Chapter 4

General Principles

In this chapter, we initiate the investigation of large deviation principles (LDPs) for families of measures on general spaces. As will be obvious in subsequent chapters, the objects on which the LDP is sought may vary considerably. Hence, it is necessary to undertake a study of the LDP in an abstract setting. We shall focus our attention on the abstract statement of the LDP as presented in Section 1.2 and give conditions for the existence of such a principle and various approaches for the identification of the resulting rate function.

Since this chapter deals with different approaches to the LDP, some of its sections are independent of the others. A rough structure of it is as follows. In Section 4.1, extensions of the basic properties of the LDP are provided. In particular, relations between the topological structure of the space, the existence of certain limits, and the existence and uniqueness of the LDP are explored. Section 4.2 describes how to move around the LDP from one space to another. Thus, under appropriate conditions, the LDP can be proved in a simple situation and then effortlessly transferred to a more complex one. Of course, one should not be misled by the word *effortlessly*: It often occurs in applications that much of the technical work to be done is checking that the conditions for such a transformation are satisfied! Sections 4.3 and 4.4 investigate the relation between the LDP and the computation of exponential integrals. Although in some applications the computation of the exponential integrals is a goal in itself, it is more often the case that such computations are an intermediate step in deriving the LDP. Such a situation has already been described, though implicitly, in the treatment of the Chebycheff upper bound in Section 2.2. This line of thought is tackled again in Section 4.5.1, in the case where $\mathcal{X}$ is a topological vector space, such that the Fenchel–Legendre transform is well-defined

A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications*, Stochastic Modelling and Applied Probability 38,
DOI 10.1007/978-3-642-03311-7_4,

and an upper bound may be derived based on it. Section 4.5.2 complements this approach by providing the tools that will enable us to exploit convexity, again in the case of topological vector spaces, to derive the lower bound. The attentive reader may have already suspected that such an attack on the LDP is possible when he followed the arguments of Section 2.3. Section 4.6 is somewhat independent of the rest of the chapter. Its goal is to show that the LDP is preserved under projective limits. Although at first sight this may not look useful for applications, it will become clear to the patient reader that this approach is quite general and may lead from finite dimensional computations to the LDP in abstract spaces. Finally, Section 4.7 draws attention to the similarity between the LDP and weak convergence in metric spaces.

Since this chapter deals with the LDP in abstract spaces, some topological and analytical preliminaries are in order. The reader may find Appendices B and C helpful reminders of a particular definition or theorem.

The convention that $\mathcal{B}$ contains the Borel σ-field $\mathcal{B}_{\mathcal{X}}$ is used throughout this chapter, except in Lemma 4.1.5, Theorem 4.2.1, Exercise 4.2.9, Exercise 4.2.32, and Section 4.6.

4.1 Existence of an LDP and Related Properties

If a set $\mathcal{X}$ is given the coarse topology $\{\emptyset, \mathcal{X}\}$, the only information implied by the LDP is that $\inf_{x\in\mathcal{X}} I(x) = 0$, and many rate functions satisfy this requirement. To avoid such trivialities, we must put some constraint on the topology of the set $\mathcal{X}$. Recall that a topological space is Hausdorff if, for every pair of distinct points x and y, there exist disjoint neighborhoods of x and y. The natural condition that prevails throughout this book is that, in addition to being Hausdorff, $\mathcal{X}$ is a *regular space* as defined next.

Definition 4.1.1 *A Hausdorff topological space $\mathcal{X}$ is regular if, for any closed set $F \subset \mathcal{X}$ and any point $x \notin F$, there exist disjoint open subsets G_1 and G_2 such that $F \subset G_1$ and $x \in G_2$.*

In the rest of the book, the term *regular* will mean Hausdorff and regular. The following observations regarding regular spaces are of crucial importance here:
(a) For any neighborhood G of $x \in \mathcal{X}$, there exists a neighborhood A of x such that $\overline{A} \subset G$.
(b) Every metric space is regular. Moreover, if a real topological vector space is Hausdorff, then it is regular. All examples of an LDP considered

in this book are either for metric spaces, or for Hausdorff real topological vector spaces.
(c) A lower semicontinuous function f satisfies, at every point x,

$$f(x) = \sup_{\{G \text{ neighborhood of } x\}} \inf_{y \in G} f(y) . \tag{4.1.2}$$

Therefore, for any $x \in \mathcal{X}$ and any $\delta > 0$, one may find a neighborhood $G = G(x, \delta)$ of x, such that $\inf_{y \in G} f(y) \geq (f(x) - \delta) \wedge 1/\delta$. Let $A = A(x, \delta)$ be a neighborhood of x such that $\overline{A} \subset G$. (Such a set exists by property (a).) One then has

$$\inf_{y \in \overline{A}} f(y) \geq \inf_{y \in G} f(y) \geq (f(x) - \delta) \wedge \frac{1}{\delta} . \tag{4.1.3}$$

The sets $G = G(x, \delta)$ frequently appear in the proofs of large deviations statements and properties. Observe that in a metric space, $G(x, \delta)$ may be taken as a ball centered at x and having a small enough radius.

4.1.1 Properties of the LDP

The first desirable consequence of the assumption that $\mathcal{X}$ is a regular topological space is the uniqueness of the rate function associated with the LDP.

Lemma 4.1.4 *A family of probability measures $\{\mu_\epsilon\}$ on a regular topological space can have at most one rate function associated with its LDP.*

Proof: Suppose there exist two rate functions $I_1(\cdot)$ and $I_2(\cdot)$, both associated with the LDP for $\{\mu_\epsilon\}$. Without loss of generality, assume that for some $x_0 \in \mathcal{X}$, $I_1(x_0) > I_2(x_0)$. Fix $\delta > 0$ and consider the open set A for which $x_0 \in A$, while $\inf_{y \in \overline{A}} I_1(y) \geq (I_1(x_0) - \delta) \wedge 1/\delta$. Such a set exists by (4.1.3). It follows by the LDP for $\{\mu_\epsilon\}$ that

$$-\inf_{y \in \overline{A}} I_1(y) \geq \limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A) \geq \liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A) \geq -\inf_{y \in A} I_2(y) .$$

Therefore,

$$I_2(x_0) \geq \inf_{y \in A} I_2(y) \geq \inf_{y \in \overline{A}} I_1(y) \geq (I_1(x_0) - \delta) \wedge \frac{1}{\delta} .$$

Since δ is arbitrary, this contradicts the assumption that $I_1(x_0) > I_2(x_0)$. □

Remarks:
(a) It is evident from the proof that if $\mathcal{X}$ is a locally compact space (e.g.,

$\mathcal{X} = \mathbb{R}^d$), the rate function is unique as soon as a weak LDP holds. As shown in Exercise 4.1.30, if $\mathcal{X}$ is a Polish space, then also the rate function is unique as soon as a weak LDP holds.
(b) The uniqueness of the rate function does not depend on the Hausdorff part of the definition of regular spaces. However, the rate function assigns the same value to any two points of $\mathcal{X}$ that are not separated. (See Exercise 4.1.9.) Thus, in terms of the LDP, such points are indistinguishable.

As shown in the next lemma, the LDP is preserved under suitable inclusions. Hence, in applications, one may first prove an LDP in a space that possesses additional structure (for example, a topological vector space), and then use this lemma to deduce the LDP in the subspace of interest.

Lemma 4.1.5 *Let $\mathcal{E}$ be a measurable subset of $\mathcal{X}$ such that $\mu_\epsilon(\mathcal{E}) = 1$ for all $\epsilon > 0$. Suppose that $\mathcal{E}$ is equipped with the topology induced by $\mathcal{X}$.*
(a) If $\mathcal{E}$ is a closed subset of $\mathcal{X}$ and $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{E}$ with rate function I, then $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{X}$ with rate function I' such that $I' = I$ on $\mathcal{E}$ and $I' = \infty$ on $\mathcal{E}^c$.
(b) If $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{X}$ with rate function I and $\mathcal{D}_I \subset \mathcal{E}$, then the same LDP holds in $\mathcal{E}$. In particular, if $\mathcal{E}$ is a closed subset of $\mathcal{X}$, then $\mathcal{D}_I \subset \mathcal{E}$ and hence the LDP holds in $\mathcal{E}$.

Proof: In the topology induced on $\mathcal{E}$ by $\mathcal{X}$, the open sets are the sets of the form $G \cap \mathcal{E}$ with $G \subseteq \mathcal{X}$ open. Similarly, the closed sets in this topology are the sets of the form $F \cap \mathcal{E}$ with $F \subseteq \mathcal{X}$ closed. Furthermore, $\mu_\epsilon(\Gamma) = \mu_\epsilon(\Gamma \cap \mathcal{E})$ for any $\Gamma \in \mathcal{B}$.
(a) Suppose that an LDP holds in $\mathcal{E}$, which is a closed subset of $\mathcal{X}$. Extend the rate function I to be a lower semicontinuous function on $\mathcal{X}$ by setting $I(x) = \infty$ for any $x \in \mathcal{E}^c$. Thus, $\inf_{x\in\Gamma} I(x) = \inf_{x\in\Gamma\cap\mathcal{E}} I(x)$ for any $\Gamma \subset \mathcal{X}$ and the large deviations lower (upper) bound holds.
(b) Suppose that an LDP holds in $\mathcal{X}$. If $\mathcal{E}$ is closed, then $\mathcal{D}_I \subset \mathcal{E}$ by the large deviations lower bound (since $\mu_\epsilon(\mathcal{E}^c) = 0$ for all $\epsilon > 0$ and $\mathcal{E}^c$ is open). Now, $\mathcal{D}_I \subset \mathcal{E}$ implies that $\inf_{x\in\Gamma} I(x) = \inf_{x\in\Gamma\cap\mathcal{E}} I(x)$ for any $\Gamma \subset \mathcal{X}$ and the large deviations lower (upper) bound holds for all measurable subsets of $\mathcal{E}$. Further, since the level sets $\Psi_I(\alpha)$ are closed subsets of $\mathcal{E}$, the rate function I remains lower semicontinuous when restricted to $\mathcal{E}$. $\square$

Remarks:
(a) The preceding lemma also holds for the weak LDP, since compact subsets of $\mathcal{E}$ are just the compact subsets of $\mathcal{X}$ contained in $\mathcal{E}$. Similarly, under the assumptions of the lemma, I is a good rate function on $\mathcal{X}$ iff it is a good rate function when restricted to $\mathcal{E}$.
(b) Lemma 4.1.5 holds without any change in the proof even when $\mathcal{B}_\mathcal{X} \not\subseteq \mathcal{B}$.

The following is an important property of good rate functions.

Lemma 4.1.6 *Let I be a good rate function.*
(a) Let $\{F_\delta\}_{\delta>0}$ be a nested family of closed sets, i.e., $F_\delta \subseteq F_{\delta'}$ if $\delta < \delta'$. Define $F_0 = \cap_{\delta>0} F_\delta$. Then

$$\inf_{y \in F_0} I(y) = \lim_{\delta \to 0} \inf_{y \in F_\delta} I(y) .$$

(b) Suppose $(\mathcal{X}, d)$ is a metric space. Then, for any set A,

$$\inf_{y \in \overline{A}} I(y) = \lim_{\delta \to 0} \inf_{y \in A^\delta} I(y) , \tag{4.1.7}$$

where

$$A^\delta \stackrel{\triangle}{=} \{y : \ d(y, A) = \inf_{z \in A} d(y, z) \le \delta\} \tag{4.1.8}$$

denotes the closed blowup of A.

Proof: (a) Since $F_0 \subseteq F_\delta$ for all $\delta > 0$, it suffices to prove that for all $\eta > 0$,

$$\gamma \stackrel{\triangle}{=} \lim_{\delta \to 0} \inf_{y \in F_\delta} I(y) \ge \inf_{y \in F_0} I(y) - \eta .$$

This inequality holds trivially when $\gamma = \infty$. If $\gamma < \infty$, fix $\eta > 0$ and let $\alpha = \gamma + \eta$. The sets $F_\delta \cap \Psi_I(\alpha)$, $\delta > 0$, are non-empty, nested, and compact. Consequently,

$$F_0 \bigcap \Psi_I(\alpha) = \bigcap_{\delta>0} F_\delta \bigcap \Psi_I(\alpha)$$

is also non-empty, and the proof of part (a) is thus completed.
(b) Note that $d(\cdot, A)$ is a continuous function and hence $\{A^\delta\}_{\delta>0}$ are nested, closed sets. Moreover,

$$\bigcap_{\delta>0} A^\delta = \{y : \ d(y, A) = 0\} = \overline{A} . \qquad \square$$

Exercise 4.1.9 Suppose that for any closed subset F of $\mathcal{X}$ and any point $x \notin F$, there exist two disjoint open sets G_1 and G_2 such that $F \subset G_1$ and $x \in G_2$. Prove that if $I(x) \neq I(y)$ for some lower semicontinuous function I, then there exist disjoint neighborhoods of x and y.

Exercise 4.1.10 [[LyS87], Lemma 2.6. See also [Puk91], Theorem (P).]
Let $\{\mu_n\}$ be a sequence of probability measures on a Polish space $\mathcal{X}$.
(a) Show that $\{\mu_n\}$ is exponentially tight if for every $\alpha < \infty$ and every $\eta > 0$, there exist $m \in \mathbb{Z}_+$ and $x_1, \ldots, x_m \in \mathcal{X}$ such that for all n,

$$\mu_n\Big(\Big[\bigcup_{i=1}^m B_{x_i,\eta}\Big]^c\Big) \le e^{-\alpha n} .$$

Hint: Observe that for every sequence $\{m_k\}$ and any $x_i^{(k)} \in \mathcal{X}$, the set $\cap_{k=1}^{\infty} \cup_{i=1}^{m_k} B_{x_i^{(k)},1/k}$ is pre-compact.
(b) Suppose that $\{\mu_n\}$ satisfies the large deviations upper bound with a good rate function. Show that for every countable dense subset of $\mathcal{X}$, e.g., $\{x_i\}$, every $\eta > 0$, every $\alpha < \infty$, and every m large enough,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n\Big(\Big[\bigcup_{i=1}^{m} B_{x_i,\eta} \Big]^c \Big) < -\alpha \,.$$

Hint: Use Lemma 4.1.6.
(c) Deduce that if $\{\mu_n\}$ satisfies the large deviations upper bound with a good rate function, then $\{\mu_n\}$ is exponentially tight.

Remark: When a non-countable family of measures $\{\mu_\epsilon, \epsilon > 0\}$ satisfies the large deviations upper bound in a Polish space with a good rate function, the preceding yields the exponential tightness of every sequence $\{\mu_{\epsilon_n}\}$, where $\epsilon_n \to 0$ as $n \to \infty$. As far as large deviations results are concerned, this is indistinguishable from exponential tightness of the whole family.

4.1.2 The Existence of an LDP

The following theorem introduces a general, indirect approach for establishing the *existence* of a *weak LDP*.

Theorem 4.1.11 *Let $\mathcal{A}$ be a base of the topology of $\mathcal{X}$. For every $A \in \mathcal{A}$, define*

$$\mathcal{L}_A \stackrel{\triangle}{=} -\liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A) \tag{4.1.12}$$

and

$$I(x) \stackrel{\triangle}{=} \sup_{\{A\in\mathcal{A}:\, x\in A\}} \mathcal{L}_A \ . \tag{4.1.13}$$

Suppose that for all $x \in \mathcal{X}$,

$$I(x) = \sup_{\{A\in\mathcal{A}:\, x\in A\}} \Big[-\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A) \Big] \,. \tag{4.1.14}$$

Then μ_ϵ satisfies the weak LDP with the rate function $I(x)$.

Remarks:
(a) Observe that condition (4.1.14) holds when the limits $\lim_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A)$ exist for all $A \in \mathcal{A}$ (with $-\infty$ as a possible value).
(b) When $\mathcal{X}$ is a locally convex, Hausdorff topological vector space, the base $\mathcal{A}$ is often chosen to be the collection of open, convex sets. For concrete examples, see Sections 6.1 and 6.3.

Proof: Since $\mathcal{A}$ is a base for the topology of $\mathcal{X}$, for any open set G and any point $x \in G$ there exists an $A \in \mathcal{A}$ such that $x \in A \subset G$. Therefore, by definition,

$$\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(G) \geq \liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A) = -\mathcal{L}_A \geq -I(x)\,.$$

As seen in Section 1.2, this is just one of the alternative statements of the large deviations lower bound.

Clearly, $I(x)$ is a nonnegative function. Moreover, if $I(x) > \alpha$, then $\mathcal{L}_A > \alpha$ for some $A \in \mathcal{A}$ such that $x \in A$. Therefore, $I(y) \geq \mathcal{L}_A > \alpha$ for every $y \in A$. Hence, the sets $\{x : I(x) > \alpha\}$ are open, and consequently I is a rate function.

Note that the lower bound and the fact that I is a rate function do not depend on (4.1.14). This condition is used in the proof of the upper bound. Fix $\delta > 0$ and a compact $F \subset \mathcal{X}$. Let I^δ be the δ-rate function, i.e., $I^\delta(x) \stackrel{\triangle}{=} \min\{I(x) - \delta, 1/\delta\}$. Then, (4.1.14) implies that for every $x \in F$, there exists a set $A_x \in \mathcal{A}$ (which may depend on δ) such that $x \in A_x$ and

$$-\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A_x) \geq I^\delta(x)\,.$$

Since F is compact, one can extract from the open cover $\cup_{x \in F} A_x$ of F a finite cover of F by the sets $A_{x_1}, \ldots, A_{x_m}$. Thus,

$$\mu_\epsilon(F) \leq \sum_{i=1}^{m} \mu_\epsilon(A_{x_i})\,,$$

and consequently,

$$\begin{aligned}
\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(F) &\leq \max_{i=1,\ldots,m} \limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A_{x_i}) \\
&\leq -\min_{i=1,\ldots,m} I^\delta(x_i) \leq -\inf_{x \in F} I^\delta(x)\,.
\end{aligned}$$

The proof of the upper bound for compact sets is completed by considering the limit as $\delta \to 0$. □

Theorem 4.1.11 is extended in the following lemma, which concerns the LDP of a family of probability measures $\{\mu_{\epsilon,\sigma}\}$ that is indexed by an additional parameter σ. For a concrete application, see Section 6.3, where σ is the initial state of a Markov chain.

Lemma 4.1.15 *Let $\mu_{\epsilon,\sigma}$ be a family of probability measures on $\mathcal{X}$, indexed by σ, whose range is the set Σ. Let $\mathcal{A}$ be a base for the topology of $\mathcal{X}$. For each $A \in \mathcal{A}$, define*

$$\mathcal{L}_A \stackrel{\triangle}{=} -\liminf_{\epsilon \to 0} \epsilon \log\Big[\inf_{\sigma \in \Sigma} \mu_{\epsilon,\sigma}(A)\Big]\,. \tag{4.1.16}$$

Let

$$I(x) = \sup_{\{A\in\mathcal{A}:\, x\in A\}} \mathcal{L}_A \,.$$

If for every $x \in \mathcal{X}$,

$$I(x) = \sup_{\{A\in\mathcal{A}:\, x\in A\}} \left\{ -\limsup_{\epsilon\to 0} \epsilon \log\Big[\sup_{\sigma\in\Sigma} \mu_{\epsilon,\sigma}(A)\Big]\right\} \,, \tag{4.1.17}$$

then, for each $\sigma \in \Sigma$*, the measures* $\mu_{\epsilon,\sigma}$ *satisfy a weak LDP with the (same) rate function* $I(\cdot)$*.*

Proof: The proof parallels that of Theorem 4.1.11. (See Exercise 4.1.29.) $\square$

It is aesthetically pleasing to know that the following partial converse of Theorem 4.1.11 holds.

Theorem 4.1.18 *Suppose that* $\{\mu_\epsilon\}$ *satisfies the LDP in a regular topological space* $\mathcal{X}$ *with rate function* I*. Then, for any base* $\mathcal{A}$ *of the topology of* $\mathcal{X}$*, and for any* $x \in \mathcal{X}$*,*

$$\begin{aligned} I(x) &= \sup_{\{A\in\mathcal{A}:\, x\in A\}} \Big\{ -\liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A) \Big\} \\ &= \sup_{\{A\in\mathcal{A}:\, x\in A\}} \left\{ -\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A) \right\} \,. \end{aligned} \tag{4.1.19}$$

Remark: As shown in Exercise 4.1.30, for a Polish space $\mathcal{X}$ suffices to assume in Theorem 4.1.18 that $\{\mu_\epsilon\}$ satisfies the weak LDP. Consequently, by Theorem 4.1.11, in this context (4.1.19) is *equivalent* to the weak LDP.

Proof: Fix $x \in \mathcal{X}$ and let

$$\ell(x) = \sup_{\{A\in\mathcal{A}:\, x\in A\}} \inf_{y\in\overline{A}} I(y) \,. \tag{4.1.20}$$

Suppose that $I(x) > \ell(x)$. Then, in particular, $\ell(x) < \infty$ and $x \in \Psi_I(\alpha)^c$ for some $\alpha > \ell(x)$. Since $\Psi_I(\alpha)^c$ is an open set and $\mathcal{A}$ is a base for the topology of the regular space $\mathcal{X}$, there exists a set $A \in \mathcal{A}$ such that $x \in A$ and $\overline{A} \subseteq \Psi_I(\alpha)^c$. Therefore, $\inf_{y\in\overline{A}} I(y) \geq \alpha$, which contradicts (4.1.20). We conclude that $\ell(x) \geq I(x)$. The large deviations lower bound implies

$$I(x) \geq \sup_{\{A\in\mathcal{A}:\, x\in A\}} \Big\{ -\liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A) \Big\} \,,$$

while the large deviations upper bound implies that for all $A \in \mathcal{A}$,

$$-\liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A) \geq -\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(\overline{A}) \geq \inf_{y\in\overline{A}} I(y) \,.$$

These two inequalities yield (4.1.19), since $\ell(x) \geq I(x)$. □

The characterization of the rate function in Theorem 4.1.11 and Lemma 4.1.15 involves the supremum over a large collection of sets. Hence, it does not yield a convenient explicit formula. As shown in Section 4.5.2, if $\mathcal{X}$ is a Hausdorff topological vector space, this rate function can sometimes be identified with the Fenchel–Legendre transform of a limiting logarithmic moment generating function. This approach requires an *a priori* proof that the rate function is *convex*. The following lemma improves on Theorem 4.1.11 by giving a sufficient condition for the convexity of the rate function. Throughout, for any sets $A_1, A_2 \in \mathcal{X}$,

$$\frac{A_1 + A_2}{2} \triangleq \{x : \; x = (x_1 + x_2)/2, \, x_1 \in A_1, x_2 \in A_2\} \, .$$

Lemma 4.1.21 *Let $\mathcal{A}$ be a base for a Hausdorff topological vector space $\mathcal{X}$, such that in addition to condition (4.1.14), for every $A_1, A_2 \in \mathcal{A}$,*

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon \left(\frac{A_1 + A_2}{2} \right) \geq -\frac{1}{2} \left(\mathcal{L}_{A_1} + \mathcal{L}_{A_2} \right) . \tag{4.1.22}$$

Then the rate function I of (4.1.13), which governs the weak LDP associated with $\{\mu_\epsilon\}$, is convex.

Proof: It suffices to show that the condition (4.1.22) yields the convexity of the rate function I of (4.1.13). To this end, fix $x_1, x_2 \in \mathcal{X}$ and $\delta > 0$. Let $x = (x_1 + x_2)/2$ and let I^δ denote the δ-rate function. Then, by (4.1.14), there exists an $A \in \mathcal{A}$ such that $x \in A$ and $-\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A) \geq I^\delta(x)$. The pair (x_1, x_2) belongs to the set $\{(y_1, y_2) : \; (y_1 + y_2)/2 \in A\}$, which is an open subset of $\mathcal{X} \times \mathcal{X}$. Therefore, there exist open sets $A_1 \subseteq \mathcal{X}$ and $A_2 \subseteq \mathcal{X}$ with $x_1 \in A_1$ and $x_2 \in A_2$ such that $(A_1 + A_2)/2 \subseteq A$. Furthermore, since $\mathcal{A}$ is a base for the topology of $\mathcal{X}$, one may take A_1 and A_2 in $\mathcal{A}$. Thus, our assumptions imply that

$$\begin{aligned} -I^\delta(x) \quad &\geq \quad \limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A) \\ &\geq \quad \limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon \left(\frac{A_1 + A_2}{2} \right) \geq -\frac{1}{2} \left(\mathcal{L}_{A_1} + \mathcal{L}_{A_2} \right) . \end{aligned}$$

Since $x_1 \in A_1$ and $x_2 \in A_2$, it follows that

$$\frac{1}{2} I(x_1) + \frac{1}{2} I(x_2) \geq \frac{1}{2} \mathcal{L}_{A_1} + \frac{1}{2} \mathcal{L}_{A_2} \geq I^\delta(x) = I^\delta \left(\frac{1}{2} x_1 + \frac{1}{2} x_2 \right) .$$

Considering the limit $\delta \searrow 0$, one obtains

$$\frac{1}{2} I(x_1) + \frac{1}{2} I(x_2) \geq I \left(\frac{1}{2} x_1 + \frac{1}{2} x_2 \right) .$$

By iterating, this inequality can be extended to any x of the form $(k/2^n)x_1 + (1-k/2^n)x_2$ with $k, n \in \mathbb{Z}_+$. The definition of a topological vector space and the lower semicontinuity of I imply that $I(\beta x_1 + (1-\beta)x_2) : [0,1] \to [0,\infty]$ is a lower semicontinuous function of β. Hence, the preceding inequality holds for all convex combinations of x_1, x_2 and the proof of the lemma is complete. $\square$

When combined with exponential tightness, Theorem 4.1.11 implies the following large deviations analog of Prohorov's theorem (Theorem D.9).

