$$

However, since $\lim_\ell \lim_\delta \lim_n \nu_n = \delta_x$, (H1) implies $\lim_\ell \lim_\delta I_{\delta,\ell}(x) \geq I(x)$ pointwise by Remark 3.6-(b), thus the conclusion.

(H2) $\Rightarrow$ (P2). Consider the sequence (ν_n) in $\mathcal{P}(X)$ defined as

$$\nu_n(dx) := \frac{\exp(-a_n\,\varphi_n(x))}{\mu_n\big(\exp(-a_n\,\varphi_n)\big)}\mu_n(dx).$$

By (3.3)

$$\tfrac{1}{a_n}\log\mu_n\big(\exp(-a_n\,\varphi_n)\big) = -\nu_n(\varphi_n) - H_n^{\mathbf{a}}(\nu_n).$$

By (3.1), (ν_n) is tight and thus precompact in $\mathcal{P}(X)$. Let ν be an arbitrary limit point of (ν_n). Taking the limsup in n, using Proposition 3.2 and (H2)

$$\varlimsup_n \tfrac{1}{a_n}\log\mu_n\big(\exp(-a_n\,\varphi_n)\big) \leq -\varliminf_n \nu_n\big(\varphi_n\big) - \varliminf_n H_n^{\mathbf{a}}(\nu_n) \leq -\nu(\Gamma\text{-}\varliminf_n \varphi_n) - \nu(I)$$

$$\leq \sup_{x\in X}\left\{-(\Gamma\text{-}\varliminf_n \varphi_n)(x) - I(x)\right\}.$$

(P2) $\Rightarrow$ (P1). Let K be a compact set in X, and for $M > 0$ let $\varphi_n \equiv \varphi = M\mathbf{1}_{K^c}$. (φ_n) satisfies (3.1). Moreover $\varphi = \Gamma\text{-}\lim_n \varphi_n$ since it is lower semicontinuous. Therefore, assuming (P2),

$$\varlimsup_n \tfrac{1}{a_n}\log\mu_n(K) \leq \varlimsup_n \tfrac{1}{a_n}\log\mu_n\big(\exp(-a_n\,M\,\mathbf{1}_{K^c})\big) \leq \sup_{x\in X}\{-M\,\mathbf{1}_{K^c}(x) - I(x)\}.$$

Letting $M \to +\infty$, (P1) follows.

(P1) $\Rightarrow$ (L). Fix $K \in \mathfrak{K}(X)$ and let (K_n) be a sequence converging to K in $\mathfrak{K}(X)$. Define $Q_m := K \bigcup \cup_{n\geq m} K_n$. Then Q_m is compact for all m, and by (P1)

$$\varliminf_n -\tfrac{1}{a_n}\log\mu_n(K_n) \geq \lim_m \varliminf_n -\tfrac{1}{a_n}\log\mu_n(Q_m) \geq \lim_m \inf_{x\in Q_m} I(x) = \inf_{x\in K} I(x)$$

where we used the lower semicontinuity of I in the last inequality.

(L) $\Rightarrow$ (P1). The weak upper bound is nothing but the Γ-liminf inequality for $l_n^{\mathbf{a}}$ along a constant sequence $K_n \equiv K$.

The implications (P3) $\Rightarrow$ (P1), (P4) $\Rightarrow$ (P2), and {Theorem 3.3-(P), (P1)} $\Rightarrow$ (P3) are trivial. On the other hand the implication {Theorem 3.3-(P), (P2)} $\Rightarrow$ (P4) follows from a standard cut-off argument. $\qquad\square$

4. Applications to large deviations

In this section a few consequences of the results of Section 3 are discussed.

The following proposition gives an explicit representation of the optimal upper and low bound rate functions, see also [5, Chapter 4.1], which will come useful in the following:

Proposition 4.1 (Existence of large deviations). *There exist $\overline{I}^{\mathbf{a}}$ and $\underline{I}^{\mathbf{a}}$ which are respectively the minimal and maximal lower semicontinuous functionals for which the weak lower bound and the upper bound hold respectively. A weak LDP holds for (μ_n) with speed (a_n) if $\overline{I}^{\mathbf{a}} = \underline{I}^{\mathbf{a}}$. The following representations of $\overline{I}^{\mathbf{a}}$ and $\underline{I}^{\mathbf{a}}$ hold:*

$$\overline{I}^{\mathbf{a}}(x) = (\Gamma\text{-}\overline{\lim_n} H_n^{\mathbf{a}})(\delta_x) = \lim_{\delta\downarrow 0}\overline{\lim_n} \frac{-1}{a_n}\log \mu_n(B_\delta(x)) = \sup_{(V_n)} (\Gamma\text{-}\overline{\lim_n} V_n)(x), \quad (4.1)$$

$$\underline{I}^{\mathbf{a}}(x) = (\Gamma\text{-}\underline{\lim_n} H_n^{\mathbf{a}})(\delta_x) = \lim_{\delta\downarrow 0}\underline{\lim_n} \frac{-1}{a_n}\log \mu_n(B_\delta(x)) = \sup_{(V_n)} (\Gamma\text{-}\underline{\lim_n} V_n)(x), \quad (4.2)$$

where the supremums are carried over all the sequences (V_n) such that $V_n\colon X \to [-\infty,+\infty]$ is measurable (or equivalently continuous and bounded) and $\mu_n(e^{a_n V_n}) \leq 1$ (or equivalently $\mu_n(e^{a_n V_n}) = 1$ or equivalently $\overline{\lim}_n a_n^{-1}\log \mu_n(e^{a_n V_n}) \leq 0$).