Lemma 4.1.23 *Suppose the topological space $\mathcal{X}$ has a countable base. For any family of probability measures $\{\mu_\epsilon\}$, there exists a sequence $\epsilon_k \to 0$ such that $\{\mu_{\epsilon_k}\}$ satisfies the weak LDP in $\mathcal{X}$. If $\{\mu_\epsilon\}$ is an exponentially tight family of probability measures, then $\{\mu_{\epsilon_k}\}$ also satisfies the LDP with a good rate function.*

Proof: Fix a countable base $\mathcal{A}$ for the topology of $\mathcal{X}$ and a sequence $\epsilon_n \to 0$. By Tychonoff's theorem (Theorem B.3), the product topology makes $\mathcal{Y} = [0,1]^{\mathcal{A}}$ into a compact metrizable space. Since $\mathcal{Y}$ is sequentially compact (Theorem B.2) and $\mu_\epsilon(\cdot)^\epsilon : \mathcal{A} \to [0,1]$ is in $\mathcal{Y}$ for each $\epsilon > 0$, the sequence $\mu_{\epsilon_n}(\cdot)^{\epsilon_n}$ has a convergent subsequence in $\mathcal{Y}$. Hence, passing to the latter subsequence, denoted ϵ_k, the limits $\lim_{k\to\infty} \epsilon_k \log \mu_{\epsilon_k}(A)$ exist for all $A \in \mathcal{A}$ (with $-\infty$ as a possible value). In particular, condition (4.1.14) holds and by Theorem 4.1.11, $\{\mu_{\epsilon_k} : k \in \mathbb{Z}_+\}$ satisfies the weak LDP. Applying Lemma 1.2.18, the LDP with a good rate function follows when $\{\mu_\epsilon\}$ is an exponentially tight family of probability measures. $\square$

The next lemma applies for tight Borel probability measures μ_ϵ on metric spaces. In this context, it allows replacement of the assumed LDP in either Lemma 4.1.4 or Theorem 4.1.18 by a weak LDP (see Exercise 4.1.30).

Lemma 4.1.24 *Suppose $\{\mu_\epsilon\}$ is a family of tight (Borel) probability measures on a metric space $(\mathcal{X}, d)$, such that the upper bound (1.2.12) holds for all compact sets and some rate function $I(\cdot)$. Then, for any base $\mathcal{A}$ of the topology of $\mathcal{X}$, and for any $x \in \mathcal{X}$,*

$$I(x) \leq \sup_{\{A\in\mathcal{A}:\, x\in A\}} \left\{ -\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A) \right\} . \tag{4.1.25}$$

Proof: We argue by contradiction, fixing a base $\mathcal{A}$ of the metric topology and $x \in \mathcal{X}$ for which (4.1.25) fails. For any $m \in \mathbb{Z}_+$, there exists some $A \in \mathcal{A}$ such that $x \in A \subset B_{x,m^{-1}}$. Hence, for some $\delta > 0$ and any $m \in \mathbb{Z}_+$,

$$\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(B_{x,m^{-1}}) > -I^\delta(x) = -\min\{I(x) - \delta, 1/\delta\} ,$$

implying that for some $\epsilon_m \to 0$,

$$\mu_{\epsilon_m}(B_{x,m^{-1}}) > e^{-I^\delta(x)/\epsilon_m} \qquad \forall m \in \mathbb{Z}_+ \,. \tag{4.1.26}$$

Recall that the probability measures μ_{ϵ_m} are regular (by Theorem C.5), hence in (4.1.26) we may replace each open set $B_{x,m^{-1}}$ by some closed subset F_m. With each μ_{ϵ_m} assumed tight, we may further replace the closed sets F_m by compact subsets $K_m \subset F_m \subset B_{x,m^{-1}}$ such that

$$\mu_{\epsilon_m}(K_m) > e^{-I^\delta(x)/\epsilon_m} \qquad \forall m \in \mathbb{Z}_+ \,. \tag{4.1.27}$$

Note that the sets $K_r^* = \{x\} \cup_{m \geq r} K_m$ are also compact. Indeed, in any open covering of K_r^* there is an open set G_o such that $x \in G_o$ and hence $\cup_{m>m_0} K_m \subset B_{x,m_0^{-1}} \subset G_o$ for some $m_o \in \mathbb{Z}_+$, whereas the compact set $\cup_{m=r}^{m_o} K_m$ is contained in the union of some G_i, $i = 1, \ldots, M$, from this cover. In view of (4.1.27), the upper bound (1.2.12) yields for $K_r^* \subset B_{x,r^{-1}}$ that,

$$\begin{aligned} -\inf_{y \in B_{x,r^{-1}}} I(y) \geq -\inf_{y \in K_r^*} I(y) &\geq \limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(K_r^*) \\ &\geq \limsup_{m \to \infty} \epsilon_m \log \mu_{\epsilon_m}(K_m) \geq -I^\delta(x) \,. \end{aligned} \tag{4.1.28}$$

By lower semicontinuity, $\lim_{r\to\infty} \inf_{y \in B_{x,r^{-1}}} I(y) = I(x) > I^\delta(x)$, in contradiction with (4.1.28). Necessarily, (4.1.25) holds for any $x \in \mathcal{X}$ and any base $\mathcal{A}$. □

Exercise 4.1.29 Prove Lemma 4.1.15 using the following steps.
(a) Check that the large deviations lower bound (for each $\sigma \in \Sigma$) and the lower semicontinuity of I may be proved exactly as done in Theorem 4.1.11.
(b) Fix $\sigma \in \Sigma$ and prove the large deviations upper bound for compact sets.

Exercise 4.1.30 Suppose a family of tight (Borel) probability measures $\{\mu_\epsilon\}$ satisfies the weak LDP in a metric space $(\mathcal{X}, d)$ with rate function $I(\cdot)$.
(a) Combine Lemma 4.1.24 with the large deviations lower bound to conclude that (4.1.19) holds for any base $\mathcal{A}$ of the topology of $\mathcal{X}$ and any $x \in \mathcal{X}$.
(b) Conclude that in this context the rate function $I(\cdot)$ associated with the weak LDP is unique.

Exercise 4.1.31 Suppose $X_i \in \mathbb{R}^{d-1}$, $d \geq 2$, with $|X_i| \leq C$ and $Y_i \in [m, M]$ for some $0 < m < M, C < \infty$, are such that $n^{-1}\sum_{i=1}^n (X_i, Y_i)$ satisfy the LDP in $\mathbb{R}^d$ with a good rate function $J(x, y)$. Let $\tau_\epsilon = \inf\{n : \sum_{i=1}^n Y_i > \epsilon^{-1}\}$. Show that $(\epsilon \sum_{i=1}^{\tau_\epsilon} X_i, (\epsilon\tau_\epsilon)^{-1})$ satisfies the LDP in $\mathbb{R}^d$ with good rate function $y^{-1}J(xy, y)$.
Hint: A convenient way to handle the move from the random variables

$n^{-1}\sum_{i=1}^{n}(X_i, Y_i)$ to $(\epsilon\sum_{i=1}^{\tau_\epsilon} X_i, (\epsilon\tau_\epsilon)^{-1})$ is in looking at small balls in $\mathbb{R}^d$ and applying the characterization of the weak LDP as in Theorem 4.1.11.

Remark: Such transformations appear, for example, in the study of regenerative (or renewal) processes [KucC91, Jia94, PuW97], and of multifractal formalism [Rei95, Zoh96].

Exercise 4.1.32 Suppose the topological space $\mathcal{X}$ has a countable base. Show that for any rate function $I(\cdot)$ such that $\inf_x I(x) = 0$, the LDP with rate function $I(\cdot)$ holds for some family of probability measures $\{\mu_\epsilon\}$ on $\mathcal{X}$.
Hint: For $\mathcal{A}$ a countable base for the topology of $\mathcal{X}$ and each $A \in \mathcal{A}$, let $x_{A,m} \in A$ be such that $I(x_{A,m}) \to \inf_{x\in A} I(x)$ as $m \to \infty$. Let $\mathcal{Y} = \{y_k : k \in \mathbb{Z}_+\}$ denote the countable set $\cup_{A\in\mathcal{A}} \cup_m x_{A,m}$. Check that $\inf_{x\in G} I(x) = \inf_{x\in\mathcal{Y}\cap G} I(x)$ for any open set $G \subset \mathcal{X}$, and try the probability measures μ_ϵ such that $\mu_\epsilon(\{y_k\}) = c_\epsilon^{-1}\exp(-k - I(y_k)/\epsilon)$ for $y_k \in \mathcal{Y}$ and $c_\epsilon = \sum_k \exp(-k - I(y_k)/\epsilon)$.

4.2 Transformations of LDPs

This section is devoted to transformations that preserve the LDP, although, possibly, changing the rate function. Once the LDP with a good rate function is established for μ_ϵ, the basic *contraction principle* yields the LDP for $\mu_\epsilon \circ f^{-1}$, where f is any continuous map. The *inverse contraction principle* deals with f which is the inverse of a continuous bijection, and this is a useful tool for strengthening the topology under which the LDP holds. These two transformations are presented in Section 4.2.1. Section 4.2.2 is devoted to exponentially good approximations and their implications; for example, it is shown that when two families of measures defined on the same probability space are exponentially equivalent, then one can infer the LDP for one family from the other. A direct consequence is Theorem 4.2.23, which extends the contraction principle to "approximately continuous" maps.

4.2.1 Contraction Principles

The LDP is preserved under continuous mappings, as the following elementary theorem shows.

Theorem 4.2.1 (Contraction principle) *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces and $f : \mathcal{X} \to \mathcal{Y}$ a continuous function. Consider a good rate function $I : \mathcal{X} \to [0,\infty]$.*
(a) For each $y \in \mathcal{Y}$, define

$$I'(y) \stackrel{\triangle}{=} \inf\{I(x) : \ x \in \mathcal{X}, \quad y = f(x)\}\,. \tag{4.2.2}$$

Then I' is a good rate function on $\mathcal{Y}$, where as usual the infimum over the empty set is taken as ∞.
(b) If I controls the LDP associated with a family of probability measures $\{\mu_\epsilon\}$ on $\mathcal{X}$, then I' controls the LDP associated with the family of probability measures $\{\mu_\epsilon \circ f^{-1}\}$ on $\mathcal{Y}$.

Proof: (a) Clearly, I' is nonnegative. Since I is a good rate function, for all $y \in f(\mathcal{X})$ the infimum in the definition of I' is obtained at some point of $\mathcal{X}$. Thus, the level sets of I', $\Psi_{I'}(\alpha) \triangleq \{y : \; I'(y) \le \alpha\}$, are

$$\Psi_{I'}(\alpha) = \{f(x) : \; I(x) \le \alpha\} = f(\Psi_I(\alpha)) \, ,$$

where $\Psi_I(\alpha)$ are the corresponding level sets of I. As $\Psi_I(\alpha) \subset \mathcal{X}$ are compact, so are the sets $\Psi_{I'}(\alpha) \subset \mathcal{Y}$.
(b) The definition of I' implies that for any $A \subset \mathcal{Y}$,

$$\inf_{y \in A} I'(y) = \inf_{x \in f^{-1}(A)} I(x) \, . \tag{4.2.3}$$

Since f is continuous, the set $f^{-1}(A)$ is an open (closed) subset of $\mathcal{X}$ for any open (closed) $A \subset \mathcal{Y}$. Therefore, the LDP for $\mu_\epsilon \circ f^{-1}$ follows as a consequence of the LDP for μ_ϵ and (4.2.3). □

Remarks:
(a) This theorem holds even when $\mathcal{B}_{\mathcal{X}} \not\subseteq \mathcal{B}$, since for any (measurable) set $A \subset \mathcal{Y}$, both $\overline{f^{-1}(A)} \subset f^{-1}(\overline{A})$ and $f^{-1}(A^o) \subset (f^{-1}(A))^o$.
(b) Note that the upper and lower bounds implied by part (b) of Theorem 4.2.1 hold even when I is not a good rate function. However, if I is *not* a good rate function, it may happen that I' is not a rate function, as the example $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, $I(x) = 0$, and $f(x) = e^x$ demonstrates.
(c) Theorem 4.2.1 holds as long as f is continuous at every $x \in \mathcal{D}_I$; namely, for every $x \in \mathcal{D}_I$ and every neighborhood G of $f(x) \in \mathcal{Y}$, there exists a neighborhood A of x such that $A \subseteq f^{-1}(G)$. This suggests that the contraction principle may be further extended to cover a certain class of "approximately continuous" maps. Such an extension will be pursued in Theorem 4.2.23.

We remind the reader that in what follows, it is always assumed that $\mathcal{B}_{\mathcal{X}} \subseteq \mathcal{B}$, and therefore open sets are always measurable. The following theorem shows that in the presence of exponential tightness, the contraction principle can be made to work in the reverse direction. This property is extremely useful for strengthening large deviations results from a coarse topology to a finer one.

Theorem 4.2.4 (Inverse contraction principle) *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces. Suppose that $g : \mathcal{Y} \to \mathcal{X}$ is a continuous bijection,* *and that $\{\nu_\epsilon\}$ is an exponentially tight family of probability measures on $\mathcal{Y}$. If $\{\nu_\epsilon \circ g^{-1}\}$ satisfies the LDP with the rate function $I : \mathcal{X} \to [0,\infty]$, then $\{\nu_\epsilon\}$ satisfies the LDP with the good rate function $I'(\cdot) \triangleq I(g(\cdot))$.*

Remarks:
(a) In view of Lemma 4.1.5, it suffices for g to be a continuous injection, for then the exponential tightness of $\{\nu_\epsilon\}$ implies that $\mathcal{D}_I \subseteq g(\mathcal{Y})$ even if the latter is not a closed subset of $\mathcal{X}$.
(b) The requirement that $\mathcal{B}_{\mathcal{Y}} \subseteq \mathcal{B}$ is relaxed in Exercise 4.2.9.

Proof: Note first that for every $\alpha < \infty$, by the continuity of g, the level set $\{y : I'(y) \le \alpha\} = g^{-1}(\Psi_I(\alpha))$ is closed. Moreover, $I' \ge 0$, and hence I' is a rate function. Next, because $\{\nu_\epsilon\}$ is an exponentially tight family, it suffices to prove a weak LDP with the rate function $I'(\cdot)$. Starting with the upper bound, fix an arbitrary compact set $K \subset \mathcal{Y}$ and apply the large deviations upper bound for $\nu_\epsilon \circ g^{-1}$ on the compact set $g(K)$ to obtain

$$\begin{aligned}\limsup_{\epsilon\to 0} \epsilon \log \nu_\epsilon(K) &= \limsup_{\epsilon\to 0} \epsilon \log[\nu_\epsilon \circ g^{-1}(g(K))] \\ &\le -\inf_{x\in g(K)} I(x) = -\inf_{y\in K} I'(y)\,,\end{aligned}$$

which is the specified upper bound for ν_ϵ.

To prove the large deviations lower bound, fix $y \in \mathcal{Y}$ with $I'(y) = I(g(y)) = \alpha < \infty$, and a neighborhood G of y. Since $\{\nu_\epsilon\}$ is exponentially tight, there exists a compact set $K_\alpha \subset \mathcal{Y}$ such that

$$\limsup_{\epsilon\to 0} \epsilon \log \nu_\epsilon(K_\alpha^c) < -\alpha\,. \tag{4.2.5}$$

Because g is a bijection, $K_\alpha^c = g^{-1} \circ g(K_\alpha^c)$ and $g(K_\alpha^c) = g(K_\alpha)^c$. By the continuity of g, the set $g(K_\alpha)$ is compact, and consequently $g(K_\alpha)^c$ is an open set. Thus, the large deviations lower bound for the measures $\{\nu_\epsilon \circ g^{-1}\}$ results in

$$-\inf_{x\in g(K_\alpha^c)} I(x) \le \liminf_{\epsilon\to 0} \epsilon \log \nu_\epsilon(K_\alpha^c) < -\alpha\,.$$

Recall that $I(g(y)) = \alpha$, and thus by the preceding inequality it must be that $y \in K_\alpha$. Since g is a continuous bijection, it is a homeomorphism between the compact sets K_α and $g(K_\alpha)$. Therefore, the set $g(G \cap K_\alpha)$ is a neighborhood of $g(y)$ in the induced topology on $g(K_\alpha) \subset \mathcal{X}$. Hence, there exists a neighborhood G' of $g(y)$ in $\mathcal{X}$ such that

$$G' \subset g(G \cap K_\alpha) \cup g(K_\alpha)^c = g(G \cup K_\alpha^c)\,,$$

where the last equality holds because g is a bijection. Consequently, for every $\epsilon > 0$,

$$\nu_\epsilon(G) + \nu_\epsilon(K_\alpha^c) \ge \nu_\epsilon \circ g^{-1}(G')\,,$$

and by the large deviations lower bound for $\{\nu_\epsilon \circ g^{-1}\}$,

$$\begin{aligned}\max\{\liminf_{\epsilon\to 0}\ \epsilon\log\nu_\epsilon(G),\ \limsup_{\epsilon\to 0}\ \epsilon\log\nu_\epsilon(K_\alpha^c)\}\\ \geq \liminf_{\epsilon\to 0}\ \epsilon\log\{\nu_\epsilon\circ g^{-1}(G')\}\\ \geq -I(g(y)) = -I'(y)\,.\end{aligned}$$

Since $I'(y)=\alpha$, it follows by combining this inequality and (4.2.5) that

$$\liminf_{\epsilon\to 0}\ \epsilon\log\nu_\epsilon(G) \geq -I'(y)\,.$$

The proof is complete, since the preceding holds for every $y\in\mathcal{Y}$ and every neighborhood G of y. □

Corollary 4.2.6 *Let $\{\mu_\epsilon\}$ be an exponentially tight family of probability measures on $\mathcal{X}$ equipped with the topology τ_1. If $\{\mu_\epsilon\}$ satisfies an LDP with respect to a Hausdorff topology τ_2 on $\mathcal{X}$ that is coarser than τ_1, then the same LDP holds with respect to the topology τ_1.*

Proof: The proof follows from Theorem 4.2.4 by using as g the natural embedding of $(\mathcal{X},\tau_1)$ onto $(\mathcal{X},\tau_2)$, which is continuous because τ_1 is finer than τ_2. Note that, since g is continuous, the measures μ_ϵ are well-defined as Borel measures on $(\mathcal{X},\tau_2)$. □

Exercise 4.2.7 Suppose that $\mathcal{X}$ is a separable regular space, and that for all $\epsilon>0$, (X_ϵ,Y_ϵ) is distributed according to the product measure $\mu_\epsilon\times\nu_\epsilon$ on $\mathcal{B}_\mathcal{X}\times\mathcal{B}_\mathcal{X}$ (namely, X_ϵ is independent of Y_ϵ). Assume that $\{\mu_\epsilon\}$ satisfies the LDP with the good rate function $I_X(\cdot)$, while ν_ϵ satisfies the LDP with the good rate function $I_Y(\cdot)$, and both $\{\mu_\epsilon\}$ and $\{\nu_\epsilon\}$ are exponentially tight.
Prove that for any continuous $F:\mathcal{X}\times\mathcal{X}\to\mathcal{Y}$, the family of laws induced on $\mathcal{Y}$ by $Z_\epsilon=F(X_\epsilon,Y_\epsilon)$ satisfies the LDP with the good rate function

$$I_Z(z) = \inf_{\{(x,y):z=F(x,y)\}} I_X(x)+I_Y(y)\,. \tag{4.2.8}$$

Hint: Recall that $\mathcal{B}_\mathcal{X}\times\mathcal{B}_\mathcal{X}=\mathcal{B}_{\mathcal{X}\times\mathcal{X}}$ by Theorem D.4. To establish the LDP for $\mu_\epsilon\times\nu_\epsilon$, apply Theorems 4.1.11 and 4.1.18.

Exercise 4.2.9 (a) Prove that Theorem 4.2.4 holds even when the exponentially tight $\{\nu_\epsilon:\epsilon>0\}$ are not Borel measures on $\mathcal{Y}$, provided $\{\nu_\epsilon\circ g^{-1}:\epsilon>0\}$ are Borel probability measures on $\mathcal{X}$.
(b) Show that in particular, Corollary 4.2.6 holds as soon as $\mathcal{B}$ contains the Borel σ-field of $(\mathcal{X},\tau_2)$ and all compact subsets of $(\mathcal{X},\tau_1)$.

4.2.2 Exponential Approximations

In order to extend the contraction principle beyond the continuous case, it is obvious that one should consider approximations by continuous functions. It is beneficial to consider a somewhat wider question, namely, when the LDP for a family of laws $\{\tilde{\mu}_\epsilon\}$ can be deduced from the LDP for a family $\{\mu_\epsilon\}$. The application to approximate contractions follows from these general results.

Definition 4.2.10 *Let $(\mathcal{Y}, d)$ be a metric space. The probability measures $\{\mu_\epsilon\}$ and $\{\tilde{\mu}_\epsilon\}$ on $\mathcal{Y}$ are called exponentially equivalent if there exist probability spaces $\{(\Omega, \mathcal{B}_\epsilon, P_\epsilon)\}$ and two families of $\mathcal{Y}$-valued random variables $\{Z_\epsilon\}$ and $\{\tilde{Z}_\epsilon\}$ with joint laws $\{P_\epsilon\}$ and marginals $\{\mu_\epsilon\}$ and $\{\tilde{\mu}_\epsilon\}$, respectively, such that the following condition is satisfied:*

For each $\delta > 0$, the set $\{\omega : \ (\tilde{Z}_\epsilon, Z_\epsilon) \in \Gamma_\delta\}$ is $\mathcal{B}_\epsilon$ measurable, and

$$\limsup_{\epsilon \to 0} \epsilon \log \, P_\epsilon(\Gamma_\delta) = -\infty \,, \tag{4.2.11}$$

where

$$\Gamma_\delta \stackrel{\triangle}{=} \{(\tilde{y}, y) : d(\tilde{y}, y) > \delta\} \subset \mathcal{Y} \times \mathcal{Y} \,. \tag{4.2.12}$$

Remarks:
(a) The random variables $\{Z_\epsilon\}$ and $\{\tilde{Z}_\epsilon\}$ in Definition 4.2.10 are called *exponentially equivalent.*
(b) It is relatively easy to check that the measurability requirement is satisfied whenever $\mathcal{Y}$ is a separable space, or whenever the laws $\{P_\epsilon\}$ are induced by separable real-valued stochastic processes and d is the supremum norm.

As far as the LDP is concerned, exponentially equivalent measures are indistinguishable, as the following theorem attests.

Theorem 4.2.13 *If an LDP with a good rate function $I(\cdot)$ holds for the probability measures $\{\mu_\epsilon\}$, which are exponentially equivalent to $\{\tilde{\mu}_\epsilon\}$, then the same LDP holds for $\{\tilde{\mu}_\epsilon\}$.*

Proof: This theorem is a consequence of the forthcoming Theorem 4.2.16. To avoid repetitions, a direct proof is omitted. □

As pointed out in the beginning of this section, an important goal in considering exponential equivalence is the treatment of approximations. To this end, the notion of exponential equivalence is replaced by the notion of exponential approximation, as follows.

Definition 4.2.14 *Let $\mathcal{Y}$ and Γ_δ be as in Definition 4.2.10. For each $\epsilon > 0$ and all $m \in \mathbb{Z}_+$, let $(\Omega, \mathcal{B}_\epsilon, P_{\epsilon,m})$ be a probability space, and let the $\mathcal{Y}$-valued random variables $\tilde{Z}_\epsilon$ and $Z_{\epsilon,m}$ be distributed according to the joint law $P_{\epsilon,m}$, with marginals $\tilde{\mu}_\epsilon$ and $\mu_{\epsilon,m}$, respectively. $\{Z_{\epsilon,m}\}$ are called exponentially good approximations of $\{\tilde{Z}_\epsilon\}$ if, for every $\delta > 0$, the set $\{\omega : (\tilde{Z}_\epsilon, Z_{\epsilon,m}) \in \Gamma_\delta\}$ is $\mathcal{B}_\epsilon$ measurable and*

$$\lim_{m\to\infty} \limsup_{\epsilon\to 0} \epsilon \log P_{\epsilon,m}(\Gamma_\delta) = -\infty . \tag{4.2.15}$$

Similarly, the measures $\{\mu_{\epsilon,m}\}$ are exponentially good approximations of $\{\tilde{\mu}_\epsilon\}$ if one can construct probability spaces $\{(\Omega, \mathcal{B}_\epsilon, P_{\epsilon,m})\}$ as above.

It should be obvious that Definition 4.2.14 reduces to Definition 4.2.10 if the laws $P_{\epsilon,m}$ do not depend on m.

The main (highly technical) result of this section is the following relation between the LDPs of exponentially good approximations.

Theorem 4.2.16 *Suppose that for every m, the family of measures $\{\mu_{\epsilon,m}\}$ satisfies the LDP with rate function $I_m(\cdot)$ and that $\{\mu_{\epsilon,m}\}$ are exponentially good approximations of $\{\tilde{\mu}_\epsilon\}$. Then*
(a) $\{\tilde{\mu}_\epsilon\}$ satisfies a weak LDP with the rate function

$$I(y) \stackrel{\triangle}{=} \sup_{\delta>0} \liminf_{m\to\infty} \inf_{z\in B_{y,\delta}} I_m(z) , \tag{4.2.17}$$

where $B_{y,\delta}$ denotes the ball $\{z : d(y,z) < \delta\}$.
(b) If $I(\cdot)$ is a good rate function and for every closed set F,

$$\inf_{y\in F} I(y) \le \limsup_{m\to\infty} \inf_{y\in F} I_m(y) , \tag{4.2.18}$$

then the full LDP holds for $\{\tilde{\mu}_\epsilon\}$ with rate function I.

Remarks:
(a) The sets Γ_δ may be replaced by sets $\tilde{\Gamma}_{\delta,m}$ such that the sets $\{\omega : (\tilde{Z}_\epsilon, Z_{\epsilon,m}) \in \tilde{\Gamma}_{\delta,m}\}$ differ from $\mathcal{B}_\epsilon$ measurable sets by $P_{\epsilon,m}$ null sets, and $\tilde{\Gamma}_{\delta,m}$ satisfy both (4.2.15) and $\Gamma_\delta \subset \tilde{\Gamma}_{\delta,m}$.
(b) If the rate functions $I_m(\cdot)$ are independent of m, and are good rate functions, then by Theorem 4.2.16, $\{\tilde{\mu}_\epsilon\}$ satisfies the LDP with $I(\cdot) = I_m(\cdot)$. In particular, Theorem 4.2.13 is a direct consequence of Theorem 4.2.16.

Proof: (a) Throughout, let $\{Z_{\epsilon,m}\}$ be the exponentially good approximations of $\{\tilde{Z}_\epsilon\}$, having the joint laws $\{P_{\epsilon,m}\}$ with marginals $\{\mu_{\epsilon,m}\}$ and $\{\tilde{\mu}_\epsilon\}$, respectively, and let Γ_δ be as defined in (4.2.12). The weak LDP is obtained by applying Theorem 4.1.11 for the base $\{B_{y,\delta}\}_{y\in\mathcal{Y},\delta>0}$ of $(\mathcal{Y},d)$. Specifically, it suffices to show that

$$I(y) = -\inf_{\delta>0} \limsup_{\epsilon\to 0} \epsilon \log \tilde{\mu}_\epsilon(B_{y,\delta}) = -\inf_{\delta>0} \liminf_{\epsilon\to 0} \epsilon \log \tilde{\mu}_\epsilon(B_{y,\delta}) \,. \quad (4.2.19)$$

To this end, fix $\delta > 0$, $y \in \mathcal{Y}$. Note that for every $m \in \mathbb{Z}_+$ and every $\epsilon > 0$,

$$\{Z_{\epsilon,m} \in B_{y,\delta}\} \subseteq \{\tilde{Z}_\epsilon \in B_{y,2\delta}\} \cup \{(\tilde{Z}_\epsilon, Z_{\epsilon,m}) \in \Gamma_\delta\} \,.$$

Hence, by the union of events bound,

$$\mu_{\epsilon,m}(B_{y,\delta}) \le \tilde{\mu}_\epsilon(B_{y,2\delta}) + P_{\epsilon,m}(\Gamma_\delta) \,.$$

By the large deviations lower bounds for $\{\mu_{\epsilon,m}\}$,

$$\begin{aligned} -\inf_{z\in B_{y,\delta}} I_m(z) &\le \liminf_{\epsilon\to 0} \epsilon \log \mu_{\epsilon,m}(B_{y,\delta}) \\ &\le \liminf_{\epsilon\to 0} \epsilon \log [\tilde{\mu}_\epsilon(B_{y,2\delta}) + P_{\epsilon,m}(\Gamma_\delta)] \qquad (4.2.20) \\ &\le \liminf_{\epsilon\to 0} \epsilon \log \tilde{\mu}_\epsilon(B_{y,2\delta}) \vee \limsup_{\epsilon\to 0} \epsilon \log P_{\epsilon,m}(\Gamma_\delta) \,. \end{aligned}$$

Because $\{\mu_{\epsilon,m}\}$ are exponentially good approximations of $\{\tilde{\mu}_\epsilon\}$,

$$\liminf_{\epsilon\to 0} \epsilon \log \tilde{\mu}_\epsilon(B_{y,2\delta}) \ge \limsup_{m\to\infty} \Big\{ -\inf_{z\in B_{y,\delta}} I_m(z) \Big\} \,.$$

Repeating the derivation leading to (4.2.20) with the roles of $Z_{\epsilon,m}$ and $\tilde{Z}_\epsilon$ reversed yields

$$\limsup_{\epsilon\to 0} \epsilon \log \tilde{\mu}_\epsilon(B_{y,\delta}) \le \liminf_{m\to\infty} \Big\{ -\inf_{z\in \overline{B}_{y,2\delta}} I_m(z) \Big\} \,.$$

Since $\overline{B}_{y,2\delta} \subset B_{y,3\delta}$, (4.2.19) follows by considering the infimum over $\delta > 0$ in the preceding two inequalities (recall the definition (4.2.17) of $I(\cdot)$). Moreover, this argument also implies that

$$I(y) = \sup_{\delta>0} \limsup_{m\to\infty} \inf_{z\in \overline{B}_{y,\delta}} I_m(z) = \sup_{\delta>0} \limsup_{m\to\infty} \inf_{z\in B_{y,\delta}} I_m(z) \,.$$

(b) Fix $\delta > 0$ and a closed set $F \subseteq \mathcal{Y}$. Observe that for $m = 1, 2, \ldots$, and for all $\epsilon > 0$,

$$\{\tilde{Z}_\epsilon \in F\} \subseteq \{Z_{\epsilon,m} \in F^\delta\} \cup \{(\tilde{Z}_\epsilon, Z_{\epsilon,m}) \in \Gamma_\delta\} \,,$$

where $F^\delta = \{z : d(z,F) \le \delta\}$ is the closed blowup of F. Thus, the large deviations upper bounds for $\{\mu_{\epsilon,m}\}$ imply that for every m,

$$\begin{aligned} \limsup_{\epsilon\to 0} \epsilon \log \tilde{\mu}_\epsilon(F) &\le \limsup_{\epsilon\to 0} \epsilon \log \mu_{\epsilon,m}(F^\delta) \vee \limsup_{\epsilon\to 0} \epsilon \log P_{\epsilon,m}(\Gamma_\delta) \\ &\le [-\inf_{y\in F^\delta} I_m(y)] \vee \limsup_{\epsilon\to 0} \epsilon \log P_{\epsilon,m}(\Gamma_\delta) \,. \end{aligned}$$

Hence, as $\{Z_{\epsilon,m}\}$ are exponentially good approximations of $\{\tilde{Z}_\epsilon\}$, considering $m \to \infty$, it follows that

$$\limsup_{\epsilon\to 0} \epsilon \log \tilde{\mu}_\epsilon(F) \le -\limsup_{m\to\infty} \inf_{y\in F^\delta} I_m(y) \le -\inf_{y\in F^\delta} I(y) ,$$

where the second inequality is just our condition (4.2.18) for the closed set F^δ. Taking $\delta \to 0$, Lemma 4.1.6 yields the large deviations upper bound and completes the proof of the full LDP . □

It should be obvious that the results on exponential approximations imply results on approximate contractions. We now present two such results. The first is related to Theorem 4.2.13 and considers approximations that are ϵ dependent. The second allows one to consider approximations that depend on an auxiliary parameter.

Corollary 4.2.21 *Suppose $f : \mathcal{X} \to \mathcal{Y}$ is a continuous map from a Hausdorff topological space $\mathcal{X}$ to the metric space $(\mathcal{Y}, d)$ and that $\{\mu_\epsilon\}$ satisfy the LDP with the good rate function $I : \mathcal{X} \to [0,\infty]$. Suppose further that for all $\epsilon > 0$, $f_\epsilon : \mathcal{X} \to \mathcal{Y}$ are measurable maps such that for all $\delta > 0$, the set $\Gamma_{\epsilon,\delta} \triangleq \{x \in \mathcal{X} : d(f(x), f_\epsilon(x)) > \delta\}$ is measurable, and*

$$\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(\Gamma_{\epsilon,\delta}) = -\infty . \tag{4.2.22}$$

Then the LDP with the good rate function $I'(\cdot)$ of (4.2.2) holds for the measures $\mu_\epsilon \circ f_\epsilon^{-1}$ on $\mathcal{Y}$.

Proof: The contraction principle (Theorem 4.2.1) yields the desired LDP for $\{\mu_\epsilon \circ f^{-1}\}$. By (4.2.22), these measures are exponentially equivalent to $\{\mu_\epsilon \circ f_\epsilon^{-1}\}$, and the corollary follows from Theorem 4.2.13. □

A special case of Theorem 4.2.16 is the following extension of the contraction principle to maps that are not continuous, but that can be approximated well by continuous maps.

Theorem 4.2.23 *Let $\{\mu_\epsilon\}$ be a family of probability measures that satisfies the LDP with a good rate function I on a Hausdorff topological space $\mathcal{X}$, and for $m = 1, 2, \ldots$, let $f_m : \mathcal{X} \to \mathcal{Y}$ be continuous functions, with $(\mathcal{Y}, d)$ a metric space. Assume there exists a measurable map $f : \mathcal{X} \to \mathcal{Y}$ such that for every $\alpha < \infty$,*

$$\limsup_{m\to\infty} \sup_{\{x: I(x)\le\alpha\}} d(f_m(x), f(x)) = 0 . \tag{4.2.24}$$

Then any family of probability measures $\{\tilde{\mu}_\epsilon\}$ for which $\{\mu_\epsilon \circ f_m^{-1}\}$ are exponentially good approximations satisfies the LDP in $\mathcal{Y}$ with the good rate function $I'(y) = \inf\{I(x) : y = f(x)\}$.

Remarks:
(a) The condition (4.2.24) implies that for every $\alpha < \infty$, the function f is continuous on the level set $\Psi_I(\alpha) = \{x : I(x) \leq \alpha\}$. Suppose that in addition,

$$\lim_{m\to\infty} \underline{\inf_{x\in\Psi_I(m)^c}} I(x) = \infty . \tag{4.2.25}$$

Then the LDP for $\mu_\epsilon \circ f^{-1}$ follows as a direct consequence of Theorem 4.2.23 by considering a sequence f_m of continuous extensions of f from $\Psi_I(m)$ to $\mathcal{X}$. (Such a sequence exists whenever $\mathcal{X}$ is a completely regular space.) That (4.2.25) need not hold true, even when $\mathcal{X} = \mathbb{R}$, may be seen by considering the following example. It is easy to check that $\mu_\epsilon = (\delta_{\{0\}} + \delta_{\{\epsilon\}})/2$ satisfies the LDP on $\mathbb{R}$ with the good rate function $I(0) = 0$ and $I(x) = \infty, x \neq 0$. On the other hand, the closure of the complement of any level set is the whole real line. If one now considers the function $f : \mathbb{R} \to \mathbb{R}$ such that $f(0) = 0$ and $f(x) = 1, x \neq 0$, then $\mu_\epsilon \circ f^{-1}$ does not satisfy the LDP with the rate function $I'(y) = \inf\{I(x) : x \in \mathbb{R},\ y = f(x)\}$, i.e., $I'(0) = 0$ and $I'(y) = \infty, y \neq 0$.
(b) Suppose for each $m \in \mathbb{Z}_+$, the family of measures $\{\mu_{\epsilon,m}\}$ satisfies the LDP on $\mathcal{Y}$ with the good rate function $I_m(\cdot)$ of (4.2.26), where the continuous functions $f_m : \mathcal{D}_I \to \mathcal{Y}$ and the measurable function $f : \mathcal{D}_I \to \mathcal{Y}$ satisfy condition (4.2.24). Then any $\{\tilde{\mu}_\epsilon\}$ for which $\{\mu_{\epsilon,m}\}$ are exponentially good approximations satisfies the LDP in $\mathcal{Y}$ with good rate function $I'(\cdot)$. This easy adaptation of the proof of Theorem 4.2.23 is left for the reader.

Proof: By assumption, the functions $f_m : \mathcal{X} \to \mathcal{Y}$ are continuous. Hence, by the contraction principle (Theorem 4.2.1), for each $m \in \mathbb{Z}_+$, the family of measures $\{\mu_\epsilon \circ f_m^{-1}\}$ satisfies the LDP on $\mathcal{Y}$ with the good rate function

$$I_m(y) = \inf\{I(x) : \ x \in \mathcal{X}, \quad y = f_m(x)\} . \tag{4.2.26}$$

Recall that the condition (4.2.24) implies that f is continuous on each level set $\Psi_I(\alpha)$. Therefore, I' is a good rate function on $\mathcal{Y}$ with level sets $f(\Psi_I(\alpha))$ (while the corresponding level set of I_m is $f_m(\Psi_I(\alpha))$).

Fix a closed set F and for any $m \in \mathbb{Z}_+$, let

$$\gamma_m \triangleq \inf_{y\in F} I_m(y) = \inf_{x\in f_m^{-1}(F)} I(x) .$$

Assume first that $\gamma \triangleq \liminf_{m\to\infty} \gamma_m < \infty$, and pass to a subsequence of m's such that $\gamma_m \to \gamma$ and $\sup_m \gamma_m = \alpha < \infty$. Since $I(\cdot)$ is a good rate function and $f_m^{-1}(F)$ are closed sets, there exist $x_m \in \mathcal{X}$ such that $f_m(x_m) \in F$ and $I(x_m) = \gamma_m \leq \alpha$. Now, the uniform convergence assumption of (4.2.24) implies that $f(x_m) \in F^\delta$ for every $\delta > 0$ and all m large enough. Therefore, $\inf_{y\in F^\delta} I'(y) \leq I'(f(x_m)) \leq I(x_m) = \gamma_m$ for all m large enough. Hence, for all $\delta > 0$,

$$\inf_{y \in F^\delta} I'(y) \le \liminf_{m \to \infty} \inf_{y \in F} I_m(y)\,.$$

(Note that this inequality trivially holds when $\gamma = \infty$.) Taking $\delta \to 0$, it follows from Lemma 4.1.6 that for every closed set F,

$$\inf_{y \in F} I'(y) \le \liminf_{m \to \infty} \inf_{y \in F} I_m(y)\,. \tag{4.2.27}$$

In particular, this inequality implies that (4.2.18) holds for the good rate function $I'(\cdot)$. Moreover, considering $F = \overline{B}_{y,\delta}$, and taking $\delta \to 0$, it follows from Lemma 4.1.6 that

$$I'(y) = \sup_{\delta > 0} \inf_{z \in \overline{B}_{y,\delta}} I'(z) \le \sup_{\delta > 0} \liminf_{m \to \infty} \inf_{z \in B_{y,\delta}} I_m(z) \stackrel{\triangle}{=} \bar{I}(y)\,.$$

Note that $\bar{I}(\cdot)$ is the rate function defined in Theorem 4.2.16, and consequently the proof is complete as soon as we show that $\bar{I}(y) \le I'(y)$ for all $y \in \mathcal{Y}$. To this end, assume with no loss of generality that $I'(y) = \alpha < \infty$. Then, $y \in f(\Psi_I(\alpha))$, i.e., there exists $x \in \Psi_I(\alpha)$ such that $f(x) = y$. Note that $y_m = f_m(x) \in f_m(\Psi_I(\alpha))$, and consequently $I_m(y_m) \le \alpha$ for all $m \in \mathbb{Z}_+$. The condition (4.2.24) then implies that $d(y, y_m) \to 0$, and hence $\bar{I}(y) \le \liminf_{m \to \infty} I_m(y_m) \le \alpha$, as required. $\square$

Exercise 4.2.28 [Based on [DV75a]] Let $\Sigma = \{1, \cdots, r\}$, and let Y_t be a Σ-valued *continuous time* Markov process with irreducible generator $A = \{a(i,j)\}$. In this exercise, you derive the LDP for the empirical measures

$$L^{\mathbf{y}}_\epsilon(i) = \epsilon \int_0^{1/\epsilon} 1_i(Y_t)dt, \quad i = 1, \ldots, r\,.$$

(a) Define

$$L^{\mathbf{y}}_{\epsilon,m}(i) = \frac{\epsilon}{m} \sum_{j=1}^{\lfloor \frac{m}{\epsilon} \rfloor} 1_i(Y_{\frac{j}{m}}), \quad i = 1, \ldots, r\,.$$

Show that $\{L^{\mathbf{y}}_{\epsilon,m}\}$ are exponentially good approximations of $\{L^{\mathbf{y}}_\epsilon\}$.
Hint: Note that

$$|L^{\mathbf{y}}_\epsilon(i) - L^{\mathbf{y}}_{\epsilon,m}(i)| \le \frac{\epsilon}{m} \left\{ \begin{array}{l} \text{total number of jumps in} \\ \text{the path } Y_t,\ t \in [0, 1/\epsilon] \end{array} \right\} \stackrel{\triangle}{=} \frac{\epsilon}{m} N_\epsilon\,,$$

and N_ϵ is stochastically dominated by a Poisson(c/ϵ) random variable for some constant $c < \infty$.
(b) Note that $L^{\mathbf{y}}_{\epsilon,m}$ is the empirical measure of a Σ-valued, discrete time Markov process with irreducible transition probability matrix $e^{A/m}$. Using Theorem

3.1.6 and Exercise 3.1.11, show that for every m, $L^{\mathbf{y}}_{\epsilon,m}$ satisfies the LDP with the good rate function

$$I_m(q) = m \sup_{u \gg 0} \sum_{j=1}^r q_j \log \left[\frac{u_j}{(e^{A/m} u)_j} \right] ,$$

where $q \in M_1(\Sigma)$.

(c) Applying Theorem 4.2.16, prove that $\{L^{\mathbf{y}}_\epsilon\}$ satisfies the LDP with the good rate function

$$I(q) = \sup_{u \gg 0} \Big\{ - \sum_{j=1}^r q_j \frac{(Au)_j}{u_j} \Big\} .$$

Hint: Check that for all $q \in M_1(\Sigma), I(q) \geq I_m(q)$, and that for each fixed $u \gg 0$, $I_m(q) \geq -\sum_j q_j \frac{(Au)_j}{u_j} - \frac{c(u)}{m}$ for some $c(u) < \infty$.

(d) Assume that A is symmetric and check that then

$$I(q) = -\sum_{i,j=1}^r \sqrt{q_i}\, a(i,j) \sqrt{q_j} .$$

Exercise 4.2.29 Suppose that for every m, the family of measures $\{\mu_{\epsilon,m}\}$ satisfies the LDP with good rate function $I_m(\cdot)$ and that $\{\mu_{\epsilon,m}\}$ are exponentially good approximations of $\{\tilde{\mu}_\epsilon\}$.

(a) Show that if $(\mathcal{Y}, d)$ is a Polish space, then $\{\tilde{\mu}_{\epsilon_n}\}$ is exponentially tight for any $\epsilon_n \to 0$. Hence, by part (a) of Theorem 4.2.16, $\{\tilde{\mu}_\epsilon\}$ satisfies the LDP with the good rate function $I(\cdot)$ of (4.2.17).

Hint: See Exercise 4.1.10.

(b) Let $\mathcal{Y} = \{1/m, m \in \mathbb{Z}_+\}$ with the metric $d(\cdot,\cdot)$ induced on $\mathcal{Y}$ by $\mathbb{R}$ and $\mathcal{Y}$-valued random variables Y_m such that $P(Y_m = 1 \text{ for every } m) = 1/2$, and $P(Y_m = 1/m \text{ for every } m) = 1/2$. Check that $Z_{\epsilon,m} \triangleq Y_m$ are exponentially good approximations of $\tilde{Z}_\epsilon \triangleq Y_{[1/\epsilon]}$ ($\epsilon \leq 1$), which for any fixed $m \in \mathbb{Z}_+$ satisfy the LDP in $\mathcal{Y}$ with the good rate function $I_m(y) = 0$ for $y = 1$, $y = 1/m$, and $I_m(y) = \infty$ otherwise. Check that in this case, the good rate function $I(\cdot)$ of (4.2.17) is such that $I(y) = \infty$ for every $y \neq 1$ and in particular, the large deviations upper bound fails for $\{\tilde{Z}_\epsilon \neq 1\}$ and this rate function.

Remark: This example shows that when $(\mathcal{Y}, d)$ is not a Polish space one can not dispense of condition (4.2.18) in Theorem 4.2.16.

Exercise 4.2.30 For any $\delta > 0$ and probability measures ν, μ on the metric space $(\mathcal{Y}, d)$ let

$$\rho_\delta(\nu,\mu) \triangleq \sup\{\nu(A) - \mu(A^\delta) : A \in \mathcal{B}_{\mathcal{Y}}\} .$$

(a) Show that if $\{\mu_{\epsilon,m}\}$ are exponentially good approximations of $\{\tilde{\mu}_\epsilon\}$ then

$$\lim_{m\to\infty} \limsup_{\epsilon\to 0} \epsilon \log \rho_\delta(\mu_{\epsilon,m}, \tilde{\mu}_\epsilon) = -\infty . \tag{4.2.31}$$

(b) Show that if $(\mathcal{Y}, d)$ is a Polish space and (4.2.31) holds for any $\delta > 0$, then $\{\mu_{\epsilon,m}\}$ are exponentially good approximations of $\{\tilde{\mu}_\epsilon\}$.
Hint: Recall the following consequence of [Str65, Theorem 11]. For any open set $\Gamma \subset \mathcal{Y}^2$ and any Borel probability measures ν, μ on the Polish space $(\mathcal{Y}, d)$ there exists a Borel probability measure P on $\mathcal{Y}^2$ with marginals μ, ν such that

$$P(\Gamma) = \sup\{\nu(G) - \mu(\{\tilde{y} : \exists y \in G, \text{ such that } (\tilde{y}, y) \in \Gamma^c\}) \ : \ G \subset \mathcal{Y} \text{ open}\} .$$

Conclude that $P_{\epsilon,m}(\Gamma_{\delta'}) \le \rho_\delta(\mu_{\epsilon,m}, \tilde{\mu}_\epsilon)$ for any m, $\epsilon > 0$, and $\delta' > \delta > 0$.

Exercise 4.2.32 Prove Theorem 4.2.13, assuming that $\{\mu_\epsilon\}$ are Borel probability measures, but $\{\tilde{\mu}_\epsilon\}$ are not necessarily such.

4.3 Varadhan's Integral Lemma

Throughout this section, $\{Z_\epsilon\}$ is a family of random variables taking values in the regular topological space $\mathcal{X}$, and $\{\mu_\epsilon\}$ denotes the probability measures associated with $\{Z_\epsilon\}$. The next theorem could actually be used as a starting point for developing the large deviations paradigm. It is a very useful tool in many applications of large deviations. For example, the asymptotics of the partition function in statistical mechanics can be derived using this theorem.

Theorem 4.3.1 (Varadhan) *Suppose that $\{\mu_\epsilon\}$ satisfies the LDP with a good rate function $I : \mathcal{X} \to [0, \infty]$, and let $\phi : \mathcal{X} \to \mathbb{R}$ be any continuous function. Assume further either the tail condition*

$$\lim_{M\to\infty} \limsup_{\epsilon\to 0} \ \epsilon \log E\left[e^{\phi(Z_\epsilon)/\epsilon} \ 1_{\{\phi(Z_\epsilon)\ge M\}}\right] = -\infty \, , \tag{4.3.2}$$

or the following moment condition for some $\gamma > 1$,

$$\limsup_{\epsilon\to 0} \epsilon \log E\left[e^{\gamma\phi(Z_\epsilon)/\epsilon}\right] < \infty \, . \tag{4.3.3}$$

Then

$$\lim_{\epsilon\to 0} \ \epsilon \log E\left[e^{\phi(Z_\epsilon)/\epsilon}\right] = \sup_{x\in\mathcal{X}} \{\phi(x) - I(x)\} \, .$$

Remark: This theorem is the natural extension of Laplace's method to infinite dimensional spaces. Indeed, let $\mathcal{X} = \mathbb{R}$ and assume for the moment that the density of μ_ϵ with respect to Lebesgue's measure is such that $d\mu_\epsilon/dx \approx e^{-I(x)/\epsilon}$. Then

$$\int_{\mathbb{R}} e^{\phi(x)/\epsilon} \mu_\epsilon(dx) \approx \int_{\mathbb{R}} e^{(\phi(x)-I(x))/\epsilon} dx \, .$$

Assume that $I(\cdot)$ and $\phi(\cdot)$ are twice differentiable, with $(\phi(x)-I(x))$ concave and possessing a unique global maximum at some $\overline{x}$. Then

$$\phi(x)-I(x)=\phi(\overline{x})-I(\overline{x})+\frac{(x-\overline{x})^2}{2}(\phi(x)-I(x))''|_{x=\xi} \ ,$$

where $\xi \in [\overline{x}, x]$. Therefore,

$$\int_{\mathbb{R}} e^{\phi(x)/\epsilon}\mu_\epsilon(dx) \approx e^{(\phi(\overline{x})-I(\overline{x}))/\epsilon}\int_{\mathbb{R}} e^{-B(x)(x-\overline{x})^2/2\epsilon}dx \ ,$$

where $B(\cdot) \geq 0$. The content of Laplace's method (and of Theorem 4.3.1) is that on a logarithmic scale the rightmost integral may be ignored.

Theorem 4.3.1 is a direct consequence of the following three lemmas.