Proof. The existence of the optimal rate functions $\overline{I}^{\mathbf{a}}$ and $\underline{I}^{\mathbf{a}}$, and the first representation formula above follows from the equivalences (P1) ⇔ (H1) in Theorems 3.4-3.5. By (3.4), it is easy to see that, given $x \in X$, the sequence $\mu_n^{B_\delta(x)}$ is an optimal recovery sequence for δ_x in the Γ-limit of $H_n^{\mathbf{a}}$, provided $\delta \downarrow 0$ after $n \to +\infty$. The second equalities in (4.1)-(4.2) then follow again by (P1) ⇔ (H1) in Theorem 3.4-3.5. The third equalities in (4.1)-(4.2) follow in the same fashion, if one remarks that the supremum in the rightest hand side is attained on the family of sequences (V_n) of the form

$$V_n(y) := \begin{cases} -\frac{1}{a_n}\log \mu_n(B_\delta(x)) & \text{if } y \in B_\delta(x) \\ -\infty & \text{if } y \notin B_\delta(x) \end{cases}$$

as δ runs in $[0, 1[$. □

The following corollaries follow easily from Proposition 4.1.

Corollary 4.2 (Improving the bounds). *Let $\mathfrak{A}$ be a set of indexes.*
   Assume that for each $\alpha \in \mathfrak{A}$, μ_n satisfies a LD lower bound with speed (a_n) and rate $\overline{I}_\alpha$. Then (μ_n) satisfies a LD lower bound with speed (a_n) and rate equal to the lower semicontinuous envelope of $x \mapsto \inf_{\alpha\in\mathfrak{A}} \overline{I}_\alpha(x)$.
   Assume that for each $\alpha \in \mathfrak{A}$, μ_n satisfies a weak LD upper bound with speed (a_n) and lower semicontinuous rate $\underline{I}_\alpha$. Then (μ_n) satisfies a weak LD upper bound with speed (a_n) and rate $x \mapsto \sup_{\alpha\in\mathfrak{A}} \underline{I}_\alpha(x)$.

Corollary 4.3 (Large deviations for double indexed sequences). *Let $(\mu_{n,m})_{n,m}$ be a double-indexed sequence, directed by $(n', m') \geq (n, m)$ if $n' > n$ or $n' = n$ and $m' \geq m$. For each fixed m let $\overline{I}_m^{\mathbf{a}}$ and $\underline{I}_m^{\mathbf{a}}$ be the optimal lower and weak upper bound rate functionals for $(\mu_{n,m})_n$.*
   Then $\overline{I}^{\mathbf{a}} = \Gamma\text{-}\overline{\lim}_m \overline{I}_m^{\mathbf{a}}$ and $\underline{I}^{\mathbf{a}} = \Gamma\text{-}\underline{\lim}_m \underline{I}_m^{\mathbf{a}}$, where $\overline{I}^{\mathbf{a}}$ and $\underline{I}^{\mathbf{a}}$ are defined as in (4.1)-(4.2) by changing the index n with (n, m) (or in other words, performing the limits $n \to \infty$ and next $m \to \infty$).

Proposition 4.4 (General contraction principle). *Let X, Y be two Polish spaces, let (μ_n) be a sequence in $\mathcal{P}(X)$ and for $n \in \mathbb{N}$ let θ_n, $\theta \colon X \to Y$ be measurable maps. Assume that $\theta_n \to \theta$ uniformly on compact sets. Define $\gamma_n = \mu_n \circ \theta_n^{-1} \in \mathcal{P}(Y)$. Then*

  (i) *If (μ_n) satisfies a LD lower bound with speed (a_n) and lower semicontinuous rate $\bar{I} \colon X \to [0, +\infty]$, then (γ_n) satisfies a LD lower bound with the same speed and rate $\bar{J} \colon Y \to [0, +\infty]$ given by*

$$\bar{J}(y) := \inf_{x \in \underline{\Lambda}^y} \bar{I}(x),$$

  where

$$\underline{\Lambda}^y := \lim_{\delta \downarrow 0} \mathrm{Interior}\big(\theta^{-1}(B_\delta(y))\big);$$

  (ii) *If (μ_n) is exponentially tight and satisfies a LD upper bound with speed (a_n) and lower semicontinuous rate $\underline{I} \colon X \to [0, +\infty]$, then (γ_n) satisfies a LD weak upper bound with the same speed and rate $\underline{J} \colon Y \to [0, +\infty]$ given by*

$$\underline{J}(y) := \inf_{x \in \overline{\Lambda}^y} \underline{I}(x),$$

  where

$$\overline{\Lambda}^y := \lim_{\delta \downarrow 0} \mathrm{Closure}\big(\theta^{-1}(B_\delta(y))\big).$$

Note in particular that $\overline{\Lambda}^y \supset \theta^{-1}(y) \supset \underline{\Lambda}^y$, with equality holding if θ is continuous (recovering the standard contraction principle).

The proof requires a similar statement concerning Γ-convergence (contraction principles are surprisingly missing from the Γ-convergence literature).

Lemma 4.5. *Let X, Y be two Polish spaces, and let (I_n) be a sequence of lower semicontinuous functions on X. Let θ_n, θ, $\bar{J}$ and $\underline{J}$ be as in Proposition 4.4. Define $J_n \colon Y \to [0, +\infty]$ as $J_n(y) = \inf_{x \in \theta_n^{-1}(y)} I_n(x)$. Then $\Gamma\text{-}\overline{\lim}_n J_n \leq \bar{J}$ and, if (I_n) is equicoercive, $\Gamma\text{-}\underline{\lim}_n J_n \geq \underline{J}$.*

Proof. Fix $y \in Y$.