Lemma 4.3.4 *If $\phi : \mathcal{X} \to \mathbb{R}$ is lower semicontinuous and the large deviations lower bound holds with $I : \mathcal{X} \to [0, \infty]$, then*

$$\liminf_{\epsilon\to 0} \ \epsilon \log E\left[e^{\phi(Z_\epsilon)/\epsilon}\right] \geq \sup_{x\in\mathcal{X}} \{\phi(x)-I(x)\} \ . \qquad (4.3.5)$$

Lemma 4.3.6 *If $\phi : \mathcal{X} \to \mathbb{R}$ is an upper semicontinuous function for which the tail condition (4.3.2) holds, and the large deviations upper bound holds with the good rate function $I : \mathcal{X} \to [0, \infty]$, then*

$$\limsup_{\epsilon\to 0} \ \epsilon \log E\left[e^{\phi(Z_\epsilon)/\epsilon}\right] \leq \sup_{x\in\mathcal{X}} \{\phi(x)-I(x)\} . \qquad (4.3.7)$$

Lemma 4.3.8 *Condition (4.3.3) implies the tail condition (4.3.2).*

Proof of Lemma 4.3.4: Fix $x \in \mathcal{X}$ and $\delta > 0$. Since $\phi(\cdot)$ is lower semicontinuous, it follows that there exists a neighborhood G of x such that $\inf_{y\in G} \phi(y) \geq \phi(x)-\delta$. Hence,

$$\begin{aligned}\liminf_{\epsilon\to 0} \ \epsilon \log E\left[e^{\phi(Z_\epsilon)/\epsilon}\right] &\geq \liminf_{\epsilon\to 0} \ \epsilon \log E\left[e^{\phi(Z_\epsilon)/\epsilon}1_{\{Z_\epsilon\in G\}}\right] \\ &\geq \inf_{y\in G}\phi(y)+\liminf_{\epsilon\to 0} \ \epsilon\log\mu_\epsilon(G) \ .\end{aligned}$$

By the large deviations lower bound and the choice of G,

$$\inf_{y\in G}\phi(y)+\liminf_{\epsilon\to 0} \ \epsilon\log\mu_\epsilon(G) \geq \inf_{y\in G}\phi(y)-\inf_{y\in G}I(y) \ \geq \phi(x)-I(x)-\delta \, .$$

The inequality (4.3.5) now follows, since $\delta > 0$ and $x \in \mathcal{X}$ are arbitrary. $\square$

Proof of Lemma 4.3.6: Consider first a function ϕ bounded above, i.e., $\sup_{x\in\mathcal{X}} \ \phi(x) \leq M < \infty$. For such functions, the tail condition (4.3.2) holds trivially. Fix $\alpha < \infty$ and $\delta > 0$, and let $\Psi_I(\alpha) = \{x : \ I(x) \le \alpha\}$ denote the compact level set of the good rate function I. Since $I(\cdot)$ is lower semicontinuous, $\phi(\cdot)$ is upper semicontinuous, and $\mathcal{X}$ is a regular topological space, for every $x \in \Psi_I(\alpha)$, there exists a neighborhood A_x of x such that

$$\inf_{y\in\overline{A_x}} I(y) \ge I(x) - \delta \quad , \quad \sup_{y\in\overline{A_x}} \phi(y) \le \phi(x) + \delta \, . \tag{4.3.9}$$

From the open cover $\cup_{x\in\Psi_I(\alpha)} A_x$ of the compact set $\Psi_I(\alpha)$, one can extract a finite cover of $\Psi_I(\alpha)$, e.g., $\cup_{i=1}^N A_{x_i}$. Therefore,

$$\begin{aligned} E\left[e^{\phi(Z_\epsilon)/\epsilon}\right] &\le \sum_{i=1}^N E\left[e^{\phi(Z_\epsilon)/\epsilon} 1_{\{Z_\epsilon\in A_{x_i}\}}\right] + \ e^{M/\epsilon}\mu_\epsilon\Big(\Big(\bigcup_{i=1}^N A_{x_i}\Big)^c\Big) \\ &\le \sum_{i=1}^N e^{(\phi(x_i)+\delta)/\epsilon}\mu_\epsilon(\,\overline{A_{x_i}}\,) + \ e^{M/\epsilon}\mu_\epsilon\Big(\Big(\bigcup_{i=1}^N A_{x_i}\Big)^c\Big) \end{aligned}$$

where the last inequality follows by (4.3.9). Applying the large deviations upper bound to the sets $\overline{A_{x_i}}$, $i = 1, \ldots, N$ and $(\cup_{i=1}^N A_{x_i})^c \subseteq \Psi_I(\alpha)^c$, one obtains (again, in view of (4.3.9)),

$$\begin{aligned} &\limsup_{\epsilon\to 0} \epsilon \log E\left[e^{\phi(Z_\epsilon)/\epsilon}\right] \\ &\le \ \max\Big\{\max_{i=1}^N\{\phi(x_i) + \delta - \inf_{y\in\overline{A_{x_i}}} I(y)\ \}, M - \inf_{y\in(\cup_{i=1}^N A_{x_i})^c} I(y)\Big\} \\ &\le \ \max\Big\{\max_{i=1}^N\{\phi(x_i) - I(x_i) + 2\delta\}, M - \alpha\Big\} \\ &\le \ \max\Big\{\sup_{x\in\mathcal{X}}\{\phi(x) - I(x)\}, M - \alpha\Big\} + 2\delta \, . \end{aligned}$$

Thus, for any $\phi(\cdot)$ bounded above, the lemma follows by taking the limits $\delta \to 0$ and $\alpha \to \infty$.

To treat the general case, set $\phi_M(x) = \phi(x) \wedge M \le \phi(x)$, and use the preceding to show that for every $M < \infty$,

$$\begin{aligned} &\limsup_{\epsilon\to 0}\ \epsilon \log E\left[e^{\phi(Z_\epsilon)/\epsilon}\right] \\ &\qquad \le \sup_{x\in\mathcal{X}}\ \{\phi(x) - I(x)\} \vee \limsup_{\epsilon\to 0}\ \epsilon \log E\left[e^{\phi(Z_\epsilon)/\epsilon}\ 1_{\{\phi(Z_\epsilon)\ge M\}}\right] . \end{aligned}$$

The tail condition (4.3.2) completes the proof of the lemma by taking the limit $M \to \infty$. □

Proof of Lemma 4.3.8: For $\epsilon > 0$, define $X_\epsilon \triangleq \exp((\phi(Z_\epsilon) - M)/\epsilon)$, and let $\gamma > 1$ be the constant given in the moment condition (4.3.3). Then

$$\begin{aligned} e^{-M/\epsilon} E\left[e^{\phi(Z_\epsilon)/\epsilon}\, 1_{\{\phi(Z_\epsilon)\geq M\}}\right] &= E\left[X_\epsilon\, 1_{\{X_\epsilon \geq 1\}}\right] \\ &\leq E\left[(X_\epsilon)^\gamma\right] = e^{-\gamma M/\epsilon} E\left[e^{\gamma\phi(Z_\epsilon)/\epsilon}\right] . \end{aligned}$$

Therefore,

$$\begin{aligned} &\limsup_{\epsilon\to 0} \epsilon \log E\left[e^{\phi(Z_\epsilon)/\epsilon}\, 1_{\{\phi(Z_\epsilon)\geq M\}}\right] \\ &\qquad \leq -(\gamma-1)M + \limsup_{\epsilon\to 0} \epsilon \log E\left[e^{\gamma\phi(Z_\epsilon)/\epsilon}\right] . \end{aligned}$$

The right side of this inequality is finite by the moment condition (4.3.3). In the limit $M \to \infty$, it yields the tail condition (4.3.2). $\square$

Exercise 4.3.10 Let $\phi : \mathcal{X} \to [-\infty, \infty]$ be an upper semicontinuous function, and let $I(\cdot)$ be a good rate function. Prove that in any closed set $F \subset \mathcal{X}$ on which ϕ is bounded above, there exists a point x_0 such that

$$\phi(x_0) - I(x_0) = \sup_{x\in F} \{\phi(x) - I(x)\} .$$

Exercise 4.3.11 [From [DeuS89b], Exercise 2.1.24]. Assume that $\{\mu_\epsilon\}$ satisfies the LDP with good rate function $I(\cdot)$ and that the tail condition (4.3.2) holds for the continuous function $\phi : \mathcal{X} \to \mathbb{R}$. Show that

$$\liminf_{\epsilon\to 0} \epsilon \log \left(\int_G e^{\phi(x)/\epsilon} d\mu_\epsilon\right) \geq \sup_{x\in G} \{\phi(x) - I(x)\}, \quad \forall G \quad \text{open}\,,$$

$$\limsup_{\epsilon\to 0} \epsilon \log \left(\int_F e^{\phi(x)/\epsilon} d\mu_\epsilon\right) \leq \sup_{x\in F} \{\phi(x) - I(x)\}, \quad \forall F \quad \text{closed}\,.$$

Exercise 4.3.12 The purpose of this exercise is to demonstrate that some tail condition like (4.3.2) is necessary for Lemma 4.3.6 to hold. In particular, this lemma may not hold for linear functions.

Consider a family of real valued random variables $\{Z_\epsilon\}$, where $\mathrm{P}(Z_\epsilon = 0) = 1 - 2p_\epsilon$, $\mathrm{P}(Z_\epsilon = -m_\epsilon) = p_\epsilon$, and $\mathrm{P}(Z_\epsilon = m_\epsilon) = p_\epsilon$.
(a) Prove that if

$$\lim_{\epsilon\to 0} \epsilon \log p_\epsilon = -\infty\,,$$

then the laws of $\{Z_\epsilon\}$ are exponentially tight, and moreover they satisfy the LDP with the convex, good rate function

$$I(x) = \begin{cases} 0 & x = 0 \\ \infty & \text{otherwise}\,. \end{cases}$$

(b) Let $m_\epsilon = -\epsilon \log p_\epsilon$ and define

$$\Lambda(\lambda) = \lim_{\epsilon \to 0} \epsilon \log E\left(e^{\lambda Z_\epsilon/\epsilon}\right) .$$

Prove that

$$\Lambda(\lambda) = \begin{cases} 0 & |\lambda| \leq 1 \\ \infty & \text{otherwise}, \end{cases}$$

and its Fenchel–Legendre transform is $\Lambda^*(x) = |x|$.
(c) Observe that $\Lambda(\lambda) \neq \sup_{x \in \mathbb{R}} \{\lambda x - I(x)\}$, and $\Lambda^*(\cdot) \neq I(\cdot)$.

4.4 Bryc's Inverse Varadhan Lemma

As will be seen in Section 4.5, in the setting of topological vector spaces, linear functionals play an important role in establishing the LDP, particularly when convexity is involved. Note, however, that Varadhan's lemma applies to nonlinear functions as well. It is the goal of this section to derive the *inverse* of Varadhan's lemma. Specifically, let $\{\mu_\epsilon\}$ be a family of probability measures on a topological space $\mathcal{X}$. For each Borel measurable function $f : \mathcal{X} \to \mathbb{R}$, define

$$\Lambda_f \stackrel{\triangle}{=} \lim_{\epsilon \to 0} \epsilon \log \int_{\mathcal{X}} e^{f(x)/\epsilon} \mu_\epsilon(dx) , \tag{4.4.1}$$

provided the limit exists. For example, when $\mathcal{X}$ is a vector space, then the $\{\Lambda_f\}$ for continuous linear functionals (i.e., for $f \in \mathcal{X}^*$) are just the values of the logarithmic moment generating function defined in Section 4.5. The main result of this section is that the LDP is a consequence of exponential tightness and the existence of the limits (4.4.1) for every $f \in \mathcal{G}$, for appropriate families of functions $\mathcal{G}$. This result is used in Section 6.4, where the smoothness assumptions of the Gärtner–Ellis theorem (Theorem 2.3.6) are replaced by mixing assumptions en route to the LDP for the empirical measures of Markov chains.

Throughout this section, it is assumed that $\mathcal{X}$ is a completely regular topological space, i.e., $\mathcal{X}$ is Hausdorff, and for any closed set $F \subset \mathcal{X}$ and any point $x \notin F$, there exists a continuous function $f : \mathcal{X} \to [0, 1]$ such that $f(x) = 1$ and $f(y) = 0$ for all $y \in F$. It is also not hard to verify that such a space is regular and that both metric spaces and Hausdorff topological vector spaces are completely regular.

The class of all bounded, real valued continuous functions on $\mathcal{X}$ is denoted throughout by $C_b(\mathcal{X})$.

Theorem 4.4.2 (Bryc) *Suppose that the family $\{\mu_\epsilon\}$ is exponentially tight and that the limit Λ_f in (4.4.1) exists for every $f \in C_b(\mathcal{X})$. Then $\{\mu_\epsilon\}$ satisfies the LDP with the good rate function*

$$I(x) = \sup_{f \in C_b(\mathcal{X})} \{f(x) - \Lambda_f\} . \tag{4.4.3}$$

Furthermore, for every $f \in C_b(\mathcal{X})$,

$$\Lambda_f = \sup_{x \in \mathcal{X}} \{f(x) - I(x)\} . \tag{4.4.4}$$

Remark: In the case where $\mathcal{X}$ is a topological vector space, it is tempting to compare (4.4.3) and (4.4.4) with the Fenchel–Legendre transform pair $\Lambda(\cdot)$ and $\Lambda^*(\cdot)$ of Section 4.5. Note, however, that here the rate function $I(x)$ need not be convex.

Proof: Since $\Lambda_0 = 0$, it follows that $I(\cdot) \geq 0$. Moreover, $I(x)$ is lower semicontinuous, since it is the supremum of continuous functions. Due to the exponential tightness of $\{\mu_\epsilon\}$, the LDP asserted follows once the weak LDP (with rate function $I(\cdot)$) is proved. Moreover, by an application of Varadhan's lemma (Theorem 4.3.1), the identity (4.4.4) then holds. It remains, therefore, only to prove the weak LDP, which is a consequence of the following two lemmas.

Lemma 4.4.5 (Upper bound) *If Λ_f exists for each $f \in C_b(\mathcal{X})$, then, for every compact $\Gamma \subset \mathcal{X}$,*

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(\Gamma) \leq - \inf_{x \in \Gamma} I(x) .$$

Lemma 4.4.6 (Lower bound) *If Λ_f exists for each $f \in C_b(\mathcal{X})$, then, for every open $G \subset \mathcal{X}$ and each $x \in G$,*

$$\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(G) \geq -I(x) .$$

Proof of Lemma 4.4.5: The proof is almost identical to the proof of part (b) of Theorem 4.5.3, substituting $f(x)$ for $\langle \lambda, x \rangle$. To avoid repetition, the details are omitted. □

Proof of Lemma 4.4.6: Fix $x \in \mathcal{X}$ and a neighborhood G of x. Since $\mathcal{X}$ is a completely regular topological space, there exists a continuous function $f : \mathcal{X} \to [0, 1]$, such that $f(x) = 1$ and $f(y) = 0$ for all $y \in G^c$. For $m > 0$, define $f_m(\cdot) \triangleq m(f(\cdot) - 1)$. Then

$$\int_{\mathcal{X}} e^{f_m(x)/\epsilon} \mu_\epsilon(dx) \leq e^{-m/\epsilon} \mu_\epsilon(G^c) + \mu_\epsilon(G) \leq e^{-m/\epsilon} + \mu_\epsilon(G) .$$

Since $f_m \in C_b(\mathcal{X})$ and $f_m(x) = 0$, it now follows that

$$\begin{aligned} & \max\{ \liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(G),\ -m \} \\ & \quad \geq \liminf_{\epsilon \to 0} \epsilon \log \int_{\mathcal{X}} e^{f_m(x)/\epsilon} \mu_\epsilon(dx) = \Lambda_{f_m} \\ & \quad = -[f_m(x) - \Lambda_{f_m}] \geq - \sup_{f \in C_b(\mathcal{X})} \{ f(x) - \Lambda_f \} = -I(x) \ , \end{aligned}$$

and the lower bound follows by letting $m \to \infty$. □

This proof works because indicators on open sets are approximated well enough by bounded continuous functions. It is clear, however, that not all of $C_b(\mathcal{X})$ is needed for that purpose. The following definition is the tool for relaxing the assumptions of Theorem 4.4.2.

Definition 4.4.7 *A class $\mathcal{G}$ of continuous, real valued functions on a topological space $\mathcal{X}$ is said to be well-separating if:*
(1) $\mathcal{G}$ contains the constant functions.
(2) $\mathcal{G}$ is closed under finite pointwise minima, i.e., $g_1, g_2 \in \mathcal{G} \Rightarrow g_1 \wedge g_2 \in \mathcal{G}$.
(3) $\mathcal{G}$ separates points of $\mathcal{X}$, i.e., given two points $x, y \in \mathcal{X}$ with $x \neq y$, and $a, b \in \mathbb{R}$, there exists a function $g \in \mathcal{G}$ such that $g(x) = a$ and $g(y) = b$.

Remark: It is easy to check that if $\mathcal{G}$ is well-separating, so is $\mathcal{G}^+$, the class of all bounded above functions in $\mathcal{G}$.

When $\mathcal{X}$ is a vector space, a particularly useful class of well-separating functions exists.

Lemma 4.4.8 *Let $\mathcal{X}$ be a locally convex, Hausdorff topological vector space. Then the class $\mathcal{G}$ of all continuous, bounded above, concave functions on $\mathcal{X}$ is well-separating.*

Proof: Let $\mathcal{X}^*$ denote the topological dual of $\mathcal{X}$, and let $\mathcal{G}_0 \triangleq \{\lambda(x) + c : \lambda \in \mathcal{X}^*, c \in \mathbb{R}\}$. Note that $\mathcal{G}_0$ contains the constant functions, and by the Hahn–Banach theorem, $\mathcal{G}_0$ separates points of $\mathcal{X}$. Since $\mathcal{G}_0$ consists of continuous, concave functions, it follows that the class of all continuous, concave functions separates points. Moreover, as the pointwise minimum of concave, continuous functions is concave and continuous, this class of functions is well-separating. Finally, by the earlier remark, it suffices to consider only the bounded above, continuous, concave functions. □

The following lemma, whose proof is deferred to the end of the section, states the specific approximation property of well-separating classes of functions that allows their use instead of $C_b(\mathcal{X})$. It will be used in the proof of Theorem 4.4.10.

Lemma 4.4.9 *Let $\mathcal{G}$ be a well-separating class of functions on $\mathcal{X}$. Then for any compact set $\Gamma \subset \mathcal{X}$, any $f \in C_b(\Gamma)$, and any $\delta > 0$, there exists an integer $d < \infty$ and functions $g_1, \ldots, g_d \in \mathcal{G}$ such that*

$$\sup_{x\in\Gamma} |f(x) - \max_{i=1}^{d} g_i(x)| \le \delta$$

and

$$\sup_{x\in\mathcal{X}} g_i(x) \le \sup_{x\in\Gamma} f(x) < \infty \, .$$

Theorem 4.4.10 *Let $\{\mu_\epsilon\}$ be an exponentially tight family of probability measures on a completely regular topological space $\mathcal{X}$, and suppose $\mathcal{G}$ is a well-separating class of functions on $\mathcal{X}$. If Λ_g exists for each $g \in \mathcal{G}$, then Λ_f exists for each $f \in C_b(\mathcal{X})$. Consequently, all the conclusions of Theorem 4.4.2 hold.*

Proof: Fix a bounded continuous function $f(x)$ with $|f(x)| \le M$. Since the family $\{\mu_\epsilon\}$ is exponentially tight, there exists a compact set Γ such that for all ϵ small enough,

$$\mu_\epsilon(\Gamma^c) \le e^{-3M/\epsilon} \, .$$

Fix $\delta > 0$ and let $g_1, \ldots, g_d \in \mathcal{G}$, $d < \infty$ be as in Lemma 4.4.9, with $h(x) \triangleq \max_{i=1}^{d} g_i(x)$. Then, for every $\epsilon > 0$,

$$\max_{i=1}^{d} \left\{ \int_{\mathcal{X}} e^{g_i(x)/\epsilon} \mu_\epsilon(dx) \right\} \le \int_{\mathcal{X}} e^{h(x)/\epsilon} \mu_\epsilon(dx) \le \sum_{i=1}^{d} \int_{\mathcal{X}} e^{g_i(x)/\epsilon} \mu_\epsilon(dx) \, .$$

Hence, by the assumption of the theorem, the limit

$$\Lambda_h = \lim_{\epsilon\to 0} \epsilon \log \int_{\mathcal{X}} e^{h(x)/\epsilon} \mu_\epsilon(dx)$$

exists, and $\Lambda_h = \max_{i=1}^{d} \Lambda_{g_i}$. Moreover, by Lemma 4.4.9, $h(x) \le M$ for all $x \in \mathcal{X}$, and $h(x) \ge (f(x) - \delta) \ge -(M + \delta)$ for all $x \in \Gamma$. Consequently, for all ϵ small enough,

$$\int_{\Gamma^c} e^{h(x)/\epsilon} \mu_\epsilon(dx) \le e^{-2M/\epsilon}$$

and

$$\int_{\Gamma} e^{h(x)/\epsilon} \mu_\epsilon(dx) \ge e^{-(M+\delta)/\epsilon} \mu_\epsilon(\Gamma) \ge \frac{1}{2} e^{-(M+\delta)/\epsilon}.$$

Hence, for any $\delta < M$,

$$\Lambda_h = \lim_{\epsilon\to 0} \epsilon \log \int_{\Gamma} e^{h(x)/\epsilon} \mu_\epsilon(dx) \, .$$

Since $\sup_{x\in\Gamma}|f(x)-h(x)|\le\delta$,

$$\begin{aligned}\limsup_{\epsilon\to 0}\epsilon\log\int_\Gamma e^{f(x)/\epsilon}\mu_\epsilon(dx) &\le \delta+\limsup_{\epsilon\to 0}\epsilon\log\int_\Gamma e^{h(x)/\epsilon}\mu_\epsilon(dx)\\ =\delta+\Lambda_h &= \delta+\liminf_{\epsilon\to 0}\epsilon\log\int_\Gamma e^{h(x)/\epsilon}\mu_\epsilon(dx)\\ &\le 2\delta+\liminf_{\epsilon\to 0}\epsilon\log\int_\Gamma e^{f(x)/\epsilon}\mu_\epsilon(dx)\,.\end{aligned}$$

Thus, taking $\delta\to 0$, it follows that

$$\lim_{\epsilon\to 0}\epsilon\log\int_\Gamma e^{f(x)/\epsilon}\mu_\epsilon(dx)$$

exists. This limit equals Λ_f, since, for all ϵ small enough,

$$\int_{\Gamma^c} e^{f(x)/\epsilon}\mu_\epsilon(dx)\le e^{-2M/\epsilon}\,,\qquad \int_\Gamma e^{f(x)/\epsilon}\mu_\epsilon(dx)\ge\frac{1}{2}\,e^{-M/\epsilon}\,. \qquad\square$$

Proof of Lemma 4.4.9: Fix $\Gamma\subset\mathcal{X}$ compact, $f\in C_b(\Gamma)$ and $\delta>0$. Let $x,y\in\Gamma$ with $x\neq y$. Since $\mathcal{G}$ separates points in Γ, there is a function $g_{x,y}(\cdot)\in\mathcal{G}$ such that $g_{x,y}(x)=f(x)$ and $g_{x,y}(y)=f(y)$. Because each of the functions $f(\cdot)-g_{x,y}(\cdot)$ is continuous, one may find for each $y\in\Gamma$ a neighborhood U_y of y such that

$$\inf_{u\in U_y}\{f(u)-g_{x,y}(u)\}\ \ge -\delta\ .$$

The neighborhoods $\{U_y\}$ form a cover of Γ; hence, Γ may be covered by a finite collection $U_{y_1},\ldots,U_{y_m}$ of such neighborhoods. For every $x\in\Gamma$, define

$$g_x(\cdot)=g_{x,y_1}(\cdot)\wedge g_{x,y_2}(\cdot)\wedge\cdots\wedge g_{x,y_m}(\cdot)\in\mathcal{G}\ .$$

Then

$$\inf_{u\in\Gamma}\{f(u)-g_x(u)\}\ \ge -\delta\ . \tag{4.4.11}$$

Recall now that, for all i, $g_{x,y_i}(x)=f(x)$ and hence $g_x(x)=f(x)$. Since each of the functions $f(\cdot)-g_x(\cdot)$ is continuous, one may find a finite cover $V_1,\ldots,V_d$ of Γ and functions $g_{x_1},\ldots,g_{x_d}\in\mathcal{G}$ such that

$$\sup_{v\in V_i}\{f(v)-g_{x_i}(v)\}\ \le\delta\ . \tag{4.4.12}$$

By the two preceding inequalities,

$$\sup_{v\in\Gamma}|f(v)-\max_{i=1}^{d} g_{x_i}(v)|\ \le\delta\ .$$

To complete the proof, observe that the constant $M \stackrel{\triangle}{=} \sup_{x\in\Gamma} f(x)$ belongs to $\mathcal{G}$, and hence so does $g_i(\cdot) = g_{x_i}(\cdot) \wedge M$, while for all $v \in \Gamma$,

$$|f(v) - \max_{i=1}^{d} g_i(v)| \le |f(v) - \max_{i=1}^{d} g_{x_i}(v)|\,. \qquad \square$$

The following variant of Theorem 4.4.2 dispenses with the exponential tightness of $\{\mu_\epsilon\}$, assuming instead that (4.4.4) holds for some good rate function $I(\cdot)$. See Section 6.6 for an application of this result.

Theorem 4.4.13 *Let $I(\cdot)$ be a good rate function. A family of probability measures $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{X}$ with the rate function $I(\cdot)$ if and only if the limit Λ_f in (4.4.1) exists for every $f \in C_b(\mathcal{X})$ and satisfies (4.4.4).*

Proof: Suppose first that $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{X}$ with the good rate function $I(\cdot)$. Then, by Varadhan's Lemma (Theorem 4.3.1), the limit Λ_f in (4.4.1) exists for every $f \in C_b(\mathcal{X})$ and satisfies (4.4.4).

Conversely, suppose that the limit Λ_f in (4.4.1) exists for every $f \in C_b(\mathcal{X})$ and satisfies (4.4.4) for some good rate function $I(\cdot)$. The relation (4.4.4) implies that $\Lambda_f - f(x) \ge -I(x)$ for any $x \in \mathcal{X}$ and any $f \in C_b(\mathcal{X})$. Therefore, by Lemma 4.4.6, the existence of Λ_f implies that $\{\mu_\epsilon\}$ satisfies the large deviations lower bound, with the good rate function $I(\cdot)$. Turning to prove the complementary upper bound, it suffices to consider closed sets $F \subset \mathcal{X}$ for which $\inf_{x\in F} I(x) > 0$. Fix such a set and $\delta > 0$ small enough so that $\alpha \stackrel{\triangle}{=} \inf_{x\in F} I^\delta(x) \in (0,\infty)$ for the δ-rate function $I^\delta(\cdot) = \min\{I(\cdot) - \delta, \frac{1}{\delta}\}$. With $\Lambda_0 = 0$, the relation (4.4.4) implies that $\Psi_I(\alpha)$ is non-empty. Since F and $\Psi_I(\alpha)$ are disjoint subsets of the completely regular topological space $\mathcal{X}$, for any $y \in \Psi_I(\alpha)$ there exists a continuous function $f_y : \mathcal{X} \to [0,1]$ such that $f_y(y) = 1$ and $f_y(x) = 0$ for all $x \in F$. The neighborhoods $U_y \stackrel{\triangle}{=} \{z : f_y(z) > 1/2\}$ form a cover of $\Psi_I(\alpha)$; hence, the compact set $\Psi_I(\alpha)$ may be covered by a finite collection $U_{y_1}, \ldots, U_{y_n}$ of such neighborhoods. For any $m \in \mathbb{Z}_+$, the non-negative function $h_m(\cdot) \stackrel{\triangle}{=} 2m \max_{i=1}^{n} f_{y_i}(\cdot)$ is continuous and bounded, with $h_m(x) = 0$ for all $x \in F$ and $h_m(y) \ge m$ for all $y \in \Psi_I(\alpha)$. Therefore, by (4.4.4),

$$\begin{aligned}\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(F) &\le \limsup_{\epsilon\to 0} \epsilon \log \int_{\mathcal{X}} e^{-h_m(x)/\epsilon} \mu_\epsilon(dx) \\ &= \Lambda_{-h_m} = -\inf_{x\in\mathcal{X}} \{h_m(x) + I(x)\}\,.\end{aligned}$$

Note that $h_m(x) + I(x) \ge m$ for any $x \in \Psi_I(\alpha)$, whereas $h_m(x) + I(x) \ge \alpha$ for any $x \notin \Psi_I(\alpha)$. Consequently, taking $m \ge \alpha$,

$$\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(F) \le -\alpha\,.$$

Since $\delta > 0$ is arbitrarily small, the large deviations upper bound holds (see (1.2.11)). $\square$

Exercise 4.4.14 Let $\{\mu_\epsilon\}$ be an exponentially tight family of probability measures on a completely regular topological space $\mathcal{X}$. Let $\mathcal{G}$ be a well-separating class of real valued, continuous functions on $\mathcal{X}$, and let $\mathcal{G}^+$ denote the functions in $\mathcal{G}$ that are bounded above.
(a) Suppose that Λ_g exists for all $g \in \mathcal{G}^+$. For $g \notin \mathcal{G}^+$, define

$$\Lambda_g = \liminf_{\epsilon \to 0} \epsilon \log \int_{\mathcal{X}} e^{g(x)/\epsilon} \mu_\epsilon(dx) \,.$$

Let $\hat{I}(x) = \sup_{g \in \mathcal{G}^+} \{g(x) - \Lambda_g\}$ and show that

$$\hat{I}(x) = \sup_{g \in \mathcal{G}} \{g(x) - \Lambda_g\} \,.$$

Hint: Observe that for every $g \in \mathcal{G}$ and every constant $M < \infty$, both $g(x) \wedge M \in \mathcal{G}^+$ and $\Lambda_{g \wedge M} \leq \Lambda_g$.
(b) Note that $\mathcal{G}^+$ is well-separating, and hence $\{\mu_\epsilon\}$ satisfies the LDP with the good rate function

$$I(x) = \sup_{f \in C_b(\mathcal{X})} \{f(x) - \Lambda_f\} \,.$$

Prove that $I(\cdot) = \hat{I}(\cdot)$.
Hint: Varadhan's lemma applies to every $g \in \mathcal{G}^+$. Consequently, $I(x) \geq \hat{I}(x)$. Fix $x \in \mathcal{X}$ and $f \in C_b(\mathcal{X})$. Following the proof of Theorem 4.4.10 with the compact set Γ enlarged to ensure that $x \in \Gamma$, show that

$$f(x) - \Lambda_f \leq \sup_{d < \infty} \sup_{g_i \in \mathcal{G}^+} \left\{ \max_{i=1}^{d} g_i(x) - \max_{i=1}^{d} \Lambda_{g_i} \right\} = \hat{I}(x) \,.$$

(c) To derive the converse of Theorem 4.4.10, suppose now that $\{\mu_\epsilon\}$ satisfies the LDP with rate function $I(\cdot)$. Use Varadhan's lemma to deduce that Λ_g exists for all $g \in \mathcal{G}^+$, and consequently by parts (a) and (b) of this exercise,

$$I(x) = \sup_{g \in \mathcal{G}} \{g(x) - \Lambda_g\} \,.$$

Exercise 4.4.15 Suppose the topological space $\mathcal{X}$ has a countable base. Let $\mathcal{G}$ be a class of continuous, bounded above, real valued functions on $\mathcal{X}$ such that for any good rate function $J(\cdot)$,

$$J(y) \leq \sup_{g \in \mathcal{G}} \inf_{x \in \mathcal{X}} \{\, g(y) - g(x) + J(x) \,\} \,. \tag{4.4.16}$$

(a) Suppose the family of probability measures $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{X}$ with a good rate function $I(\cdot)$. Then, by Varadhan's Lemma, Λ_g exists for

$g \in \mathcal{G}$ and is given by (4.4.4). Show that $I(\cdot) = \hat{I}(\cdot) \triangleq \sup_{g \in \mathcal{G}} \{g(\cdot) - \Lambda_g\}$.
(b) Suppose $\{\mu_\epsilon\}$ is an exponentially tight family of probability measures, such that Λ_g exists for any $g \in \mathcal{G}$. Show that $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{X}$ with the good rate function $\hat{I}(\cdot)$.
Hint: By Lemma 4.1.23 for any sequence $\epsilon_n \to 0$, there exists a subsequence $n(k) \to \infty$ such that $\{\mu_{\epsilon_{n(k)}}\}$ satisfies the LDP with a good rate function. Use part (a) to show that this good rate function is independent of $\epsilon_n \to 0$.
(c) Show that (4.4.16) holds if for any compact set $K \subset \mathcal{X}$, $y \notin K$ and $\alpha, \delta > 0$, there exists $g \in \mathcal{G}$ such that $\sup_{x \in \mathcal{X}} g(x) \leq g(y) + \delta$ and $\sup_{x \in K} g(x) \leq g(y) - \alpha$.
Hint: Consider $g \in \mathcal{G}$ corresponding to $K = \Psi_J(\alpha)$, $\alpha \nearrow J(y)$ and $\delta \to 0$.
(d) Use part (c) to verify that (4.4.16) holds for $\mathcal{G} = C_b(\mathcal{X})$ and $\mathcal{X}$ a completely regular topological space, thus providing an alternative proof of Theorem 4.4.2 under somewhat stronger conditions.
Hint: See the construction of $h_m(\cdot)$ in Theorem 4.4.13.

Exercise 4.4.17 Complete the proof of Lemma 4.4.5.

4.5 LDP in Topological Vector Spaces

In Section 2.3, it was shown that when a limiting logarithmic moment generating function exists for a family of $\mathbb{R}^d$-valued random variables, then its Fenchel–Legendre transform is the natural candidate rate function for the LDP associated with these variables. The goal of this section is to extend this result to topological vector spaces. As will be seen, convexity plays a major role as soon as the linear structure is introduced. For this reason, after the upper bound is established for all compact sets in Section 4.5.1, Section 4.5.2 turns to the study of some generalities involving the convex duality of Λ and Λ^*. These convexity considerations play an essential role in applications. Finally, Section 4.5.3 is devoted to a direct derivation of a weak version of the Gärtner–Ellis theorem in an abstract setup (Theorem 4.5.20), and to a Banach space variant of it.

Throughout this section, $\mathcal{X}$ is a Hausdorff (real) topological *vector* space. Recall that such spaces are regular, so the results of Sections 4.1 and 4.3 apply. The dual space of $\mathcal{X}$, namely, the space of all *continuous linear functionals* on $\mathcal{X}$, is denoted throughout by $\mathcal{X}^*$. Let Z_ϵ be a family of random variables taking values in $\mathcal{X}$, and let $\mu_\epsilon \in M_1(\mathcal{X})$ denote the probability measure associated with Z_ϵ. By analogy with the $\mathbb{R}^d$ case presented in Section 2.3, the logarithmic moment generating function $\Lambda_{\mu_\epsilon} : \mathcal{X}^* \to (-\infty, \infty]$ is defined to be

$$\Lambda_{\mu_\epsilon}(\lambda) = \log E\left[e^{\langle \lambda, Z_\epsilon \rangle}\right] = \log \int_{\mathcal{X}} e^{\lambda(x)} \mu_\epsilon(dx)\,, \quad \lambda \in \mathcal{X}^*\,,$$

where for $x \in \mathcal{X}$ and $\lambda \in \mathcal{X}^*$, $\langle \lambda, x \rangle$ denotes the value of $\lambda(x) \in \mathbb{R}$.

Let

$$\bar{\Lambda}(\lambda) \stackrel{\triangle}{=} \limsup_{\epsilon \to 0} \epsilon \Lambda_{\mu_\epsilon}\left(\frac{\lambda}{\epsilon}\right) , \qquad (4.5.1)$$

using the notation $\Lambda(\lambda)$ whenever the *limit exists*. In most of the examples considered in Chapter 2, when $\epsilon \Lambda_{\mu_\epsilon}(\cdot/\epsilon)$ converges pointwise to $\Lambda(\cdot)$ for $\mathcal{X} = \mathbb{R}^d$ and an LDP holds for $\{\mu_\epsilon\}$, the rate function associated with this LDP is the Fenchel–Legendre transform of $\Lambda(\cdot)$. In the current setup, the Fenchel–Legendre transform of a function $f : \mathcal{X}^* \to [-\infty, \infty]$ is defined as

$$f^*(x) \stackrel{\triangle}{=} \sup_{\lambda \in \mathcal{X}^*} \{\langle \lambda, x \rangle - f(\lambda)\} , \quad x \in \mathcal{X} . \qquad (4.5.2)$$

Thus, $\bar{\Lambda}^*$ denotes the Fenchel–Legendre transform of $\bar{\Lambda}$, and Λ^* denotes that of Λ when the latter exists for all $\lambda \in \mathcal{X}^*$.

4.5.1 A General Upper Bound

As in the $\mathbb{R}^d$ case, $\bar{\Lambda}^*$ plays a prominent role in the LDP bounds.

Theorem 4.5.3
(a) $\bar{\Lambda}(\cdot)$ of (4.5.1) is convex on $\mathcal{X}^*$ and $\bar{\Lambda}^*(\cdot)$ is a convex rate function.
(b) For any compact set $\Gamma \subset \mathcal{X}$,

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(\Gamma) \le - \inf_{x \in \Gamma} \bar{\Lambda}^*(x) . \qquad (4.5.4)$$

Remarks:
(a) In Theorem 2.3.6, which corresponds to $\mathcal{X} = \mathbb{R}^d$, it was assumed, for the purpose of establishing exponential tightness, that $0 \in \mathcal{D}_\Lambda^o$. In the abstract setup considered here, the exponential tightness does not follow from this assumption, and therefore must be handled on a case-by-case basis. (See, however, [deA85a] for a criterion for exponential tightness which is applicable in a variety of situations.)
(b) Note that any bound of the form $\bar{\Lambda}(\lambda) \le K(\lambda)$ for all $\lambda \in \mathcal{X}^*$ implies that the Fenchel–Legendre transform $K^*(\cdot)$ may be substituted for $\bar{\Lambda}^*(\cdot)$ in (4.5.4). This is useful in situations in which $\bar{\Lambda}(\lambda)$ is easy to bound but hard to compute.
(c) The inequality (4.5.4) may serve as the upper bound related to a weak LDP. Thus, when $\{\mu_\epsilon\}$ is an exponentially tight family of measures, (4.5.4) extends to all closed sets. If in addition, the large deviations lower bound is also satisfied with $\bar{\Lambda}^*(\cdot)$, then this is a good rate function that controls the large deviations of the family $\{\mu_\epsilon\}$.

Proof: (a) The proof is similar to the proof of these properties in the special case $\mathcal{X} = \mathbb{R}^d$, which is presented in the context of the Gärtner–Ellis theorem.

Using the linearity of (λ/ϵ) and applying Hölder's inequality, one shows that the functions $\Lambda_{\mu_\epsilon}(\lambda/\epsilon)$ are convex. Thus, $\bar{\Lambda}(\cdot) = \limsup_{\epsilon\to 0} \epsilon\Lambda_{\mu_\epsilon}(\cdot/\epsilon)$, is also a convex function. Since $\Lambda_{\mu_\epsilon}(0) = 0$ for all $\epsilon > 0$, it follows that $\bar{\Lambda}(0) = 0$. Consequently, $\bar{\Lambda}^*(\cdot)$ is a nonnegative function. Since the supremum of a family of continuous functions is lower semicontinuous, the lower semicontinuity of $\bar{\Lambda}^*(\cdot)$ follows from the continuity of $g_\lambda(x) = \langle \lambda, x\rangle - \bar{\Lambda}(\lambda)$ for every $\lambda \in \mathcal{X}^*$. The convexity of $\bar{\Lambda}^*(\cdot)$ is a direct consequence of its definition via (4.5.2).
(b) The proof of the upper bound (4.5.4) is a repeat of the relevant part of the proof of Theorem 2.2.30. In particular, fix a compact set $\Gamma \subset \mathcal{X}$ and a $\delta > 0$. Let I^δ be the δ-rate function associated with $\bar{\Lambda}^*$, i.e., $I^\delta(x) \triangleq \min\{\bar{\Lambda}^*(x) - \delta, 1/\delta\}$. Then, for any $x \in \Gamma$, there exists a $\lambda_x \in \mathcal{X}^*$ such that

$$\langle \lambda_x, x\rangle - \bar{\Lambda}(\lambda_x) \geq I^\delta(x)\,.$$

Since λ_x is a continuous functional, there exists a neighborhood of x, denoted A_x, such that

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

For any $\theta \in \mathcal{X}^*$, by Chebycheff's inequality,

$$\mu_\epsilon(A_x) \leq E\left[e^{\langle \theta, Z_\epsilon\rangle - \langle \theta, x\rangle}\right] \exp\left(-\inf_{y\in A_x}\{\langle \theta, y\rangle - \langle \theta, x\rangle\}\right)\,.$$

Substituting $\theta = \lambda_x/\epsilon$ yields

$$\epsilon \log \mu_\epsilon(A_x) \leq \delta - \left\{\langle \lambda_x, x\rangle - \epsilon\Lambda_{\mu_\epsilon}\left(\frac{\lambda_x}{\epsilon}\right)\right\}\,.$$

A finite cover, $\cup_{i=1}^N A_{x_i}$, can be extracted from the open cover $\cup_{x\in\Gamma} A_x$ of the compact set Γ. Therefore, by the union of events bound,

$$\epsilon \log \mu_\epsilon(\Gamma) \leq \epsilon \log N + \delta - \min_{i=1,\ldots,N}\left\{\langle \lambda_{x_i}, x_i\rangle - \epsilon\Lambda_{\mu_\epsilon}\left(\frac{\lambda_{x_i}}{\epsilon}\right)\right\}\,.$$

Thus, by (4.5.1) and the choice of λ_x,

$$\begin{aligned}\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(\Gamma) &\leq \delta - \min_{i=1,\ldots,N}\{\langle \lambda_{x_i}, x_i\rangle - \bar{\Lambda}(\lambda_{x_i})\} \\ &\leq \delta - \min_{i=1,\ldots,N} I^\delta(x_i)\,.\end{aligned}$$

Moreover, $x_i \in \Gamma$ for each i, yielding the inequality

$$\limsup_{\epsilon \to 0} \; \epsilon \log \, \mu_\epsilon(\Gamma) \le \delta - \inf_{x \in \Gamma} I^\delta(x) \, .$$

The proof of the theorem is complete by taking $\delta \to 0$. $\square$

Exercise 4.5.5 An upper bound, valid for all ϵ, is developed in this exercise. This bound may be made specific in various situations (*c.f.* Exercise 6.2.19).
(a) Let $\mathcal{X}$ be a Hausdorff topological vector space and $V \subset \mathcal{X}$ a compact, convex set. Prove that for any $\epsilon > 0$,

$$\mu_\epsilon(V) \le \exp\left(-\frac{1}{\epsilon} \inf_{x \in V} \Lambda_\epsilon^*(x) \right) \, , \tag{4.5.6}$$

where

$$\Lambda_\epsilon^*(x) = \sup_{\lambda \in \mathcal{X}^*} \left\{ \langle \lambda, x \rangle - \epsilon \Lambda_{\mu_\epsilon}\left(\frac{\lambda}{\epsilon} \right) \right\} \, .$$

Hint: Recall the following version of the *min–max theorem* ([Sio58], Theorem 4.2'). Let $f(x, \lambda)$ be concave in λ and convex and lower semicontinuous in x. Then

$$\sup_{\lambda \in \mathcal{X}^*} \inf_{x \in V} f(x, \lambda) = \inf_{x \in V} \sup_{\lambda \in \mathcal{X}^*} f(x, \lambda) \, .$$

To prove (4.5.6), first use Chebycheff's inequality and then apply the min–max theorem to the function

$$f(x, \lambda) = [\langle \lambda, x \rangle - \epsilon \Lambda_{\mu_\epsilon}(\lambda/\epsilon)].$$

(b) Suppose that $\mathcal{E}$ is a convex metric subspace of $\mathcal{X}$ (in a metric compatible with the induced topology). Assume that all balls in $\mathcal{E}$ are convex, pre-compact subsets of $\mathcal{X}$. Show that for every measurable set $A \in \mathcal{E}$,

$$\mu_\epsilon(A) \le \inf_{\delta > 0} \left\{ m(A, \delta) \exp\left(-\frac{1}{\epsilon} \inf_{x \in A^\delta} \Lambda_\epsilon^*(x) \right) \right\} \, , \tag{4.5.7}$$

where A^δ is the closed δ blowup of A, and $m(A, \delta)$ denotes the *metric entropy* of A, i.e., the minimal number of balls of radius δ needed to cover A.

4.5.2 Convexity Considerations

The implications of the existence of an LDP with a convex rate function to the structure of Λ and Λ^* are explored here. Building on Varadhan's lemma and Theorem 4.5.3, it is first shown that when the quantities $\epsilon \Lambda_{\mu_\epsilon}(\lambda/\epsilon)$ are uniformly bounded (in ϵ) and an LDP holds with a good convex rate function, then $\epsilon\Lambda_{\mu_\epsilon}(\cdot/\epsilon)$ converges pointwise to $\Lambda(\cdot)$ and the rate function equals $\Lambda^*(\cdot)$. Consequently, the assumptions of Lemma 4.1.21 together with the exponential tightness of $\{\mu_\epsilon\}$ and the uniform boundedness mentioned earlier, suffice to establish the LDP with rate function $\Lambda^*(\cdot)$. Alternatively, if the relation (4.5.15) between I and Λ^* holds, then $\Lambda^*(\cdot)$ controls a weak LDP even when $\Lambda(\lambda) = \infty$ for some λ and $\{\mu_\epsilon\}$ are not exponentially tight. This statement is the key to Cramér's theorem at its most general.

Before proceeding with the attempt to identify the rate function of the LDP as $\Lambda^*(\cdot)$, note that while $\Lambda^*(\cdot)$ is always convex by Theorem 4.5.3, the rate function may well be non-convex. For example, such a situation may occur when contractions using non-convex functions are considered. However, it may be expected that $I(\cdot)$ is identical to $\Lambda^*(\cdot)$ when $I(\cdot)$ is convex.

An instrumental tool in the identification of I as Λ^* is the following duality property of the Fenchel–Legendre transform, whose proof is deferred to the end of this section.

Lemma 4.5.8 (Duality lemma) *Let $\mathcal{X}$ be a locally convex Hausdorff topological vector space. Let $f : \mathcal{X} \to (-\infty, \infty]$ be a lower semicontinuous, convex function, and define*

$$g(\lambda) = \sup_{x\in\mathcal{X}} \{\langle \lambda, x\rangle - f(x)\}\,.$$

Then $f(\cdot)$ is the Fenchel–Legendre transform of $g(\cdot)$, i.e.,

$$f(x) = \sup_{\lambda\in\mathcal{X}^*} \{\langle \lambda, x\rangle - g(\lambda)\}\,. \tag{4.5.9}$$

Remark: This lemma has the following geometric interpretation. For every hyperplane defined by λ, $g(\lambda)$ is the largest amount one may push up the tangent before it hits $f(\cdot)$ and becomes a tangent hyperplane. The duality lemma states the "obvious result" that to reconstruct $f(\cdot)$, one only needs to find the tangent at x and "push it down" by $g(\lambda)$. (See Fig. 4.5.2.)

The first application of the duality lemma is in the following theorem, where convex rate functions are identified as $\Lambda^*(\cdot)$.

Theorem 4.5.10 *Let $\mathcal{X}$ be a locally convex Hausdorff topological vector space. Assume that μ_ϵ satisfies the LDP with a good rate function I. Suppose in addition that*

$$\bar{\Lambda}(\lambda) \stackrel{\triangle}{=} \limsup_{\epsilon\to 0}\, \epsilon\Lambda_{\mu_\epsilon}(\lambda/\epsilon) < \infty, \quad \forall \lambda \in \mathcal{X}^*\,. \tag{4.5.11}$$

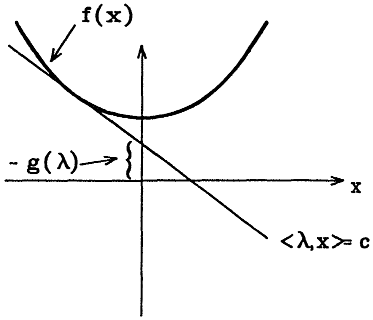

Figure 4.5.1: Duality lemma.

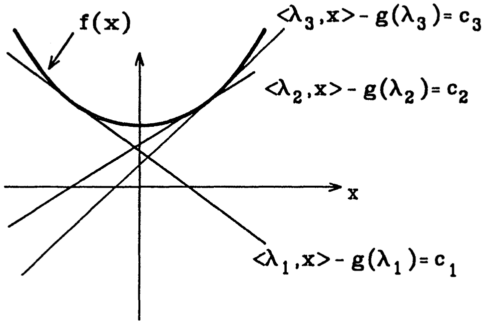

Figure 4.5.2: Duality reconstruction. $c_i = f(x_i)$ and x_i is the point of tangency of the line with slope λ_i to the graph of $f(\cdot)$.

(a) For each $\lambda \in \mathcal{X}^*$, the limit $\Lambda(\lambda) = \lim_{\epsilon \to 0} \epsilon \Lambda_{\mu_\epsilon}(\lambda/\epsilon)$ exists, is finite, and satisfies

$$\Lambda(\lambda) = \sup_{x \in \mathcal{X}} \{\langle \lambda, x \rangle - I(x)\} . \tag{4.5.12}$$

(b) If I is convex, then it is the Fenchel–Legendre transform of Λ, namely,

$$I(x) = \Lambda^*(x) \stackrel{\triangle}{=} \sup_{\lambda \in \mathcal{X}^*} \{\langle \lambda, x \rangle - \Lambda(\lambda)\} .$$

(c) If I is not convex, then Λ^* is the affine regularization of I, i.e., $\Lambda^*(\cdot) \leq I(\cdot)$, and for any convex rate function f, $f(\cdot) \leq I(\cdot)$ implies $f(\cdot) \leq \Lambda^*(\cdot)$. (See Fig. 4.5.3.)

Remark: The weak* topology on $\mathcal{X}^*$ makes the functions $\langle \lambda, x \rangle - I(x)$ continuous in λ for all $x \in \mathcal{X}$. By part (a), $\Lambda(\cdot)$ is lower semicontinuous with respect to this topology, which explains why lower semicontinuity of $\Lambda(\cdot)$ is necessary in Rockafellar's lemma (Lemma 2.3.12).

Proof: (a) Fix $\lambda \in \mathcal{X}^*$ and $\gamma > 1$. By assumption, $\bar{\Lambda}(\gamma\lambda) < \infty$, and Varadhan's lemma (Theorem 4.3.1) applies for the continuous function

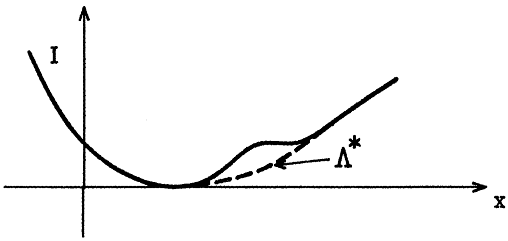

Figure 4.5.3: Λ^* as affine regularization of I.