Γ-*limsup inequality.* If $\underline{\Lambda}^y = \emptyset$ there is nothing to prove. Otherwise, take $\varepsilon > 0$ and $x_\varepsilon \in \underline{\Lambda}^y$ such that $\bar{J}(y) \geq I(x_\varepsilon) - \varepsilon$. Then there is a sequence $(x_{n,\varepsilon})$ converging to x_ε in X such that $\overline{\lim}_n I_n(x_{n,\varepsilon}) \leq I(x_\varepsilon)$. Since $x_\varepsilon \in \underline{\Lambda}^y$, for all $\delta > 0$ and $n \geq n_\delta$ large enough, $x_{n,\varepsilon} \in \mathrm{Interior}\big(\theta^{-1}(B_\delta(x))\big)$. So that $\theta(x_{n,\varepsilon}) \to y$, and setting $y_{n,\varepsilon} = \theta_n(x_{n,\varepsilon})$ one has $\lim_n y_{n,\varepsilon} = y$. On the other hand, $\overline{\lim}_n J_n(y_{n,\varepsilon}) \leq \overline{\lim}_n I(x_{n,\varepsilon}) \leq \bar{J}(y) + \varepsilon$. Thus there is a subsequence (ε_n) such that $\underline{\lim}_n J_n(y_n) \leq \bar{J}_n(y) + \varepsilon$ with $y_n = y_{n,\varepsilon_n}$.

Γ-liminf inequality. Let (y_n) be a sequence converging to y. Up to passing to a subsequence (still label n here), we can assume $\sup_n J_n(y_n) < +\infty$, the inequality being trivial otherwise. In particular, $\theta_n^{-1}(y_n) \neq \emptyset$. For $\varepsilon > 0$, let $x_{n,\varepsilon} \in \theta_n^{-1}(y_n)$ be such that $I_n(x_{n,\varepsilon}) \leq J_n(y_n) + \varepsilon$. Since I_n is equicoercive and $J_n(y_n)$ uniformly bounded, $(x_{n,\varepsilon})$ is precompact. It is easy to check that any limit point of $(x_{n,\varepsilon})$ is in $\overline{\Lambda}^y$ (which is nonempty under the above assumptions). In particular, by the Γ-liminf inequality for (I_n)

$$\varliminf_n J_n(y_n) \geq \varliminf_n I_n(x_{n,\varepsilon}) - \varepsilon \geq \underline{J}(y) - \varepsilon$$

and we get the statement since (y_n) and $\varepsilon > 0$ where arbitrary. $\qquad \square$

Proof of Theorem 4.4. Let $\vartheta \colon \mathcal{P}(X) \to \mathcal{P}(Y)$ be defined by $\vartheta(\mu) = \mu \circ \theta^{-1}$, and let ϑ_n be defined similarly. It is easy to see that $\vartheta_n \to \vartheta$ uniformly on compact subsets of the Polish space $\mathcal{P}(X)$. Then (3.5) implies that for $\beta \in \mathcal{P}(Y)$

$$\frac{1}{a_n} H(\beta | \gamma_n) = \frac{1}{a_n} H(\beta | \vartheta_n(\mu_n)) = \inf_{\nu \in \vartheta_n^{-1}(\beta)} \frac{1}{a_n} H(\nu | \mu_n).$$

Therefore, by Lemma 4.5 (applied to the Polish space $\mathcal{P}(X)$ and maps ϑ_n) and the equivalence (P1)-(H1) in Theorem 3.4, (γ_n) satisfies a LD lower bound with speed (a_n) and rate

$$\tilde{J}(y) := \inf_{\nu \in \underline{\Delta}^y} \int_X \nu(dx) \, \bar{I}(x)$$

with

$$\underline{\Delta}^y = \lim_{r \downarrow 0} \operatorname{Interior}\left(\vartheta^{-1}(\mathcal{B}_r(\delta_y))\right),$$

where $\mathcal{B}_r(\delta_y)$ is the open ball of radius $r > 0$ centered in δ_y in $\mathcal{P}(X)$ (with respect to a fixed compatible distance on $\mathcal{P}(X)$). However, since $\delta_x \in \underline{\Delta}^y$ iff $x \in \underline{\Lambda}^y$, it is easy to see that $\tilde{J} = \bar{J}$. Namely the statement (i) holds.

In order to prove (ii), note that by Theorem 3.3 the sequence of functionals $\frac{1}{a_n} H(\cdot | \mu_n)$ is equicoercive. One can then apply the Γ-liminf statement in Lemma 4.5, to prove (ii) following exactly the same lines as in (i). $\qquad \square$

The following result appears to be new in such a generality.

Theorem 4.6 (Large deviations for coupled systems). *For $n \in \mathbb{N}$, $i = 1, 2$ let $(\Omega^i, \mathfrak{F}^i, \mathbb{P}_n^i)$ be standard probability spaces, and let $(\Omega, \mathfrak{F}, \mathbb{P}_n)$ be their product space. Let X, Y be Polish spaces with compatible distances d_X and d_Y. Assume that for each n there are measurable maps $F_n \colon Y \times \Omega^1 \to X$, $G_n \colon X \times \Omega^2 \to Y$, $\xi_n \colon \Omega \to X$ and $\eta_n \colon \Omega \to Y$ such that $\mathbb{P}_n$-a.s.*

$$\xi_n(\omega^1, \omega^2) = F_n(\eta_n(\omega^1, \omega^2), \omega^1)$$

$$\eta_n(\omega^1, \omega^2) = G_n(\xi_n(\omega^1, \omega^2), \omega^2).$$

For fixed $x \in X$, $y \in Y$ define $f_n^y(\omega^1) = F_n(y, \omega^1)$, $g_n^x(\omega^2) = G_n(x, \omega^2)$, and let $\mu_n^y := \mathbb{P}_n^1 \circ (f_n^y)^{-1} \in \mathcal{P}(X)$ and $v_n^x := \mathbb{P}_n^2 \circ (g_n^x)^{-1} \in \mathcal{P}(Y)$ be the laws of f_n^y and g_n^x respectively. Assume that for fixed $x \in X$, $y \in Y$ there exists a positive function $q \equiv q^{x,y} \in C_b(\mathbb{R}^+; \mathbb{R}^+)$ with $q(0) = 0$ such that