$\lambda : \mathcal{X} \to \mathbb{R}$. Thus, $\Lambda(\lambda) = \lim_{\epsilon\to 0} \epsilon\Lambda_{\mu_\epsilon}(\lambda/\epsilon)$ exists, and satisfies the identity (4.5.12). By the assumption (4.5.11), $\Lambda(\cdot) < \infty$ everywhere. Since $\Lambda(0) = 0$ and $\Lambda(\cdot)$ is convex by part (a) of Theorem 4.5.3, it also holds that $\Lambda(\lambda) > -\infty$ everywhere.
(b) This is a direct consequence of the duality lemma (Lemma 4.5.8), applied to the lower semicontinuous, convex function I.
(c) The proof of this part of the theorem is left as Exercise 4.5.18. □

Corollary 4.5.13 *Suppose that both condition (4.5.11) and the assumptions of Lemma 4.1.21 hold for the family $\{\mu_\epsilon\}$, which is exponentially tight. Then $\{\mu_\epsilon\}$ satisfies in $\mathcal{X}$ the LDP with the good, convex rate function Λ^*.*

Proof: By Lemma 4.1.21, $\{\mu_\epsilon\}$ satisfies a weak LDP with a convex rate function. As $\{\mu_\epsilon\}$ is exponentially tight, it is deduced that it satisfies the *full* LDP with a convex, good rate function. The corollary then follows from parts (a) and (b) of Theorem 4.5.10. □

Theorem 4.5.10 is not applicable when $\Lambda(\cdot)$ exists but is infinite at some $\lambda \in \mathcal{X}^*$, and moreover, it requires the *full* LDP with a convex, *good* rate function. As seen in the case of Cramér's theorem in $\mathbb{R}$, these conditions are not necessary. The following theorem replaces the finiteness conditions on Λ by an appropriate inequality on *open* half-spaces. Of course, there is a price to pay: The resulting Λ^* may not be a good rate function and only the weak LDP is proved.

Theorem 4.5.14 *Suppose that $\{\mu_\epsilon\}$ satisfies a weak LDP with a convex rate function $I(\cdot)$, and that $\mathcal{X}$ is a locally convex, Hausdorff topological vector space. Assume that for each $\lambda \in \mathcal{X}^*$, the limits $\Lambda_\lambda(t) = \lim_{\epsilon\to 0} \epsilon\Lambda_{\mu_\epsilon}(t\lambda/\epsilon)$ exist as extended real numbers, and that $\Lambda_\lambda(t)$ is a lower semicontinuous function of $t \in \mathbb{R}$. Let $\Lambda^*_\lambda(\cdot)$ be the Fenchel–Legendre* *transform of* $\Lambda_\lambda(\cdot)$*, i.e.,*

$$\Lambda_\lambda^*(z) \stackrel{\triangle}{=} \sup_{\theta \in \mathbb{R}} \{ \theta z - \Lambda_\lambda(\theta) \} .$$

If for every $\lambda \in \mathcal{X}^*$ *and every* $a \in \mathbb{R}$*,*

$$\inf_{\{x : (\langle \lambda, x \rangle - a) > 0\}} I(x) \le \inf_{z > a} \Lambda_\lambda^*(z) , \tag{4.5.15}$$

then $I(\cdot) = \Lambda^*(\cdot)$*, and consequently,* Λ^* *controls a weak LDP associated with* $\{\mu_\epsilon\}$*.*

Proof: Fix $\lambda \in \mathcal{X}^*$. By the inequality (4.5.15),

$$\begin{aligned} \sup_{x \in \mathcal{X}} \{ \langle \lambda, x \rangle - I(x) \} &= \sup_{a \in \mathbb{R}} \sup_{\{x : (\langle \lambda, x \rangle - a) > 0\}} \{ \langle \lambda, x \rangle - I(x) \} \\ &\ge \sup_{a \in \mathbb{R}} \left\{ a - \inf_{\{x : (\langle \lambda, x \rangle - a) > 0\}} I(x) \right\} \qquad (4.5.16) \\ &\ge \sup_{a \in \mathbb{R}} \left\{ a - \inf_{z > a} \Lambda_\lambda^*(z) \right\} = \sup_{z \in \mathbb{R}} \left\{ z - \Lambda_\lambda^*(z) \right\} . \end{aligned}$$

Note that $\Lambda_\lambda(\cdot)$ is convex with $\Lambda_\lambda(0) = 0$ and is assumed lower semicontinuous. Therefore, it can not attain the value $-\infty$. Hence, by applying the duality lemma (Lemma 4.5.8) to $\Lambda_\lambda : \mathbb{R} \to (-\infty, \infty]$, it follows that

$$\Lambda_\lambda(1) = \sup_{z \in \mathbb{R}} \{ z - \Lambda_\lambda^*(z) \} .$$

Combining this identity with (4.5.16) yields

$$\sup_{x \in \mathcal{X}} \{ \langle \lambda, x \rangle - I(x) \} \ge \Lambda_\lambda(1) = \Lambda(\lambda) .$$

The opposite inequality follows by applying Lemma 4.3.4 to the continuous linear functional $\lambda \in \mathcal{X}^*$. Thus, the identity (4.5.12) holds for all $\lambda \in \mathcal{X}^*$, and the proof of the theorem is completed by applying the duality lemma (Lemma 4.5.8) to the convex rate function I. $\square$

Proof of Lemma 4.5.8: Consider the sets $\mathcal{X} \times \mathbb{R}$ and $\mathcal{X}^* \times \mathbb{R}$. Each of these can be made into a locally convex, Hausdorff topological vector space in the obvious way. If f is identically ∞, then g is identically $-\infty$ and the lemma trivially holds. Assume otherwise and define

$$\begin{aligned} \mathcal{E} &= \{ (x, \alpha) \in \mathcal{X} \times \mathbb{R} : f(x) \le \alpha \} , \\ \mathcal{E}^* &= \{ (\lambda, \beta) \in \mathcal{X}^* \times \mathbb{R} : g(\lambda) \le \beta \} . \end{aligned}$$

Note that for any $(\lambda, \beta) \in \mathcal{E}^*$ and any $x \in \mathcal{X}$,

$$f(x) \ge \langle \lambda, x \rangle - \beta .$$

Therefore, it also holds that

$$f(x) \geq \sup_{(\lambda,\beta)\in\mathcal{E}^*} \{\langle\lambda, x\rangle - \beta\} = \sup_{\lambda\in\mathcal{X}^*} \{\langle\lambda, x\rangle - g(\lambda)\} .$$

It thus suffices to show that for any $(x, \alpha) \notin \mathcal{E}$ (i.e., $f(x) > \alpha$), there exists a $(\lambda, \beta) \in \mathcal{E}^*$ such that

$$\langle\lambda, x\rangle - \beta > \alpha , \tag{4.5.17}$$

in order to complete the proof of the lemma.

Since f is a lower semicontinuous function, the set $\mathcal{E}$ is closed (alternatively, the set $\mathcal{E}^c$ is open). Indeed, whenever $f(x) > \gamma$, there exists a neighborhood V of x such that $\inf_{y\in V} f(y) > \gamma$, and thus $\mathcal{E}^c$ contains a neighborhood of (x, γ). Moreover, since $f(\cdot)$ is convex and not identically ∞, the set $\mathcal{E}$ is a non-empty convex subset of $\mathcal{X} \times \mathbb{R}$.

Fix $(x, \alpha) \notin \mathcal{E}$. The product space $\mathcal{X}\times\mathbb{R}$ is locally convex and therefore, by the Hahn–Banach theorem (Theorem B.6), there exists a hyperplane in $\mathcal{X} \times \mathbb{R}$ that strictly separates the non-empty, closed, and convex set $\mathcal{E}$ and the point (x, α) in its complement. Hence, as the topological dual of $\mathcal{X} \times \mathbb{R}$ is $\mathcal{X}^* \times \mathbb{R}$, for some $\mu \in \mathcal{X}^*$, $\rho \in \mathbb{R}$, and $\gamma \in \mathbb{R}$,

$$\sup_{(y,\xi)\in\mathcal{E}} \{\langle\mu, y\rangle - \rho\xi\} \leq \gamma < \langle\mu, x\rangle - \rho\alpha .$$

In particular, since f is not identically ∞, it follows that $\rho \geq 0$, for otherwise a contradiction results when $\xi \to \infty$. Moreover, by considering $(y, \xi) = (x, f(x))$, the preceding inequality implies that $\rho > 0$ whenever $f(x) < \infty$.

Suppose first that $\rho > 0$. Then, (4.5.17) holds for the point $(\mu/\rho, \gamma/\rho)$. This point must be in $\mathcal{E}^*$, for otherwise there exists a $y_0 \in \mathcal{X}$ such that $\langle\mu, y_0\rangle - \rho f(y_0) > \gamma$, contradicting the previous construction of the separating hyperplane (since $(y_0, f(y_0)) \in \mathcal{E}$). In particular, since $f(x) < \infty$ for some $x \in \mathcal{X}$ it follows that $\mathcal{E}^*$ is non-empty.

Now suppose that $\rho = 0$ so that

$$\sup_{\{y:f(y)<\infty\}} \{\langle\mu, y\rangle - \gamma\} \leq 0 ,$$

while $\langle\mu, x\rangle - \gamma > 0$. Consider the points

$$(\lambda_\delta, \beta_\delta) \stackrel{\triangle}{=} \left(\frac{\mu}{\delta} + \lambda_0, \frac{\gamma}{\delta} + \beta_0\right) , \quad \forall\delta > 0 ,$$

where (λ_0, β_0) is an arbitrary point in $\mathcal{E}^*$. Then, for all $y \in \mathcal{X}$,

$$\langle\lambda_\delta, y\rangle - \beta_\delta = \frac{1}{\delta}(\langle\mu, y\rangle - \gamma) + (\langle\lambda_0, y\rangle - \beta_0) \leq f(y) .$$

Therefore, $(\lambda_\delta, \beta_\delta) \in \mathcal{E}^*$ for any $\delta > 0$. Moreover,

$$\lim_{\delta \to 0} (\langle \lambda_\delta, x \rangle - \beta_\delta) = \lim_{\delta \to 0} \left\{ \frac{1}{\delta} (\langle \mu, x \rangle - \gamma) + (\langle \lambda_0, x \rangle - \beta_0) \right\} = \infty .$$

Thus, for any $\alpha < \infty$, there exists $\delta > 0$ small enough so that $\langle \lambda_\delta, x \rangle - \beta_\delta > \alpha$. This completes the proof of (4.5.17) and of Lemma 4.5.8. $\square$

Exercise 4.5.18 Prove part (c) of Theorem 4.5.10.

Exercise 4.5.19 Consider the setup of Exercise 4.2.7, except that now $\mathcal{X} = \mathcal{Y}$ is a locally convex, separable, Hausdorff topological vector space. Let $Z_\epsilon = X_\epsilon + Y_\epsilon$.
(a) Prove that if I_X and I_Y are convex, then so is I_Z.
(b) Deduce that if in addition, the condition (4.5.11) holds for both μ_ϵ—the laws of X_ϵ and ν_ϵ—the laws of Y_ϵ, then I_Z is the Fenchel–Legendre transform of $\Lambda_X(\cdot) + \Lambda_Y(\cdot)$.

4.5.3 Abstract Gärtner–Ellis Theorem

Having seen a general upper bound in Section 4.5.1, we turn next to sufficient conditions for the existence of a complementary lower bound. To this end, recall that a point $x \in \mathcal{X}$ is called an *exposed* point of $\bar{\Lambda}^*$ if there exists an *exposing hyperplane* $\lambda \in \mathcal{X}^*$ such that

$$\langle \lambda, x \rangle - \bar{\Lambda}^*(x) > \langle \lambda, z \rangle - \bar{\Lambda}^*(z) , \quad \forall z \neq x .$$

An exposed point of $\bar{\Lambda}^*$ is, in convex analysis parlance, an exposed point of the epigraph of $\bar{\Lambda}^*$. For a geometrical interpretation, see Fig. 2.3.2.

Theorem 4.5.20 (Baldi) *Suppose that $\{\mu_\epsilon\}$ are exponentially tight probability measures on $\mathcal{X}$.*
(a) For every closed set $F \subset \mathcal{X}$,

$$\limsup_{\epsilon \to 0} \epsilon \log \ \mu_\epsilon(F) \leq - \inf_{x \in F} \bar{\Lambda}^*(x) .$$

(b) Let $\mathcal{F}$ be the set of exposed points of $\bar{\Lambda}^*$ with an exposing hyperplane λ for which

$$\Lambda(\lambda) = \lim_{\epsilon \to 0} \epsilon \Lambda_{\mu_\epsilon} \left(\frac{\lambda}{\epsilon} \right) \text{ exists and } \bar{\Lambda}(\gamma \lambda) < \infty \quad \text{for some} \quad \gamma > 1 . \quad (4.5.21)$$

Then, for every open set $G \subset \mathcal{X}$,

$$\liminf_{\epsilon \to 0} \epsilon \log \ \mu_\epsilon(G) \geq - \inf_{x \in G \cap \mathcal{F}} \bar{\Lambda}^*(x) .$$

(c) If for every open set G,

$$\inf_{x \in G \cap \mathcal{F}} \bar{\Lambda}^*(x) = \inf_{x \in G} \bar{\Lambda}^*(x) , \tag{4.5.22}$$

then $\{\mu_\epsilon\}$ satisfies the LDP with the good rate function $\bar{\Lambda}^*$.

Proof: (a) The upper bound is a consequence of Theorem 4.5.3 and the assumed exponential tightness.
(b) If $\bar{\Lambda}(\lambda) = -\infty$ for some $\lambda \in \mathcal{X}^*$, then $\bar{\Lambda}^*(\cdot) \equiv \infty$ and the large deviations lower bound trivially holds. So, without loss of generality, it is assumed throughout that $\bar{\Lambda} : \mathcal{X}^* \to (-\infty, \infty]$. Fix an open set G, an exposed point $y \in G \cap \mathcal{F}$, and $\delta > 0$ arbitrarily small. Let η be an exposing hyperplane for $\bar{\Lambda}^*$ at y such that (4.5.21) holds. The proof is now a repeat of the proof of (2.3.13). Indeed, by the continuity of η, there exists an open subset of G, denoted B_δ, such that $y \in B_\delta$ and

$$\sup_{z \in B_\delta} \{\langle \eta, z - y \rangle\} < \delta .$$

Observe that $\Lambda(\eta) < \infty$ in view of (4.5.21). Hence, by (4.5.1), $\Lambda_{\mu_\epsilon}(\eta/\epsilon) < \infty$ for all ϵ small enough. Thus, for all $\epsilon > 0$ small enough, define the probability measures $\tilde{\mu}_\epsilon$ via

$$\frac{d\tilde{\mu}_\epsilon}{d\mu_\epsilon}(z) = \exp\left[\left\langle \frac{\eta}{\epsilon}, z \right\rangle - \Lambda_{\mu_\epsilon}\left(\frac{\eta}{\epsilon}\right)\right] . \tag{4.5.23}$$

Using this definition,

$$\begin{aligned}\epsilon \log \mu_\epsilon(B_\delta) &= \epsilon\, \Lambda_{\mu_\epsilon}\left(\frac{\eta}{\epsilon}\right) - \langle \eta, y\rangle + \epsilon \log \int_{z \in B_\delta} \exp\left(\left\langle \frac{\eta}{\epsilon}, y - z\right\rangle\right) \tilde{\mu}_\epsilon(dz) \\ &\geq \epsilon \Lambda_{\mu_\epsilon}\left(\frac{\eta}{\epsilon}\right) - \langle \eta, y\rangle - \delta + \epsilon \log \tilde{\mu}_\epsilon(B_\delta) .\end{aligned}$$

Therefore, by (4.5.21),

$$\begin{aligned}\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(G) &\geq \lim_{\delta \to 0} \liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(B_\delta) \\ &\geq \Lambda(\eta) - \langle \eta, y\rangle + \lim_{\delta \to 0} \liminf_{\epsilon \to 0} \epsilon \log \tilde{\mu}_\epsilon(B_\delta) \\ &\geq -\bar{\Lambda}^*(y) + \lim_{\delta \to 0} \liminf_{\epsilon \to 0} \epsilon \log \tilde{\mu}_\epsilon(B_\delta) .\end{aligned} \tag{4.5.24}$$

Recall that $\{\mu_\epsilon\}$ are exponentially tight, so for each $\alpha < \infty$, there exists a compact set K_α such that

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(K_\alpha^c) < -\alpha .$$

If for all $\delta > 0$ and all $\alpha < \infty$,

$$\limsup_{\epsilon \to 0} \epsilon \log \tilde{\mu}_\epsilon(B_\delta^c \cap K_\alpha) < 0 , \tag{4.5.25}$$

and for all α large enough,

$$\limsup_{\epsilon \to 0} \epsilon \log \tilde{\mu}_\epsilon(K_\alpha^c) < 0 , \tag{4.5.26}$$

then $\tilde{\mu}_\epsilon(B_\delta) \to 1$ when $\epsilon \to 0$ and part (b) of the theorem follows by (4.5.24), since $y \in G \cap \mathcal{F}$ is arbitrary.

To establish (4.5.25), let $\Lambda_{\tilde{\mu}_\epsilon}(\cdot)$ denote the logarithmic moment generating function associated with the law $\tilde{\mu}_\epsilon$. By the definition (4.5.23), for every $\theta \in \mathcal{X}^*$,

$$\epsilon \Lambda_{\tilde{\mu}_\epsilon}\left(\frac{\theta}{\epsilon}\right) = \epsilon \Lambda_{\mu_\epsilon}\left(\frac{\theta+\eta}{\epsilon}\right) - \epsilon \Lambda_{\mu_\epsilon}\left(\frac{\eta}{\epsilon}\right) .$$

Hence, by (4.5.1) and (4.5.21),

$$\tilde{\Lambda}(\theta) \stackrel{\triangle}{=} \limsup_{\epsilon \to 0} \epsilon \Lambda_{\tilde{\mu}_\epsilon}\left(\frac{\theta}{\epsilon}\right) = \bar{\Lambda}(\theta+\eta) - \Lambda(\eta) .$$

Let $\tilde{\Lambda}^*$ denote the Fenchel–Legendre transform of $\tilde{\Lambda}$. It follows that for all $z \in \mathcal{X}$,

$$\tilde{\Lambda}^*(z) = \bar{\Lambda}^*(z) + \Lambda(\eta) - \langle \eta, z \rangle \geq \bar{\Lambda}^*(z) - \bar{\Lambda}^*(y) - \langle \eta, z - y \rangle .$$

Since η is an exposing hyperplane for $\bar{\Lambda}^*$ at y, this inequality implies that $\tilde{\Lambda}^*(z) > 0$ for all $z \neq y$. Theorem 4.5.3, applied to the measures $\tilde{\mu}_\epsilon$ and the compact sets $B_\delta^c \cap K_\alpha$, now yields

$$\limsup_{\epsilon \to 0} \epsilon \log \tilde{\mu}_\epsilon(B_\delta^c \cap K_\alpha) \leq - \inf_{z \in B_\delta^c \cap K_\alpha} \tilde{\Lambda}^*(z) < 0 ,$$

where the strict inequality follows because $\tilde{\Lambda}^*(\cdot)$ is a lower semicontinuous function and $y \in B_\delta$.

Turning now to establish (4.5.26), consider the open half-spaces

$$H_\rho = \{ z \in \mathcal{X} : \langle \eta, z \rangle - \rho < 0 \} .$$

By Chebycheff's inequality, for any $\beta > 0$,

$$\begin{aligned} \epsilon \log \tilde{\mu}_\epsilon(H_\rho^c) &= \epsilon \log \int_{\{z : \langle \eta, z \rangle \geq \rho\}} \tilde{\mu}_\epsilon(dz) \\ &\leq \epsilon \log \left[\int_{\mathcal{X}} \exp\left(\frac{\beta \langle \eta, z \rangle}{\epsilon}\right) \tilde{\mu}_\epsilon(dz) \right] - \beta\rho \\ &= \epsilon \Lambda_{\tilde{\mu}_\epsilon}\left(\frac{\beta\eta}{\epsilon}\right) - \beta\rho . \end{aligned}$$

Hence,

$$\limsup_{\epsilon\to 0} \epsilon \log \tilde{\mu}_\epsilon(H_\rho^c) \le \inf_{\beta>0} \{\tilde{\Lambda}(\beta\eta) - \beta\rho\} \,.$$

Due to condition (4.5.21), $\tilde{\Lambda}(\beta\eta) < \infty$ for some $\beta > 0$, implying that for large enough ρ,

$$\limsup_{\epsilon\to 0} \epsilon \log \tilde{\mu}_\epsilon(H_\rho^c) < 0 \,.$$

Now, for every α and every $\rho > 0$,

$$\begin{aligned}
&\limsup_{\epsilon\to 0} \epsilon \log \tilde{\mu}_\epsilon(K_\alpha^c \cap H_\rho) \\
&\quad = \limsup_{\epsilon\to 0} \epsilon \log \int_{K_\alpha^c \cap H_\rho} \exp\left[\left\langle \frac{\eta}{\epsilon}, z\right\rangle - \Lambda_{\mu_\epsilon}\left(\frac{\eta}{\epsilon}\right)\right] \mu_\epsilon(dz) \\
&\quad < \rho - \Lambda(\eta) - \alpha \,.
\end{aligned}$$

Finally, (4.5.26) follows by combining the two preceding inequalities.
(c) Starting with (4.5.22), the LDP is established by combining parts (a) and (b). □

In the following corollary, the smoothness of $\Lambda(\cdot)$ yields the identity (4.5.22) for exponentially tight probability measures on a Banach space, resulting in the LDP. Its proof is based on a theorem of Brønsted and Rockafellar whose proof is not reproduced here. Recall that a function $f : \mathcal{X}^* \to \mathbb{R}$ is *Gateaux differentiable* if, for every $\lambda, \theta \in \mathcal{X}^*$, the function $f(\lambda + t\theta)$ is differentiable with respect to t at $t = 0$.

Corollary 4.5.27 *Let $\{\mu_\epsilon\}$ be exponentially tight probability measures on the Banach space $\mathcal{X}$. Suppose that $\Lambda(\cdot) = \lim_{\epsilon\to 0} \epsilon\Lambda_{\mu_\epsilon}(\cdot/\epsilon)$ is finite valued, Gateaux differentiable, and lower semicontinuous in $\mathcal{X}^*$ with respect to the weak* topology. Then $\{\mu_\epsilon\}$ satisfies the LDP with the good rate function Λ^*.*

Remark: For a somewhat stronger version, see Corollary 4.6.14.

Proof: By Baldi's theorem (Theorem 4.5.20), it suffices to show that for any $x \in \mathcal{D}_{\Lambda^*}$, there exists a sequence of exposed points x_k such that $x_k \to x$ and $\Lambda^*(x_k) \to \Lambda^*(x)$. Let $\lambda \in \partial\Lambda^*(x)$ iff

$$\langle \lambda, x\rangle - \Lambda^*(x) = \sup_{z\in\mathcal{X}} \{\langle \lambda, z\rangle - \Lambda^*(z)\} \,,$$

and define

$$\mathrm{dom}\, \partial\Lambda^* \stackrel{\triangle}{=} \{x : \exists \lambda \in \partial\Lambda^*(x)\} \,.$$

Note that it may be assumed that the convex, lower semicontinuous function $\Lambda^* : \mathcal{X} \to [0,\infty]$ is proper (i.e., $\mathcal{D}_{\Lambda^*}$ is not empty). Therefore, by the Brønsted–Rockafellar theorem (see [BrR65], Theorem 2), for every $x \in \mathcal{D}_{\Lambda^*}$, there exists a sequence $x_k \to x$ such that $x_k \in \operatorname{dom} \partial\Lambda^*$ and $\Lambda^*(x_k) \to \Lambda^*(x)$.

It is therefore enough to prove that when Λ is Gateaux differentiable and weak* lower semicontinuous, any point in $\operatorname{dom} \partial\Lambda^*$ is also an exposed point. To this end, fix $x \in \operatorname{dom}\partial\Lambda^*$ and $\lambda \in \partial\Lambda^*(x)$. Observe that $\mathcal{X}^*$ when equipped with the weak* topology is a locally convex, Hausdorff topological vector space with $\mathcal{X}$ being its topological dual. Hence, it follows by applying the duality lemma (Lemma 4.5.8) for the convex, lower semicontinuous function $\Lambda : \mathcal{X}^* \to \mathbb{R}$ that

$$\Lambda(\lambda) = \sup_{z\in\mathcal{X}}\{\langle\lambda, z\rangle - \Lambda^*(z)\} = \langle\lambda, x\rangle - \Lambda^*(x) \,.$$

Therefore, for any $t > 0$, and any $\theta \in \mathcal{X}^*$,

$$\langle\theta, x\rangle \;\leq \frac{1}{t}[\Lambda(\lambda + t\theta) - \Lambda(\lambda)] \,.$$

Thus, by the Gateaux differentiability of Λ, it follows that

$$\langle\theta, x\rangle \;\leq \lim_{t\searrow 0} \frac{1}{t}[\Lambda(\lambda + t\theta) - \Lambda(\lambda)] \stackrel{\triangle}{=} D\Lambda(\theta) \,.$$

Moreover, $D\Lambda(\theta) = -D\Lambda(-\theta)$, and consequently $\langle\theta, x\rangle = D\Lambda(\theta)$ for all $\theta \in \mathcal{X}^*$. Similarly, if there exists $y \in \mathcal{X}$, $y \neq x$, such that

$$\langle\lambda, x\rangle - \Lambda^*(x) = \langle\lambda, y\rangle - \Lambda^*(y) \,,$$

then, by exactly the same argument, $\langle\theta, y\rangle = D\Lambda(\theta)$ for all $\theta \in \mathcal{X}^*$. Since $\langle\theta, x - y\rangle = 0$ for all $\theta \in \mathcal{X}^*$, it follows that $x = y$. Hence, x is an exposed point and the proof is complete. $\square$

4.6 Large Deviations for Projective Limits

In this section, we develop a method of lifting a collection of LDPs in “small” spaces into the LDP in the “large” space $\mathcal{X}$, which is their projective limit. (See definition below.) The motivation for such an approach is as follows. Suppose we are interested in proving the LDP associated with a sequence of random variables $X_1, X_2, \ldots$ in some abstract space $\mathcal{X}$. The identification of $\mathcal{X}^*$ (if $\mathcal{X}$ is a topological vector space) and the computation of the Fenchel–Legendre transform of the moment generating function may involve the solution of variational problems in an infinite dimensional setting. Moreover, proving exponential tightness in $\mathcal{X}$, the main tool of getting at the upper bound, may be a difficult task. On the other hand, the evaluation of the limiting logarithmic moment generating function involves probabilistic computations at the level of real-valued random variables, albeit an infinite number of such computations. It is often relatively easy to derive the LDP for every finite collection of these real-valued random variables. Hence, it is reasonable to inquire if this implies that the laws of the original, $\mathcal{X}$-valued random variables satisfy the LDP.

An affirmative result is derived shortly in a somewhat abstract setting that will serve us well in diverse situations. The idea is to identify $\mathcal{X}$ with the projective limit of a family of spaces $\{\mathcal{Y}_j\}_{j\in J}$ with the hope that the LDP for any given family $\{\mu_\epsilon\}$ of probability measures on $\mathcal{X}$ follows as the consequence of the fact that the LDP holds for any of the projections of μ_ϵ to $\{\mathcal{Y}_j\}_{j\in J}$.

To make the program described precise, we first review a few standard topological definitions. Let $(J,\leq)$ be a partially ordered, right-filtering set. (The latter notion means that for any i,j in J, there exists $k\in J$ such that both $i\leq k$ and $j\leq k$.) Note that J need not be countable. A projective system $(\mathcal{Y}_j,p_{ij})_{i\leq j\in J}$ consists of Hausdorff topological spaces $\{\mathcal{Y}_j\}_{j\in J}$ and continuous maps $p_{ij}:\mathcal{Y}_j\to\mathcal{Y}_i$ such that $p_{ik}=p_{ij}\circ p_{jk}$ whenever $i\leq j\leq k$ ($\{p_{jj}\}_{j\in J}$ are the appropriate identity maps). The *projective limit* of this system, denoted by $\mathcal{X}=\varprojlim\mathcal{Y}_j$, is the subset of the topological product space $\mathcal{Y}=\prod_{j\in J}\mathcal{Y}_j$, consisting of all the elements $\mathbf{x}=(y_j)_{j\in J}$ for which $y_i=p_{ij}(y_j)$ whenever $i\leq j$, equipped with the topology induced by $\mathcal{Y}$. Projective limits of closed subsets $F_j\subseteq\mathcal{Y}_j$ are defined analogously and denoted $F=\varprojlim F_j$. The canonical projections of $\mathcal{X}$, which are the restrictions $p_j:\mathcal{X}\to\mathcal{Y}_j$ of the coordinate maps from $\mathcal{Y}$ to $\mathcal{Y}_j$, are continuous. Some properties of projective limits are recalled in Appendix B.

The following theorem yields the LDP in $\mathcal{X}$ as a consequence of the LDPs associated with $\{\mu_\epsilon\circ p_j^{-1},\epsilon>0\}$. In order to have a specific example in mind, think of $\mathcal{X}$ as the space of all maps $f:[0,1]\to\mathbb{R}$ such that $f(0)=0$, equipped with the topology of pointwise convergence. Then $p_j:\mathcal{X}\to\mathbb{R}^d$ is the projection of functions onto their values at the time instances $0\leq t_1<t_2<\cdots<t_d\leq 1$, with the partial ordering induced on the set $J=\cup_{d=1}^\infty\{(t_1,\ldots,t_d):0\leq t_1<t_2<\cdots<t_d\leq 1\}$ by inclusions. For details of this construction, see Section 5.1.

Theorem 4.6.1 (Dawson–Gärtner) *Let $\{\mu_\epsilon\}$ be a family of probability measures on $\mathcal{X}$, such that for any $j\in J$ the Borel probability measures $\mu_\epsilon\circ p_j^{-1}$ on $\mathcal{Y}_j$ satisfy the LDP with the good rate function $I_j(\cdot)$. Then $\{\mu_\epsilon\}$ satisfies the LDP with the good rate function*

$$I(\mathbf{x})=\sup_{j\in J}\{\,I_j(p_j(\mathbf{x}))\,\}\,,\qquad \mathbf{x}\in\mathcal{X}\,. \tag{4.6.2}$$

Remark: Throughout this section, we drop the blanket assumption that $\mathcal{B}_{\mathcal{X}} \subseteq \mathcal{B}$. This is natural in view of the fact that the set J need not be countable. It is worthwhile to note that $\mathcal{B}$ is required to contain all sets $p_j^{-1}(B_j)$, where $B_j \in \mathcal{B}_{\mathcal{Y}_j}$.

Proof: Clearly, $I(\mathbf{x})$ is nonnegative. For any $\alpha \in [0, \infty)$ and $j \in J$, let $\Psi_{I_j}(\alpha)$ denote the *compact* level set of I_j, i.e., $\Psi_{I_j}(\alpha) \triangleq \{y_j : I_j(y_j) \leq \alpha\}$. Recall that for any $i \leq j \in J$, $p_{ij} : \mathcal{Y}_j \to \mathcal{Y}_i$ is a continuous map and $\mu_\epsilon \circ p_i^{-1} = (\mu_\epsilon \circ p_j^{-1}) \circ p_{ij}^{-1}$. Hence, by the contraction principle (Theorem 4.2.1), $I_i(y_i) = \inf_{y_j \in p_{ij}^{-1}(y_i)} I_j(y_j)$, or alternatively, $\Psi_{I_i}(\alpha) = p_{ij}(\Psi_{I_j}(\alpha))$. Therefore,

$$\Psi_I(\alpha) = \mathcal{X} \cap \prod_{j \in J} \Psi_{I_j}(\alpha) = \varprojlim \Psi_{I_j}(\alpha) , \tag{4.6.3}$$

and $I(\mathbf{x})$ is a good rate function, since by Tychonoff's theorem (Theorem B.3), the projective limit of compact subsets of $\mathcal{Y}_j$, $j \in J$, is a compact subset of $\mathcal{X}$.

In order to prove the large deviations lower bound, it suffices to show that for every measurable set $A \subset \mathcal{X}$ and each $\mathbf{x} \in A^o$, there exists a $j \in J$ such that

$$\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A) \geq -I_j(p_j(\mathbf{x})) .$$

Since the collection $\{p_j^{-1}(U_j) : U_j \subset \mathcal{Y}_j \text{ is open}\}$ is a base of the topology of $\mathcal{X}$, there exists some $j \in J$ and an open set $U_j \subset \mathcal{Y}_j$ such that $\mathbf{x} \in p_j^{-1}(U_j) \subset A^o$. Thus, by the large deviations lower bound for $\{\mu_\epsilon \circ p_j^{-1}\}$,

$$\begin{aligned}
\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A) &\geq \liminf_{\epsilon \to 0} \epsilon \log(\mu_\epsilon \circ p_j^{-1}(U_j)\,) \\
&\geq -\inf_{y \in U_j} I_j(y) \geq -I_j(p_j(\mathbf{x})) ,
\end{aligned}$$

as desired.

Considering the large deviations upper bound, fix a measurable set $A \subset \mathcal{X}$ and let $A_j \triangleq p_j(\,\overline{A}\,)$. Then, $A_i = p_{ij}(A_j)$ for any $i \leq j$, implying that $p_{ij}(\overline{A_j}) \subseteq \overline{A_i}$ (since p_{ij} are continuous). Hence, $\overline{A} \subseteq \varprojlim \overline{A_j}$. To prove the converse inclusion, fix $\mathbf{x} \in (\overline{A})^c$. Since $(\overline{A})^c$ is an open subset of $\mathcal{X}$, there exists some $j \in J$ and an open set $U_j \subseteq \mathcal{Y}_j$ such that $\mathbf{x} \in p_j^{-1}(U_j) \subseteq (\overline{A})^c$. Consequently, for this value of j, $p_j(\mathbf{x}) \in U_j \subseteq A_j^c$, implying that $p_j(\mathbf{x}) \notin \overline{A_j}$. Hence,

$$\overline{A} = \varprojlim \overline{A_j} . \tag{4.6.4}$$

Combining this identity with (4.6.3), it follows that for every $\alpha < \infty$,

$$\overline{A} \cap \Psi_I(\alpha) = \varprojlim \left(\overline{A_j} \cap \Psi_{I_j}(\alpha) \right) .$$

Fix $\alpha < \inf_{x\in\overline{A}} I(x)$, for which $\overline{A} \cap \Psi_I(\alpha) = \emptyset$. Then, by Theorem B.4, $\overline{A_j} \cap \Psi_{I_j}(\alpha) = \emptyset$ for some $j \in J$. Therefore, as $A \subseteq p_j^{-1}(\,\overline{A_j}\,)$, by the LDP upper bound associated with the Borel measures $\{\mu_\epsilon \circ p_j^{-1}\}$,

$$\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A) \leq \limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon \circ p_j^{-1}(\overline{A_j}) \leq -\alpha \,.$$

This inequality holds for every measurable A and $\alpha < \infty$ such that $\overline{A} \cap \Psi_I(\alpha) = \emptyset$. Consequently, it yields the LDP upper bound for $\{\mu_\epsilon\}$. $\square$

The following lemma is often useful for simplifying the formula (4.6.2) of the Dawson–Gärtner rate function.

Lemma 4.6.5 *If $I(\cdot)$ is a good rate function on $\mathcal{X}$ such that*

$$I_j(y) = \inf\{I(\mathbf{x}) :\ \mathbf{x} \in \mathcal{X}, \quad y = p_j(\mathbf{x})\} \,, \tag{4.6.6}$$

for any $y \in \mathcal{Y}_j$, $j \in J$, then the identity (4.6.2) holds.

Proof: Fix $\alpha \in [0,\infty)$ and let A denote the compact level set $\Psi_I(\alpha)$. Since $p_j : \mathcal{X} \to \mathcal{Y}_j$ is continuous for any $j \in J$, by (4.6.6) $A_j \triangleq \Psi_{I_j}(\alpha) = p_j(A)$ is a compact subset of $\mathcal{Y}_j$. With $A_i = p_{ij}(A_j)$ for any $i \leq j$, the set $\{\mathbf{x} : \sup_{j\in J} I_j(p_j(\mathbf{x})) \leq \alpha\}$ is the projective limit of the closed sets A_j, and as such it is merely the closed set $A = \Psi_I(\alpha)$ (see (4.6.4)). The identity (4.6.2) follows since $\alpha \in [0,\infty)$ is arbitrary. $\square$

The preceding theorem is particularly suitable for situations involving topological vector spaces that satisfy the following assumptions.

Assumption 4.6.7 *Let $\mathcal{W}$ be an infinite dimensional real vector space, and $\mathcal{W}'$ its algebraic dual, i.e., the space of all linear functionals $\lambda \mapsto \langle \lambda, x\rangle : \mathcal{W} \to \mathbb{R}$. The topological (vector) space $\mathcal{X}$ consists of $\mathcal{W}'$ equipped with the $\mathcal{W}$-topology, i.e., the weakest topology such that for each $\lambda \in \mathcal{W}$, the linear functional $x \mapsto \langle \lambda, x\rangle : \mathcal{X} \to \mathbb{R}$ is continuous.*

Remark: The $\mathcal{W}$-topology of $\mathcal{W}'$ makes $\mathcal{W}$ into the topological dual of $\mathcal{X}$, i.e., $\mathcal{W} = \mathcal{X}^*$.

For any $d \in \mathbb{Z}_+$ and $\lambda_1, \ldots, \lambda_d \in \mathcal{W}$, define the projection $p_{\lambda_1,\ldots,\lambda_d} : \mathcal{X} \to \mathbb{R}^d$ by $p_{\lambda_1,\ldots,\lambda_d}(x) = (\langle \lambda_1, x\rangle, \langle \lambda_2, x\rangle, \ldots, \langle \lambda_d, x\rangle)$.

Assumption 4.6.8 *Let $(\mathcal{X}, \mathcal{B}, \mu_\epsilon)$ be probability spaces such that:*
(a) $\mathcal{X}$ satisfies Assumption 4.6.7.
(b) For any $\lambda \in \mathcal{W}$ and any Borel set B in $\mathbb{R}$, $p_\lambda^{-1}(B) \in \mathcal{B}$.

Remark: Note that if $\{\mu_\epsilon\}$ are Borel measures, then Assumption 4.6.8 reduces to Assumption 4.6.7.

Theorem 4.6.9 *Let Assumption 4.6.8 hold. Further assume that for every $d \in \mathbb{Z}_+$ and every $\lambda_1, \ldots, \lambda_d \in \mathcal{W}$, the measures $\{\mu_\epsilon \circ p^{-1}_{\lambda_1,\ldots,\lambda_d}, \epsilon > 0\}$ satisfy the LDP with the good rate function $I_{\lambda_1,\ldots,\lambda_d}(\cdot)$. Then $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{X}$, with the good rate function*

$$I(x) = \sup_{d \in \mathbb{Z}_+} \sup_{\lambda_1,\ldots,\lambda_d \in \mathcal{W}} I_{\lambda_1,\ldots,\lambda_d}((\langle \lambda_1, x\rangle, \langle \lambda_2, x\rangle, \ldots, \langle \lambda_d, x\rangle)) . \quad (4.6.10)$$

Remark: In most applications, one is interested in obtaining an LDP on $\mathcal{E}$ that is a non-closed subset of $\mathcal{X}$. Hence, the relatively effortless projective limit approach is then followed by an application specific check that $\mathcal{D}_I \subset \mathcal{E}$, as needed for Lemma 4.1.5. For example, in the study of empirical measures on a Polish space Σ, it is known *a priori* that $\mu_\epsilon(M_1(\Sigma)) = 1$ for all $\epsilon > 0$, where $M_1(\Sigma)$ is the space of Borel probability measures on Σ, equipped with the $B(\Sigma)$-topology, and $B(\Sigma) = \{f : \Sigma \to \mathbb{R}, f$ bounded, Borel measurable$\}$. Identifying each $\nu \in M_1(\Sigma)$ with the linear functional $f \mapsto \int_\Sigma f d\nu$, $\forall f \in B(\Sigma)$, it follows that $M_1(\Sigma)$ is homeomorphic to $\mathcal{E} \subset \mathcal{X}$, where here $\mathcal{X}$ denotes the algebraic dual of $B(\Sigma)$ equipped with the $B(\Sigma)$-topology. Thus, $\mathcal{X}$ satisfies Assumption 4.6.7, and $\mathcal{E}$ is not a closed subset of $\mathcal{X}$. It is worthwhile to note that in this setup, μ_ϵ is not necessarily a Borel probability measure.

Proof: Let $\mathcal{V}$ be the system of all finite dimensional linear subspaces of $\mathcal{W}$, equipped with the partial ordering defined by inclusion. To each $V \in \mathcal{V}$, attach its (finite dimensional) algebraic dual V' equipped with the V-topology. The latter are clearly Hausdorff topological spaces. For any $V \subseteq U$ and any linear functional $f : U \to \mathbb{R}$, let $p_{V,U}(f) : V \to \mathbb{R}$ be the restriction of f on the subspace V. The projections $p_{V,U} : U' \to V'$ thus defined are continuous, and compatible with the inclusion ordering of $\mathcal{V}$. Let $\tilde{\mathcal{X}}$ be the projective limit of the system $(V', p_{V,U})$. Consider the map $x \mapsto \tilde{x} = (p_V(x)) \in \tilde{\mathcal{X}}$, where for each $V \in \mathcal{V}$, $p_V(x) \in V'$ is the linear functional $\lambda \mapsto \langle \lambda, x\rangle$, $\forall \lambda \in V$. This map is a bijection between $\mathcal{W}'$ and $\tilde{\mathcal{X}}$, since the consistency conditions in the definition of $\tilde{\mathcal{X}}$ imply that any $\tilde{x} \in \tilde{\mathcal{X}}$ is determined by its values on the *one-dimensional* linear subspaces of $\mathcal{W}$, and any such collection of values determines a point in $\tilde{\mathcal{X}}$. By Assumption 4.6.7, $\mathcal{X}$ consists of the vector space $\mathcal{W}'$ equipped with the $\mathcal{W}$-topology that is generated by the sets $\{x : |\langle \lambda, x\rangle - \rho| < \delta\}$ for $\lambda \in \mathcal{W}, \rho \in \mathbb{R}, \delta > 0$. It is not hard to check that the image of these sets under the map $x \mapsto \tilde{x}$ generates the projective topology of $\tilde{\mathcal{X}}$. Consequently, this map is a homeomorphism between $\mathcal{X}$ and $\tilde{\mathcal{X}}$. Hence, if for every $V \in \mathcal{V}$, $\{\mu_\epsilon \circ p_V^{-1}, \epsilon > 0\}$ satisfies the LDP in V' with the good rate function $I_V(\cdot)$, then by Theorem 4.6.1, $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{X}$ with the good rate function $\sup_{V \in \mathcal{V}} I_V(p_V(\cdot))$.

Fix $d \in \mathbb{Z}_+$ and $V \in \mathcal{V}$, a d-dimensional linear subspace of $\mathcal{W}$. Let $\lambda_1, \ldots, \lambda_d$ be any algebraic base of V. Observe that the map $f \mapsto (f(\lambda_1), \ldots, f(\lambda_d))$ is a homeomorphism between V' and $\mathbb{R}^d$ under which the image of $p_V(x) \in V'$ is $p_{\lambda_1,\ldots,\lambda_d}(x) = (\langle \lambda_1, x\rangle, \ldots, \langle \lambda_d, x\rangle) \in \mathbb{R}^d$. Consequently, by our assumptions, the family of Borel probability measures $\{\mu_\epsilon \circ p_V^{-1}, \epsilon > 0\}$ satisfies the LDP in V', and moreover, $I_V(p_V(x)) = I_{\lambda_1,\ldots,\lambda_d}((\langle \lambda_1, x\rangle, \ldots, \langle \lambda_d, x\rangle))$. The proof is complete, as the preceding holds for every $V \in \mathcal{V}$, while because of the contraction principle (Theorem 4.2.1), there is no need to consider only linearly independent $\lambda_1, \ldots, \lambda_d$ in (4.6.10). $\square$

When using Theorem 4.6.9, either the convexity of $I_{\lambda_1,\ldots,\lambda_d}(\cdot)$ or the existence and smoothness of the limiting logarithmic moment generating function $\Lambda(\cdot)$ are relied upon in order to identify the good rate function of (4.6.10) with $\Lambda^*(\cdot)$, in a manner similar to that encountered in Section 4.5.2. This is spelled out in the following corollary.

Corollary 4.6.11 *Let Assumption 4.6.8 hold.*
(a) Suppose that for each $\lambda \in \mathcal{W}$, the limit

$$\Lambda(\lambda) = \lim_{\epsilon \to 0} \epsilon \log \int_{\mathcal{X}} e^{\epsilon^{-1}\langle \lambda, x\rangle} \mu_\epsilon(dx) \tag{4.6.12}$$

exists as an extended real number, and moreover that for any $d \in \mathbb{Z}_+$ and any $\lambda_1, \ldots, \lambda_d \in \mathcal{W}$, the function

$$g((t_1, \ldots, t_d)) \triangleq \Lambda(\sum_{i=1}^d t_i \lambda_i) : \mathbb{R}^d \to (-\infty, \infty]$$

is essentially smooth, lower semicontinuous, and finite in some neighborhood of 0.
Then $\{\mu_\epsilon\}$ satisfies the LDP in $(\mathcal{X}, \mathcal{B})$ with the convex, good rate function

$$\Lambda^*(x) = \sup_{\lambda \in \mathcal{W}} \{\langle \lambda, x\rangle - \Lambda(\lambda)\}. \tag{4.6.13}$$

(b) Alternatively, if for any $\lambda_1, \ldots, \lambda_d \in \mathcal{W}$, there exists a compact set $K \subset \mathbb{R}^d$ such that $\mu_\epsilon \circ p^{-1}_{\lambda_1,\ldots,\lambda_d}(K) = 1$, and moreover $\{\mu_\epsilon \circ p^{-1}_{\lambda_1,\ldots,\lambda_d}, \epsilon > 0\}$ satisfies the LDP with a convex rate function, then $\Lambda : \mathcal{W} \to \mathbb{R}$ exists, is finite everywhere, and $\{\mu_\epsilon\}$ satisfies the LDP in $(\mathcal{X}, \mathcal{B})$ with the convex, good rate function $\Lambda^*(\cdot)$ as defined in (4.6.13).

Remark: Since $\mathcal{X}$ satisfies Assumption 4.6.7, the only continuous linear functionals on $\mathcal{X}$ are of the form $x \mapsto \langle \lambda, x\rangle$, where $\lambda \in \mathcal{W}$. Consequently, $\mathcal{X}^*$ may be identified with $\mathcal{W}$, and $\Lambda^*(\cdot)$ is the Fenchel–Legendre transform of $\Lambda(\cdot)$ as defined in Section 4.5.

Proof: (a) Fix $d \in \mathbb{Z}_+$ and $\lambda_1, \ldots, \lambda_d \in \mathcal{W}$. Note that the limiting logarithmic moment generating function associated with $\{\mu_\epsilon \circ p^{-1}_{\lambda_1,\ldots,\lambda_d}, \epsilon > 0\}$ is $g((t_1, \ldots, t_d))$. Hence, by our assumptions, the Gärtner–Ellis theorem (Theorem 2.3.6) implies that these measures satisfy the LDP in $\mathbb{R}^d$ with the good rate function $I_{\lambda_1,\ldots,\lambda_d} = g^* : \mathbb{R}^d \to [0, \infty]$, where

$$I_{\lambda_1,\ldots,\lambda_d}((\langle \lambda_1, x\rangle, \langle \lambda_2, x\rangle, \ldots, \langle \lambda_d, x\rangle)) = \sup_{t_1,\ldots,t_d \in \mathbb{R}} \Big\{ \sum_{i=1}^d t_i \langle \lambda_i, x\rangle - \Lambda\Big(\sum_{i=1}^d t_i \lambda_i\Big)\Big\} \, .$$

Consequently, for every $x \in \mathcal{X}$,

$$I_{\lambda_1,\ldots,\lambda_d}((\langle \lambda_1, x\rangle, \langle \lambda_2, x\rangle, \ldots, \langle \lambda_d, x\rangle)) \le \Lambda^*(x) = \sup_{\lambda \in \mathcal{W}} I_\lambda(\langle \lambda, x\rangle) \, .$$

Since the preceding holds for every $\lambda_1, \ldots, \lambda_d \in \mathcal{W}$, the LDP of $\{\mu_\epsilon\}$ with the good rate function $\Lambda^*(\cdot)$ is a direct consequence of Theorem 4.6.9.
(b) Fix $d \in \mathbb{Z}_+$ and $\lambda_1, \ldots, \lambda_d \in \mathcal{W}$. Since $\mu_\epsilon \circ p^{-1}_{\lambda_1,\ldots,\lambda_d}$ are supported on a compact set K, they satisfy the boundedness condition (4.5.11). Hence, by our assumptions, Theorem 4.5.10 applies. It then follows that the limiting moment generating function $g(\cdot)$ associated with $\{\mu_\epsilon \circ p^{-1}_{\lambda_1,\ldots,\lambda_d}, \epsilon > 0\}$ exists, and the LDP for these probability measures is controlled by $g^*(\cdot)$. With $I_{\lambda_1,\ldots,\lambda_d} = g^*$ for any $\lambda_1, \ldots, \lambda_d \in \mathcal{W}$, the proof is completed as in part (a). □

The following corollary of the projective limit approach is a somewhat stronger version of Corollary 4.5.27.

Corollary 4.6.14 *Let $\{\mu_\epsilon\}$ be an exponentially tight family of Borel probability measures on the locally convex Hausdorff topological vector space $\mathcal{E}$. Suppose $\Lambda(\cdot) = \lim_{\epsilon\to 0} \epsilon \Lambda_{\mu_\epsilon}(\cdot/\epsilon)$ is finite valued and Gateaux differentiable. Then $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{E}$ with the convex, good rate function Λ^*.*

Proof: Let $\mathcal{W}$ be the topological dual of $\mathcal{E}$. Suppose first that $\mathcal{W}$ is an infinite dimensional vector space, and define $\mathcal{X}$ according to Assumption 4.6.7. Let $i : \mathcal{E} \to \mathcal{X}$ denote the map $x \mapsto i(x)$, where $i(x)$ is the linear functional $\lambda \mapsto \langle \lambda, x\rangle$, $\forall \lambda \in \mathcal{W}$. Since $\mathcal{E}$ is a locally convex topological vector space, by the Hahn–Banach theorem, $\mathcal{W}$ is separating. Therefore, $\mathcal{E}$ when equipped with the *weak* topology is Hausdorff, and i is a homeomorphism between this topological space and $i(\mathcal{E}) \subset \mathcal{X}$. Consequently, $\{\mu_\epsilon \circ i^{-1}\}$ are Borel probability measures on $\mathcal{X}$ such that $\mu_\epsilon \circ i^{-1}(i(\mathcal{E})) = 1$ for all $\epsilon > 0$. All the conditions in part (a) of Corollary 4.6.11 hold for $\{\mu_\epsilon \circ i^{-1}\}$, since we assumed that $\Lambda : \mathcal{W} \to \mathbb{R}$ exists, and is a finite valued, Gateaux differentiable function. Hence, $\{\mu_\epsilon \circ i^{-1}\}$ satisfies the LDP in $\mathcal{X}$ with the convex, good rate function $\Lambda^*(\cdot)$. Recall that $i : \mathcal{E} \to \mathcal{X}$ is a continuous injection with respect to the weak topology on $\mathcal{E}$, and hence it is also continuous with respect to the original topology on $\mathcal{E}$. Now, the exponential tightness of $\{\mu_\epsilon\}$, Theorem 4.2.4, and the remark following it, imply that $\{\mu_\epsilon\}$ satisfies the LDP in $\mathcal{E}$ with the good rate function $\Lambda^*(\cdot)$.