(i) (μ_n^y) *satisfies a weak LDP with speed* (a_n) *and lower semicontinuous rate* $x \mapsto K^y(x)$;

(ii) (ν_n^x) *satisfies a weak LDP with speed* (a_n) *and lower semicontinuous rate* $y \mapsto J^x(y)$;

(iii) *For each* $\varepsilon > 0$

$$\varlimsup_n \frac{1}{a_n} \log \mathbb{P}_n\Big(d_X(\xi_n, x) + d_Y(\eta_n, y) \geq q\big(d_X(f_n^y, x) + d_Y(g_n^x, y)\big) + \varepsilon\Big) = -\infty;$$

(iv) *For each* $\varepsilon > 0$

$$\varlimsup_n \frac{1}{a_n} \log \mathbb{P}_n\Big(d_X(f_n^y, x) + d_Y(g_n^x, y) \geq q\big(d_X(\xi_n, x) + d_Y(\eta_n, y)\big) + \varepsilon\Big) = -\infty.$$

Define $I: X \times Y \to [0, +\infty]$ *as the lower semicontinuous envelope of the map* $(x, y) \mapsto K^y(x) + J^x(y)$ *and let* $\gamma_n := \mathbb{P}_n \circ (\xi_n, \eta_n)^{-1}$ *be the law* (ξ_n, η_n). *Then* (γ_n) *satisfies a weak LDP with speed* (a_n) *and rate* I.

In the above theorem, (iii) and (iv) are basically uniform regularity requirements on F_n, G_n. (iii) is only used in the lower bound, (iv) in the upper bound. Theorem 4.6 applies in the following kind of situations. Suppose we have a weak solution to the system of SDEs on $\mathbb{R} \times \mathbb{R}$

$$\begin{cases} \dot{\xi} = b_n(\xi, \eta) + \frac{1}{n}\dot{W}^1 \\ \dot{\eta} = c_n(\xi, \eta) + \frac{1}{n}\dot{W}^2 \end{cases} \tag{4.3}$$

where W^1 and W^2 are independent Brownian motions. If one knows the LD on $C([0, T]; \mathbb{R})$ of the solutions to

$$\dot{\zeta} = b_n(\zeta, y) + \frac{1}{n}\dot{W}^1$$
$$\dot{\zeta} = c_n(x, \zeta) + \frac{1}{n}\dot{W}^2$$

for fixed x, $y \in C([0, T]; \mathbb{R})$, and if conditions (iii)-(iv) is satisfied (which happens under uniform Lipschitz conditions on b_n and c_n), then one gets the LD for the law of the original coupled system (4.3). While this kind of statement can be quite standard for finite-dimensional systems, Theorem 4.6 also applies for instance in the stochastic PDEs framework, and when considering asymptotics other than the small noise limit (*e.g.*, slow-fast random dynamics).

Proof of Theorem 4.6. Fix $x \in X$, $y \in Y$ and let $q \equiv q^{x,y}$ be as in the hypothesis.

Lower bound. By (i), (ii) and Theorem 3.4-(H1) there exist sequences (κ_n) in $\mathcal{P}(X)$, (λ_n) in $\mathcal{P}(Y)$ such that

$$\begin{aligned} \kappa_n \to \delta_x \quad &\text{and} \quad \varlimsup_n \frac{1}{a_n} H(\kappa_n | \mu_n^y) \leq K^y(x) \\ \lambda_n \to \delta_y \quad &\text{and} \quad \varlimsup_n \frac{1}{a_n} H(\lambda_n | \mu_n^y) \leq J^x(y). \end{aligned} \tag{4.4}$$

By (3.6) there exist probabilities $(\mathbb{Q}_n^1)$, $(\mathbb{Q}_n^2)$ on $(\Omega^1, \mathfrak{F}^1)$, $(\Omega^2, \mathfrak{F}^2)$ respectively, such that $\kappa_n = \mathbb{Q}_n^1 \circ (f_n^y)^{-1}$, $\lambda_n = \mathbb{Q}_n^2 \circ (g_n^x)^{-1}$ and

$$H(\kappa_n | \mu_n^y) = H(\mathbb{Q}_n^1 | \mathbb{P}_n^1) \qquad\qquad H(\lambda_n | \nu_n^x) = H(\mathbb{Q}_n^2 | \mathbb{P}_n^2). \qquad (4.5)$$

Set now $\mathbb{Q}_n = \mathbb{Q}_n^1 \otimes \mathbb{Q}_n^2$, and define $\beta_n \in \mathcal{P}(X \times Y)$ as the law of (ξ_n, η_n) under $\mathbb{Q}_n$, $\beta_n := \mathbb{Q}_n \circ (\xi_n, \eta_n)^{-1}$. Then patching (4.4) and (4.5) together

$$\begin{aligned}
\overline{\lim_n} \frac{1}{a_n} H(\beta_n | \gamma_n) &\leq \overline{\lim_n} \frac{1}{a_n} H(\mathbb{Q}_n | \mathbb{P}_n) = \overline{\lim_n} \frac{1}{a_n} H(\mathbb{Q}_n^1 | \mathbb{P}_n^1) + \frac{1}{a_n} H(\mathbb{Q}_n^1 | \mathbb{P}_n^1) \\
&\leq \overline{\lim_n} \frac{1}{a_n} H(\kappa_n | \mu_n^y) + \overline{\lim_n} \frac{1}{a_n} H(\lambda_n | \nu_n^x) \leq K^y(x) + J^x(y).
\end{aligned} \qquad (4.6)$$