We now turn to settle the (trivial) case where $\mathcal{W}$ is a d-dimensional vector space for some $d < \infty$. Observe that then $\mathcal{X}$ is of the same dimension as $\mathcal{W}$. The finite dimensional topological vector space $\mathcal{X}$ can be represented as $\mathbb{R}^d$. Hence, our assumptions about the function $\Lambda(\cdot)$ imply the LDP in $\mathcal{X}$ associated with $\{\mu_\epsilon \circ i^{-1}\}$ by a direct application of the Gärtner–Ellis theorem (Theorem 2.3.6). The LDP (in $\mathcal{E}$) associated with $\{\mu_\epsilon\}$ follows exactly as in the infinite dimensional case. $\square$

Exercise 4.6.15 Suppose that all the conditions of Corollary 4.6.14 hold except for the exponential tightness of $\{\mu_\epsilon\}$. Prove that $\{\mu_\epsilon\}$ satisfies a weak LDP with respect to the *weak* topology on $\mathcal{E}$, with the rate function $\Lambda^*(\cdot)$ defined in (4.6.13).
Hint: Follow the proof of the corollary and observe that the LDP of $\{\mu_\epsilon \circ i^{-1}\}$ in $\mathcal{X}$ still holds. Note that if $K \subset \mathcal{E}$ is weakly compact, then $i(K) \subset i(\mathcal{E})$ is a compact subset of $\mathcal{X}$.

4.7 The LDP and Weak Convergence in Metric Spaces

Throughout this section $(\mathcal{X}, d)$ is a metric space and all probability measures are Borel. For $\delta > 0$, let

$$A^{\delta,o} \triangleq \{y : \ d(y,A) \triangleq \inf_{z\in A} d(y,z) < \delta\} \tag{4.7.1}$$

denote the open blowups of A (compare with (4.1.8)), with $A^{-\delta} = ((A^c)^{\delta,o})^c$ a closed set (possibly empty). The proof of the next lemma which summarizes immediate relations between these sets is left as Exercise 4.7.18.

Lemma 4.7.2 *For any $\delta > 0$, $\eta > 0$ and $\Gamma \subset \mathcal{X}$*
(a) $(\Gamma^{-\delta})^{\delta,o} \subset \Gamma \subset (\Gamma^{\delta,o})^{-\delta}$.
(b) $\Gamma^{-(\delta+\eta)} \subset (\Gamma^{-\delta})^{-\eta}$ *and* $(\Gamma^{\delta,o})^{\eta,o} \subset \Gamma^{(\delta+\eta),o}$.
(c) $G^{-\delta}$ increases to G for any open set G and $F^{\delta,o}$ decreases to F for any closed set F.

Let $\mathcal{Q}(\mathcal{X})$ denote the collection of set functions $\nu : \mathcal{B}_{\mathcal{X}} \to [0,1]$ such that:

(a) $\nu(\emptyset) = 0$.

(b) $\nu(\Gamma) = \inf\{\nu(G) : \ \Gamma \subset G \text{ open}\}$ for any $\Gamma \in \mathcal{B}_{\mathcal{X}}$.

(c) $\nu(\cup_{i=1}^{\infty}\Gamma_i) \leq \sum_{i=1}^{\infty} \nu(\Gamma_i)$ for any $\Gamma_i \in \mathcal{B}_{\mathcal{X}}$.

(d) $\nu(G) = \lim_{\delta\to 0} \nu(G^{-\delta})$ for any open set $G \subset \mathcal{X}$.

Condition (b) implies the monotonicity property $\nu(A) \leq \nu(B)$ whenever $A \subset B$.

The following important subset of $\mathcal{Q}(\mathcal{X})$ represents the rate functions.

Definition 4.7.3 *A set function $\nu : \mathcal{B}_{\mathcal{X}} \to [0,1]$ is called a sup-measure if $\nu(\Gamma) = \sup_{y\in\Gamma} \nu(\{y\})$ for any $\Gamma \in \mathcal{B}_{\mathcal{X}}$ and $\nu(\{y\})$ is an upper semicontinuous function of $y \in \mathcal{X}$. With a sup-measure ν uniquely characterized by the rate function $I(y) = -\log \nu(\{y\})$, we adopt the notation $\nu = e^{-I}$.*

The next lemma explains why $\mathcal{Q}(\mathcal{X})$ is useful for exploring similarities between the LDP and the well known theory of weak convergence of probability measures.

Lemma 4.7.4 *$\mathcal{Q}(\mathcal{X})$ contains all sup-measures and all set functions of the form μ^ϵ for μ a probability measure on $\mathcal{X}$ and $\epsilon \in (0,1]$.*

Proof: Conditions (a) and (c) trivially hold for any sup-measure. Since any point y in an open set G is also in $G^{-\delta}$ for some $\delta = \delta(y) > 0$, all sup-measures satisfy condition (d). For (b), let $\nu(\{y\}) = e^{-I(y)}$. Fix $\Gamma \in \mathcal{B}_{\mathcal{X}}$ and $G(x,\delta)$ as in (4.1.3), such that

$$e^{\delta}\nu(\{x\}) = e^{-(I(x)-\delta)} \geq e^{-\inf_{y\in G(x,\delta)} I(y)} = \sup_{y\in G(x,\delta)} \nu(\{y\}) \,.$$

It follows that for the open set $G_\delta = \cup_{x\in\Gamma} G(x,\delta)$,

$$e^{\delta}\nu(\Gamma) = e^{\delta} \sup_{x\in\Gamma} \nu(\{x\}) \geq \sup_{y\in G_\delta} \nu(\{y\}) = \nu(G_\delta) \,.$$

Taking $\delta \to 0$, we have condition (b) holding for an arbitrary $\Gamma \in \mathcal{B}_{\mathcal{X}}$.

Turning to the second part of the lemma, note that conditions (a)–(d) hold when ν is a probability measure. Suppose next that $\nu(\cdot) = f(\mu(\cdot))$ for a probability measure μ and $f \in C_b([0,1])$ non-decreasing such that $f(0) = 0$ and $f(p+q) \leq f(p) + f(q)$ for $0 \leq p \leq 1-q \leq 1$. By induction, $f(\sum_{i=1}^k p_i) \leq \sum_{i=1}^k f(p_i)$ for all $k \in \mathbb{Z}_+$ and non-negative p_i such that $\sum_{i=1}^k p_i \leq 1$. The continuity of $f(\cdot)$ at 0 extends this property to $k =$

∞. Therefore, condition (c) holds for ν by the subadditivity of μ and monotonicity of $f(\cdot)$. The set function ν inherits conditions (b) and (d) from μ by the continuity and monotonicity of $f(\cdot)$. Similarly, it inherits condition (a) because $f(0) = 0$. In particular, this applies to $f(p) = p^\epsilon$ for all $\epsilon \in (0, 1]$. $\square$

The next definition of convergence in $\mathcal{Q}(\mathcal{X})$ coincides by the Portmanteau theorem with weak convergence when restricted to probability measures ν_ϵ, ν_0 (see Theorem D.10 for $\mathcal{X}$ Polish).

Definition 4.7.5 *$\nu_\epsilon \to \nu_0$ in $\mathcal{Q}(\mathcal{X})$ if for any closed set $F \subset \mathcal{X}$*

$$\limsup_{\epsilon \to 0} \nu_\epsilon(F) \le \nu_0(F) \,, \tag{4.7.6}$$

and for any open set $G \subset \mathcal{X}$,

$$\liminf_{\epsilon \to 0} \nu_\epsilon(G) \ge \nu_0(G) \,. \tag{4.7.7}$$

For probability measures ν_ϵ, ν_0 the two conditions (4.7.6) and (4.7.7) are equivalent. However, this is not the case in general. For example, if $\nu_0(\cdot) \equiv 0$ (an element of $\mathcal{Q}(\mathcal{X})$), then (4.7.7) holds for any ν_ϵ but (4.7.6) fails unless $\nu_\epsilon(\mathcal{X}) \to 0$.

For a family of probability measures $\{\mu_\epsilon\}$, the convergence of $\nu_\epsilon = \mu_\epsilon^\epsilon$ to a sup-measure $\nu_0 = e^{-I}$ is exactly the LDP statement (compare (4.7.6) and (4.7.7) with (1.2.12) and (1.2.13), respectively).

With this in mind, we next extend the definition of tightness and uniform tightness from $M_1(\mathcal{X})$ to $\mathcal{Q}(\mathcal{X})$ in such a way that a sup-measure $\nu = e^{-I}$ is tight if and only if the corresponding rate function is good, and exponential tightness of $\{\mu_\epsilon\}$ is essentially the same as uniform tightness of the set functions $\{\mu_\epsilon^\epsilon\}$.

Definition 4.7.8 *A set function $\nu \in \mathcal{Q}(\mathcal{X})$ is tight if for each $\eta > 0$, there exists a compact set $K_\eta \subset \mathcal{X}$ such that $\nu(K_\eta^c) < \eta$. A collection $\{\nu_\epsilon\} \subset \mathcal{Q}(\mathcal{X})$ is uniformly tight if the set K_η may be chosen independently of ϵ.*

The following lemma provides a useful consequence of tightness in $\mathcal{Q}(\mathcal{X})$.

Lemma 4.7.9 *If $\nu \in \mathcal{Q}(\mathcal{X})$ is tight, then for any $\Gamma \in \mathcal{B}_\mathcal{X}$,*

$$\nu(\overline{\Gamma}) = \lim_{\delta \to 0} \nu(\Gamma^{\delta,o}) \,. \tag{4.7.10}$$

Remark: For sup-measures this is merely part (b) of Lemma 4.1.6.

Proof: Fix a non-empty set $\Gamma \in \mathcal{B}_{\mathcal{X}}$, $\eta > 0$ and a compact set $K = K_\eta$ for which $\nu(K_\eta^c) < \eta$. For any open set $G \subset \mathcal{X}$ such that $\overline{\Gamma} \subset G$, either $K \subset G$ or else the non-empty compact set $K \cap G^c$ and the closed set $\overline{\Gamma}$ are disjoint, with $\inf_{x \in K \cap G^c} d(x, \overline{\Gamma}) > 0$. In both cases, $\Gamma^{\delta,o} \cap K = \overline{\Gamma}^{\delta,o} \cap K \subset G$ for some $\delta > 0$, and by properties (b), (c) and monotonicity of set functions in $\mathcal{Q}(\mathcal{X})$,

$$\begin{aligned} \nu(\overline{\Gamma}) = \inf\{\nu(G) : \overline{\Gamma} \subset G \text{ open }\} &\geq \lim_{\delta \to 0} \nu(\Gamma^{\delta,o} \cap K) \\ &\geq \lim_{\delta \to 0} \nu(\Gamma^{\delta,o}) - \eta \geq \nu(\overline{\Gamma}) - \eta \,. \end{aligned}$$

The limit as $\eta \to 0$ yields (4.7.10). □

For $\tilde{\nu}, \nu \in \mathcal{Q}(\mathcal{X})$, let

$$\begin{aligned} \rho(\tilde{\nu}, \nu) \stackrel{\triangle}{=} \inf\{\delta > 0 : \ & \tilde{\nu}(F) \leq \nu(F^{\delta,o}) + \delta \ \ \forall\, F \subset \mathcal{X} \text{ closed}, \\ & \tilde{\nu}(G) \geq \nu(G^{-\delta}) - \delta \ \ \forall G \subset \mathcal{X} \text{ open}\,\} \end{aligned} \tag{4.7.11}$$

When $\rho(\cdot,\cdot)$ is restricted to $M_1(\mathcal{X}) \times M_1(\mathcal{X})$, it coincides with the Lévy metric (see Theorem D.8). Indeed, in this special case, if $\delta > 0$ is such that $\tilde{\nu}(F) \leq \nu(F^{\delta,o}) + \delta$ for a closed set $F \subset \mathcal{X}$, then $\tilde{\nu}(G) \geq \nu((F^{\delta,o})^c) - \delta = \nu(G^{-\delta}) - \delta$ for the open set $G = F^c$.

The next theorem shows that in analogy with the theory of weak convergence, $(\mathcal{Q}(\mathcal{X}), \rho)$ is a metric space for which convergence to a tight limit point is characterized by Definition 4.7.5.

Theorem 4.7.12
(a) $\rho(\cdot,\cdot)$ is a metric on $\mathcal{Q}(\mathcal{X})$.
(b) For ν_0 tight, $\rho(\nu_\epsilon, \nu_0) \to 0$ if and only if $\nu_\epsilon \to \nu_0$ in $\mathcal{Q}(\mathcal{X})$.

Remarks:
(a) By Theorem 4.7.12, the Borel probability measures $\{\mu_\epsilon\}$ satisfy the LDP in $(\mathcal{X}, d)$ with good rate function $I(\cdot)$ if and only if $\rho(\mu_\epsilon^\epsilon, e^{-I}) \to 0$.
(b) In general, one can not dispense of tightness of $\nu_0 = e^{-I}$ when relating the $\rho(\nu_\epsilon, \nu_0)$ convergence to the LDP. Indeed, with μ_1 a probability measure on $\mathbb{R}$ such that $d\mu_1/dx = C/(1+|x|^2)$ it is easy to check that $\mu_\epsilon(\cdot) \stackrel{\triangle}{=} \mu_1(\cdot/\epsilon)$ satisfies the LDP in $\mathbb{R}$ with rate function $I(\cdot) \equiv 0$ while considering the open sets $G_x = (x, \infty)$ for $x \to \infty$ we see that $\rho(\mu_\epsilon^\epsilon, e^{-I}) = 1$ for all $\epsilon > 0$.
(c) By part (a) of Lemma 4.7.2, $F \subset G^{-\delta}$ for the open set $G = F^{\delta,o}$ and $F^{\delta,o} \subset G$ for the closed set $F = G^{-\delta}$. Therefore, the monotonicity of the set functions $\tilde{\nu}, \nu \in \mathcal{Q}(\mathcal{X})$, results with

$$\begin{aligned} \rho(\tilde{\nu}, \nu) = \inf\{\delta > 0 : \quad & \tilde{\nu}(F) \leq \nu(F^{\delta,o}) + \delta \quad \text{and} \\ & \nu(F) \leq \tilde{\nu}(F^{\delta,o}) + \delta \quad \forall F \subset \mathcal{X} \quad \text{closed}\,\}. \end{aligned} \tag{4.7.13}$$

Proof: (a) The alternative definition (4.7.13) of ρ shows that it is a non-negative, symmetric function, such that $\rho(\nu,\nu)=0$ (by the monotonicity of set functions in $\mathcal{Q}(\mathcal{X})$). If $\rho(\tilde\nu,\nu)=0$, then by (4.7.11), for any open set $G\subset\mathcal{X}$,

$$\tilde\nu(G)\geq\limsup_{\delta\to 0}[\nu(G^{-\delta})-\delta]=\nu(G)$$

(see property (d) of set functions in $\mathcal{Q}(\mathcal{X})$). Since ρ is symmetric, by same reasoning also $\nu(G)\geq\tilde\nu(G)$, so that $\tilde\nu(G)=\nu(G)$ for every open set $G\subset\mathcal{X}$. Thus, by property (b) of set functions in $\mathcal{Q}(\mathcal{X})$ we conclude that $\tilde\nu=\nu$.

Fix $\tilde\nu,\nu,\omega\in\mathcal{Q}(\mathcal{X})$ and $\delta>\rho(\tilde\nu,\omega)$, $\eta>\rho(\omega,\nu)$. Then, by (4.7.11) and part (b) of Lemma 4.7.2, for any closed set $F\subset\mathcal{X}$,

$$\tilde\nu(F)\leq\omega(F^{\delta,o})+\delta\leq\nu((F^{\delta,o})^{\eta,o})+\delta+\eta\leq\nu(F^{(\delta+\eta),o})+\delta+\eta\,.$$

By symmetry of ρ we can reverse the roles of $\tilde\nu$ and ν, hence concluding by (4.7.13) that $\rho(\tilde\nu,\nu)\leq\delta+\eta$. Taking $\delta\to\rho(\tilde\nu,\omega)$ and $\eta\to\rho(\omega,\nu)$ we have the triangle inequality $\rho(\tilde\nu,\nu)\leq\rho(\tilde\nu,\omega)+\rho(\omega,\nu)$.

(b) Suppose $\rho(\nu_\epsilon,\nu_0)\to 0$ for tight $\nu_0\in\mathcal{Q}(\mathcal{X})$. By (4.7.11), for any open set $G\subset\mathcal{X}$,

$$\liminf_{\epsilon\to 0}\nu_\epsilon(G)\geq\lim_{\delta\to 0}(\nu_0(G^{-\delta})-\delta)=\nu_0(G)\,,$$

yielding the lower bound (4.7.7). Similarly, by (4.7.11) and Lemma 4.7.9, for any closed set $F\subset\mathcal{X}$

$$\limsup_{\epsilon\to 0}\nu_\epsilon(F)\leq\lim_{\delta\to 0}\nu_0(F^{\delta,o})=\nu_0(F)\,.$$

Thus, the upper bound (4.7.6) holds for any closed set $F\subset\mathcal{X}$ and so $\nu_\epsilon\to\nu_0$.

Suppose now that $\nu_\epsilon\to\nu_0$ for tight $\nu_0\in\mathcal{Q}(\mathcal{X})$. Fix $\eta>0$ and a compact set $K=K_\eta$ such that $\nu_0(K^c)<\eta$. Extract a finite cover of K by open balls of radius $\eta/2$, each centered in K. Let $\{\Gamma_i;i=0,\ldots,M\}$ be the *finite* collection of all unions of elements of this cover, with $\Gamma_0\supset K$ denoting the union of all the elements of the cover. Since $\nu_0(\Gamma_0^c)<\eta$, by (4.7.6) also $\nu_\epsilon(\Gamma_0^c)\leq\eta$ for some $\epsilon_0>0$ and all $\epsilon\leq\epsilon_0$. For any closed set $F\subset\mathcal{X}$ there exists an $i\in\{0,\ldots,M\}$ such that

$$(F\cap\Gamma_0)\subset\overline{\Gamma}_i\subset F^{2\eta,o} \tag{4.7.14}$$

(take for Γ_i the union of those elements of the cover that intersect $F\cap\Gamma_0$). Thus, for $\epsilon\leq\epsilon_0$, by monotonicity and subadditivity of ν_ϵ, ν_0 and by the choice of K,

$$\nu_\epsilon(F)\leq\nu_\epsilon(F\cap\Gamma_0)+\nu_\epsilon(\Gamma_0^c)\leq\max_{0\leq i\leq M}\{\nu_\epsilon(\overline{\Gamma}_i)-\nu_0(\overline{\Gamma}_i)\}+\nu_0(F^{2\eta,o})+\eta.$$

With ϵ_0, M, and $\{\Gamma_i\}$ independent of F, since $\nu_\epsilon \to \nu_0$, it thus follows that

$$\limsup_{\epsilon\to 0} \sup_{F \text{ closed}} (\nu_\epsilon(F) - \nu_0(F^{2\eta,o})) \le \eta. \tag{4.7.15}$$

For an open set $G \subset \mathcal{X}$, let $F = G^{-2\eta}$ and note that (4.7.14) still holds with Γ_i replacing $\overline{\Gamma}_i$. Hence, reverse the roles of ν_0 and ν_ϵ to get for all $\epsilon \le \epsilon_0$,

$$\nu_0(G^{-2\eta}) \le \max_{0\le i\le M}\{\nu_0(\Gamma_i) - \nu_\epsilon(\Gamma_i)\} + \nu_\epsilon((G^{-2\eta})^{2\eta,o}) + \eta. \tag{4.7.16}$$

Recall that $(G^{-2\eta})^{2\eta,o} \subset G$ by Lemma 4.7.2. Hence, by (4.7.7), (4.7.16), and monotonicity of ν_ϵ

$$\limsup_{\epsilon\to 0} \sup_{G \text{ open}} (\nu_0(G^{-2\eta}) - \nu_\epsilon(G)) \le \eta\ . \tag{4.7.17}$$

Combining (4.7.15) and (4.7.17), we see that $\rho(\nu_\epsilon, \nu_0) \le 2\eta$ for all ϵ small enough. Taking $\eta \to 0$, we conclude that $\rho(\nu_\epsilon, \nu_0) \to 0$. □

Exercise 4.7.18 Prove Lemma 4.7.2.

4.8 Historical Notes and References

A statement of the LDP in a general setup appears in various places, *c.f.* [Var66, FW84, St84, Var84]. As mentioned in the historical notes referring to Chapter 2, various forms of this principle in specific applications have appeared earlier. The motivation for Theorem 4.1.11 and Lemma 4.1.21 comes from the analysis of [Rue67] and [Lan73].

Exercise 4.1.10 is taken from [LyS87]. Its converse, Lemma 4.1.23, is proved in [Puk91]. In that paper and in its follow-up [Puk94a], Pukhalskii derives many other parallels between exponential convergence in the form of large deviations and weak convergence. Our exposition of Lemma 4.1.23 follows that of [deA97a]. Other useful criteria for exponential tightness exist; see, for example, Theorem 3.1 in [deA85a].

The contraction principle was used by Donsker and Varadhan [DV76] in their treatment of Markov chains empirical measures. Statements of approximate contraction principles play a predominant role in Azencott's study of the large deviations for sample paths of diffusion processes [Aze80]. A general approximate contraction principle appears also in [DeuS89b]. The concept of exponentially good approximation is closely related to the comparison principle of [BxJ88, BxJ96]. In particular, the latter motivates

Exercises 4.2.29 and 4.2.30. For the extension of most of the results of Section 4.2.2 to $\mathcal{Y}$ a completely regular topological space, see [EicS96]. Finally, the inverse contraction principle in the form of Theorem 4.2.4 and Corollary 4.2.6 is taken from [Io91a].

The original version of Varadhan's lemma appears in [Var66]. As mentioned in the text, this lemma is related to *Laplace's method* in an abstract setting. See [Mal82] for a simple application in $\mathbb{R}^1$. For more on this method and its refinements, see the historical notes of Chapters 5 and 6. The inverse to Varadhan's lemma stated here is a modification of [Bry90], which also proves a version of Theorem 4.4.10.

The form of the upper bound presented in Section 4.5.1 dates back (for the empirical mean of real valued i.i.d. random variables) to Cramér and Chernoff. The bound of Theorem 4.5.3 appears in [Gär77] under additional restrictions, which are removed by Stroock [St84] and de Acosta [deA85a]. A general procedure for extending the upper bound from compact sets to closed sets without an exponential tightness condition is described in [DeuS89b], Chapter 5.1. For another version geared towards weak topologies see [deA90]. Exercise 4.5.5 and the specific computation in Exercise 6.2.19 are motivated by the derivation in [ZK95].

Convex analysis played a prominent role in the derivation of the LDP. As seen in Chapter 2, convex analysis methods had already made their entrance in $\mathbb{R}^d$. They were systematically used by Lanford and Ruelle in their treatment of thermodynamical limits via sub-additivity, and later applied in the derivation of Sanov's theorem (*c.f.* the historical notes of Chapter 6). Indeed, the statements here build on [DeuS89b] with an eye to the weak LDP presented by Bahadur and Zabell [BaZ79]. The extension of the Gärtner–Ellis theorem to the general setup of Section 4.5.3 borrows mainly from [Bal88] (who proved implicitly Theorem 4.5.20) and [Io91b]. For other variants of Corollaries 4.5.27 and 4.6.14, see also [Kif90a, deA94c, OBS96].

The projective limits approach to large deviations was formalized by Dawson and Gärtner in [DaG87], and was used in the context of obtaining the LDP for the empirical process by Ellis [Ell88] and by Deuschel and Stroock [DeuS89b]. It is a powerful tool for proving large deviations statements, as demonstrated in Section 5.1 (when combined with the inverse contraction principle) and in Section 6.4. The identification Lemma 4.6.5 is taken from [deA97a], where certain variants and generalizations of Theorem 4.6.1 are also provided. See also [deA94c] for their applications.

Our exposition of Section 4.7 is taken from [Jia95] as is Exercise 4.1.32. In [OBV91, OBV95, OBr96], O'Brien and Vervaat provide a comprehensive abstract unified treatment of weak convergence and of large deviation theory, a small part of which inspired Lemma 4.1.24 and its consequences.

http://www.springer.com/978-3-642-03310-0

Large Deviations Techniques and Applications
Dembo, A.; Zeitouni, O.
2010, XVI, 396 p., Softcover
ISBN: 978-3-642-03310-0