In particular, if $K^y(x) + J^x(y) < +\infty$, $\frac{1}{a_n} H(\mathbb{Q}_n | \mathbb{P}_n)$ is uniformly bounded. Thus by (iii) and (3.7), for each $\varepsilon > 0$

$$\overline{\lim_n} \mathbb{Q}_n \Big(d_X(\xi_n, x) + d_Y(\eta_n, y) \geq q \big(d_X(f_n^y, x) + d_Y(g_n^x, y) \big) + \varepsilon \Big) = 0. \qquad (4.7)$$

By (4.4) and $q(0) = 0$, for all $\varepsilon' > 0$

$$\overline{\lim_n} \mathbb{Q}_n \big(q(d_X(f_n^y, x) + d_Y(g_n^x, y)) > \varepsilon' \big) = 0. \qquad (4.8)$$

(4.7) and (4.8) yield $\beta_n \to \delta_{(x,y)}$ in $\mathcal{P}(X \times Y)$. Inequality (4.6) and Theorem 3.4-(H1) imply that the lower bound holds with rate $K^y(x) + J^x(y)$, and by Corollary 4.2, it holds with its lower semicontinuous envelope I.

Upper bound. Assume that $\beta_n \in \mathcal{P}(X \times Y)$ is such that $\beta_n \to \delta_{(x,y)}$. By Theorem 3.5, we need to prove that $\underline{\lim}_n a_n^{-1} H(\beta_n | \gamma_n) \geq I(x, y)$. Up to passing to a subsequence, one can assume $a_n^{-1} H(\beta_n | \gamma_n)$ to be bounded uniformly in n, so that by the Remark 3.7 there exists a probability $\mathbb{Q}_n$ on $(\Omega, \mathfrak{F})$ such that $\beta_n = \mathbb{Q}_n \circ (\xi_n, \eta_n)^{-1}$ and $H(\beta_n | \gamma_n) = H(\mathbb{Q}_n | \mathbb{P}_n)$. Let $\mathbb{Q}_n^1(d\omega^1)$ and $\mathbb{Q}_n^2(d\omega^2)$ be the marginals of $\mathbb{Q}_n$ on Ω^1 and Ω^2 respectively. For all $\varepsilon > 0$

$$\begin{aligned}
&\mathbb{Q}_n^1(d_X(f_n^y, x) > \varepsilon) + \mathbb{Q}_n^2(d_Y(g_n^x, y) > \varepsilon) \\
&= \mathbb{Q}_n(d_X(f_n^y, x) > \varepsilon) + \mathbb{Q}_n(d_Y(g_n^x, y) > \varepsilon) \\
&\leq 2 \mathbb{Q}_n \big(d_X(f_n^y, x) + d_Y(g_n^x, y) > \varepsilon \big) \\
&\leq 2 \mathbb{Q}_n \Big(d_X(f_n^y, x) + d_Y(g_n^x, y) > q \big(d_X(\xi_n, x) + d_Y(\eta_n, y) \big) + \varepsilon/2 \Big) \\
&\quad + 2 \mathbb{Q}_n \Big(q \big(d_X(\xi_n, x) + d_Y(\eta_n, y) \big) > \varepsilon/2 \Big).
\end{aligned} \qquad (4.9)$$

The last line of (4.9) vanishes as $n \to +\infty$, since $\beta_n \to \delta_{x,y}$. On the other hand, by (3.7) and hypothesis (iv)

$$\overline{\lim_n} \mathbb{Q}_n \Big(d_X(f_n^y, x) + d_Y(g_n^x, y) > q \big(d_X(\xi_n, x) + d_Y(\eta_n, y) \big) + \varepsilon/2 \Big)$$

$$\leq \overline{\lim_n} \frac{\dfrac{\log 2}{a_n} + \dfrac{1}{a_n} H(\mathbb{Q}_n | \mathbb{P}_n)}{-\dfrac{1}{a_n} \log \mathbb{P}_n \Big(d_X(f_n^y, x) + d_Y(g_n^x, y) > q \big(d_X(\xi_n, x) + d_Y(\eta_n, y) \big) + \varepsilon/2 \Big)} = 0.$$

Thus by (4.9), for all $\varepsilon > 0$

$$\lim_n \mathbb{Q}_n^1(d_X(f_n^y, x) > \varepsilon) = \lim_n \mathbb{Q}_n^2(d_Y(g_n^x, y) > \varepsilon) = 0$$

and, letting $\kappa_n = \mathbb{Q}_n^1 \circ (f_n^y)^{-1} \in \mathcal{P}(X), \lambda_n = \mathbb{Q}_n^2 \circ (g_n^x)^{-1} \in \mathcal{P}(Y)$, we gather

$$\lim_n \kappa_n = \delta_x \quad \text{in } \mathcal{P}(X) \qquad \lim_n \lambda_n = \delta_y \quad \text{in } \mathcal{P}(Y). \qquad (4.10)$$

Disintegrate now $\mathbb{Q}_n$ as $\mathbb{Q}_n(d\omega^1, d\omega^2) = \mathbb{Q}_n^1(d\omega^1)\mathbf{Q}_n(\omega^1; d\omega^2)$. Then by explicit calculations and Jensen inequality

$$
\begin{aligned}
H(\beta_n|\gamma_n) = H(\mathbb{Q}_n|\mathbb{P}_n) &= H(\mathbb{Q}_n^1|\mathbb{P}_n^1) + \int_{\Omega^1} \mathbb{Q}_n(d\omega^1) \, H(\mathbf{Q}_n(\omega^1; \cdot)|\mathbb{P}_n^2) \\
&\geq H(\mathbb{Q}_n^1|\mathbb{P}_n^1) + H\left(\int_{\Omega^1} \mathbb{Q}_n(d\omega^1) \, \mathbf{Q}_n(\omega^1; \cdot) \Big| \mathbb{P}_n^2\right) \\
&= H(\mathbb{Q}_n^1|\mathbb{P}_n^1) + H(\mathbb{Q}_n^2|\mathbb{P}_n^2) + \geq H(\kappa_n^y|\mu_n^y) + H(\lambda_n^y|\nu_n^x).
\end{aligned}
\qquad (4.11)
$$

By (4.10), hypothesis (i), (ii) and Theorem 3.5-(H1) we get

$$\varliminf_n \frac{1}{a_n} H(\beta_n|\gamma_n) \geq \varliminf_n \frac{1}{a_n} H(\kappa_n^y|\mu_n^y) + \varliminf_n \frac{1}{a_n} H(\lambda_n^y|\nu_n^x) \geq K^y(x) + J^x(y) \geq I(x, y)$$

concluding the proof. $\square$

5. LD and Γ-convergence topology

We say that the speed $\mathbf{a}$ is trivial for the LD of (μ_n) if the functionals $\bar{I}^{\mathbf{a}}, \underline{I}^{\mathbf{a}}$ only take the values 0 and $+\infty$. Assume, for the sake of simplicity, that (μ_n) converges to μ in $\mathcal{P}(X)$, and note that

$$\text{Support}(\mu) \subset \text{Closure}\Big(\varliminf_n \text{Support}(\mu_n)\Big). \qquad (5.1)$$

If the inclusion (5.1) is actually an equality, which means that the measures μ_n do not feature any concentration phenomena in the limit $n \to \infty$, then it is easy to check that

$$\bar{I}^{\mathbf{a}}(x) = \underline{I}^{\mathbf{a}}(x) = \begin{cases} 0 & \text{if } x \in \text{Support}(\mu) \\ +\infty & \text{otherwise} \end{cases}$$

regardless of the speed (a_n). That is, the LD of (μ_n) are trivial. On the other hand, if the inclusion (5.1) is strict, one can prove that there exists a non-trivial speed (a_n). This remark suggests that, when considering LD as a notion of convergence on the space of couples $(a, \mu) \in \mathbb{R}^+ \times \mathcal{P}(X)$, one should identify the singular measures, since no speed (a_n) can catch the concentration speed of the support of

Dirac masses. More precisely, recall (2.2), and define the equivalence relation on $\mathbb{R}^+ \times \mathcal{P}(X)$

$$(a, \mu) \sim (a', \mu') \Leftrightarrow \big\{(a, \mu) = (a', \mu') \text{ or } \exists x \in X \ : \ \mu = \mu' = \delta_x\big\} \Leftrightarrow l_{a,\mu} = l_{a',\mu'},$$

and let

$$\mathcal{U}(X) := \mathbb{R}^+ \times \mathcal{P}(X)/\sim$$
$$\mathcal{V}(X) := \big\{I \colon X \to [0, +\infty], \ I \text{ is lower semicontinuous}\big\}$$
$$\mathcal{W}(X) := \mathcal{U}(X) \cup \mathcal{V}(X)$$
$$\mathcal{L}(X) := \big\{l \colon \mathfrak{K}(X) \to [0, +\infty], \ l \text{ is lower semicontinuous}, \ l(K) \le l(K') \text{ if } K \supset K'\big\}.$$

We want to look at LD as a notion of convergence in $\mathcal{W}(X)$. To our aim, $\mathcal{L}(X)$ is naturally equipped with the topology of Γ-convergence [2, Chapter 10] on $\mathfrak{K}(X)$. One can prove that $\mathcal{L}(X)$ is a T_1, supercompact space (an easy extension of [2, Theorem 10.6]). The maps (2.2), (2.3) define an injection $\mathcal{W}(X) \hookrightarrow \mathcal{L}(X)$, and we equip $\mathcal{W}(X)$ with the induced topology. We say that a subset $\mathcal{W}(X)$ is equicoercive, if its homeomorphic image in $\mathcal{L}(X)$ is equicoercive. Note in particular that a sequence (a_n, μ_n) is equicoercive iff (μ_n) is exponentially tight with speed (a_n).

The following theorem is a consequence of the equivalence between the (L) and (P) statements in Proposition 3.2 and Theorems 3.4-3.5, and the metrizability properties of the topology of Γ-convergence for equicoercive subsets [2, Theorem 10.22].

Theorem 5.1. *Let (w_n) be a sequence converging to w in $\mathcal{W}(X)$. Then*

  (i) *Up to $\sim$ identification, if $w_n = (a_n, \mu_n)$ and (a_n) is bounded, then $a_n \to a$ and $\mu_n \to \mu$ and $w = (a, \mu) \in \mathcal{U}(X)$;*
  (ii) *If $w_n = (a_n, \mu_n)$ and $a_n \to +\infty$, then $w = I \in \mathcal{V}(X)$ and μ_n satisfies a weak LDP with speed (a_n) and rate I;*
  (iii) *If $w_n = I_n$, then $w = I$ and I_n Γ-converges to I.*

The relative topology induced by $\mathcal{W}(X)$ on an equicoercive subset $\mathcal{E}$ is metrizable.

Roughly speaking, the previous theorem states that the topology induced by $\mathcal{E}$ on measures is the usual topology of narrow convergence, the topology it induces on functionals is the topology of Γ-convergence. However, while the space of functionals on X is compact under this topology, it can happen that measures converge to functionals, and this is the case if and only if a LDP holds. It is worth to remark that up to identification $\mathcal{W}(X)$ can be regarded as a subset of the set $\mathcal{Q}(X)$ introduced in [5, Chapter 4.7], and while the topology of $\mathcal{L}(X)$ does not induce the topology on $\mathcal{Q}(X)$ therein considered, the two topologies coincide on $\mathcal{W}(X)$.

6. Second order Sanov theorem

In this section we give a simple application of the results in Section 3. Sanov theorem states that, if the random variables $(x_i)_{i \in \mathbb{N}}$ are i.i.d. with law $\mu \in \mathcal{P}(X)$,

then the law of their empirical measure satisfies a LDP with speed (n) and rate $H(\cdot|\mu)$. The result also holds if the law μ_n of the random variables depends on n, provided $\mu_n \to \mu$. However, if (μ_n) concentrates in the sense of (5.1), the LD can admit a nontrivial "second order" expansion. To fix ideas, suppose that $\mu_n \to \mu = \delta_x$ for some $x \in X$. Then the functional $H(\cdot|\delta_x)$ is trivial (in the sense of Section 5), and the speed (n) is not the interesting one. More in general, several non-trivial LDPs may hold, as shown in the following theorem (this result already appeared in the literature in the context of diffusion processes [4], and closer to the framework of this paper in [9]).

Theorem 6.1. *Let (μ_n) be a sequence converging to μ in $\mathcal{P}(X)$, and define the empirical measure $\pi_n \colon X^n \to \mathcal{P}(X)$ as*

$$\pi_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}.$$

Then the law $\mathbb{P}_n := \mu_n^{\otimes n} \circ \pi_n^{-1}$ of π_n under the product measure $\mu_n^{\otimes n}$ satisfies a LDP on $\mathcal{P}(X)$ with speed (n) and rate $H(\cdot|\mu)$.

Assume furthermore that (μ_n) satisfies a LDP on $\mathcal{P}(X)$ with speed (a_n) and lower semicontinuous, coercive rate $I \colon \mathcal{P}(X) \to [0, +\infty]$. Then the law of π_n satisfies a LDP on $\mathcal{P}(X)$ with speed $(n\, a_n)$ and lower semicontinuous, coercive rate $\mathcal{I} \colon \mathcal{P}(X) \to [0, +\infty]$

$$\mathcal{I}(\nu) := \int_X \nu(dx)\, I(x).$$

Proof. Fix $\nu \in \mathcal{P}(X)$.
Lower bound with speed (n). By Proposition 3.2-(H), there exists $\nu_n \to \nu$ such that $\overline{\lim}_n H(\nu_n|\mu_n) \leq H(\nu|\mu_n)$. Take $\mathbb{Q}_n := \nu_n^{\otimes n} \circ \pi_n^{-1}$. Then $\mathbb{Q}_n \to \delta_\nu$ and

$$\frac{1}{n} H(\mathbb{Q}_n|\mathbb{P}_n) = \frac{1}{n} H\big(\nu_n^{\otimes n} \circ \pi_n^{-1}|\mu_n^{\otimes n} \circ \pi_n^{-1}\big) \leq \frac{1}{n} H\big(\nu_n^{\otimes n}|\mu_n^{\otimes n}\big) = H\big(\nu_n|\mu_n\big) \quad (6.1)$$

so that we conclude by Theorem 3.4-(H1).
Lower bound with speed $(n\, a_n)$. By Theorem 3.4-(H2), there exists $\nu_n \to \nu$ such that $\overline{\lim}_n \frac{1}{a_n} H(\nu_n|\mu_n) \leq \mathcal{I}(\nu)$. Take $\mathbb{Q}_n := \nu_n^{\otimes n} \circ \pi_n^{-1}$. By the same calculation as in (6.1) and Theorem 3.4-(H1) we conclude.
Weak upper bound with speed (n). Let $(\mathbb{Q}_n)$ be a sequence in $\mathcal{P}(\mathcal{P}(X))$ such that $\mathbb{Q}_n \to \delta_\nu$. We want to prove $\underline{\lim}_n H(\mathbb{Q}_n|\mathbb{P}_n) \geq H(\nu|\mu)$. One can assume $\mathbb{Q}_n = \gamma_n \circ \pi_n^{-1}$ for some $\gamma_n \in \mathcal{P}(X^n)$ the relative entropy being infinite otherwise, see (3.5). Since $\pi_n(\mathbf{x}) = \pi_n(\mathbf{x}')$ iff $\mathbf{x}'$ is obtained from $\mathbf{x}$ by an index permutation, γ_n can be assumed invariant under index permutation as well (see (3.6)), to obtain

$$H(\mathbb{Q}_n|\mathbb{P}_n) = H\big(\gamma_n \circ \pi_n^{-1}|\mu_n^{\otimes n} \circ \pi_n^{-1}\big) = H\big(\gamma_n|\mu_n^{\otimes n}\big).$$

Let $\nu_n \in \mathcal{P}(X)$ be the one-dimensional marginal of γ_n. By the explicit representation (3.3) of the relative entropy and its convexity, reasoning as in (4.11)

$$H(\mathbb{Q}_n|\mathbb{P}_n) = H\big(\gamma_n|\mu_n^{\otimes n}\big) \geq H\big(\nu_n^{\otimes n}|\mu_n^{\otimes n}\big) = n\, H(\nu_n|\mu_n).$$

On the other hand $\mathbb{Q}_n \to \delta_\nu$ implies $\nu_n \to \nu$, so that by Proposition 3.2-(H)

$$\varliminf_n \tfrac{1}{n} H(\mathbb{Q}_n | \mathbb{P}_n) \geq \varliminf_n H(\nu_n | \mu_n) \geq H(\nu | \mu).$$

Weak upper bound with speed $(n\, a_n)$. Following the same strategy of the bound with speed (n), we obtain

$$\varliminf_n \tfrac{1}{n\, a_n} H(\mathbb{Q}_n | \mathbb{P}_n) \geq \varliminf_n \tfrac{1}{a_n} H(\nu_n | \mu_n) \geq \int \nu(dx)\, I(x),$$

where in the last inequality we used Theorem 3.5-(H2). We conclude by applying Theorem 3.5-(H1) to $(\mathbb{Q}_n)$.

Exponential tightness. With the same notation of the upper bound proofs, if $\frac{1}{n} H(\mathbb{Q}_n | \mathbb{P}_n)$ is uniformly bounded then $H(\nu_n | \mu_n)$ also is. Since μ_n is tight, Proposition 3.1 implies that ν_n is tight and thus $\mathbb{Q}_n$ is tight as well. By Theorem 3.3 we conclude that $(\mathbb{P}_n)$ is exponentially tight with speed (n). Since (μ_n) satisfies a LDP with speed $(n\, a_n)$ and I is coercive, (μ_n) is exponentially tight with this speed [5, Ex. 4.1.10], and the same proof yields that $(\mathbb{P}_n)$ is exponentially tight with speed $(n\, a_n)$. $\qquad\square$

An example of application of Theorem 6.1, is the extension of well known results about LD for the empirical measure of independent random walks or diffusion processes. Let $(X_i^n)_{i \in \mathbb{N}}$ be a family of stochastic process $X_i^n \in D([0, +\infty[; \mathbb{R}^d)$, all starting at $\mathbf{0}$ (the case of different initial conditions for each X_i^n could also be fitted in this framework, but we keep the notation simple). If X_i^n converges in law to some limit process X (*e.g.*, X_i^n is a parabolically-rescaled symmetric random walk converging to a brownian motion or a Levy process in case of heavy tails), then the LD of the empirical measure π_n happen with speed (n), for instance the results in [8] can be recovered from the first part of Theorem 6.1 by a contraction principle. However, if X_i^n converges to a deterministic trajectory (*e.g.*, X_i^n is a hyperbolically-rescaled asymmetric random walk, converging to a uniform motion), then the LD happen with a faster speed. For instance one can recover the results in [7] by the second part of Theorem 6.1 by a contraction principle.

A most interesting open problem related to the above framework is the analysis of the LD for the empirical measure of a totally asymmetric simple exclusion process on $\mathbb{Z}$ (TASEP). Indeed, the law of the path of a particle X_i is independent of the law of the other particles conditionally to the path of the particle at its right (provided the TASEP moves right). TASEP would therefore fit in the framework of Theorem 6.1, except that (x_i) is now a Markov chain (not an i.i.d. sequence), and μ_n is replaced by a jump kernel $\mu_n(x, dy)$. In this case, one still expects the presence of multiple non-trivial speeds for the LD in the same fashion of Theorem 6.1; however the Markov equivalent of Theorem 6.1 features a richer description, and it is still subject of investigation.

References

[1] P. BILLINGSLEY, "Convergence of Probability Measures", New York, John Wiley and Sons, 1999, 2nd Edition.

[2] G. DAL MASO, "An Introduction to Γ-Convergence", Progress in Nonlinear Differential Equations and their Applications, Vol. 8, Birkhauser, 1993.

[3] P. DUPUIS and R. S. ELLIS, "A Weak Convergence Approach to the Theory of Large Deviations", Wiley-Interscience 1997.

[4] D. A. DAWSON and J. GÄRTNER, *Multilevel large deviations and interacting diffusions*, Probab. Theory Related Fields **98** (1994), 423-487.

[5] A. DEMBO and O. ZEITOUNI, "Large Deviations Techniques and Applications", Jones and Bartlett Publishers, 1993.

[6] L. JENSEN, "Large Deviations of the Asymmetric Simple Exclusion Process in One Dimension", PhD Thesis, New York University, 2000.

[7] C. KIPNIS and C. LÉONARD, *Grandes Déviations pour un système hydrodynamique asymétrique de particules indépendentes*, Ann. Inst. H. Poincaré, Probab. Statist. **31** (1995), 233-248.

[8] C. KIPNIS and S. OLLA, *Large deviations from the hydrodynamical limit for a system of independent brownian particles*, Stochastics Stochastics Rep. **33** (1990), 17–25.

[9] C. LÉONARD, *From the Schrödinger problem to the Monge-Kantorovich problem*, J. Funct. Anal. **262** (2012), 1879-1920.

[10] M. MARIANI, *Large deviations principles for stochastic scalar conservation laws*, Probab. Theory Related Fields **147** (2010), 607–648.

[11] R. T. ROCKAFELLAR and T. WETS, "Variational Analysis", Springer-Verlag, 1998.

[12] D. W. STROOCK, "Probability Theory. An Analytic View", Cambridge University Press, 1993.

Faculty of Mathematics
National Research University
Higher School of Economics
6 Usacheva St.
119048 Moscow, Russia
mmariani@hse.ru