Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Large deviations for empirical measures generated by Gibbs measures with singular energy functionals

Paul Dupuis[1], Vaios Laschos[2], Kavita Ramanan[1]

submitted: December 21, 2015

[1] Division of Applied Mathematics
Brown University
182 George Street
Providence, RI 02912
USA

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: Vaios.Laschos@wias-berlin.de

No. 2203
Berlin 2015

2010 *Mathematics Subject Classification.* Primary: 60F10, 60K35; Secondary: 60B20.

Key words and phrases. Large deviations principle, empirical measures, Gibbs measures, interacting particle systems, singular potential, rate function, weak topology, Wasserstein topology, relative entropy, Coulomb gases, random matrices.

Research supported in part by AFOSRFA9550-12-1-0399.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

We establish large deviation principles (LDPs) for empirical measures associated with a sequence of Gibbs distributions on n-particle configurations, each of which is defined in terms of an inverse temperature β_n and an energy functional that is the sum of a (possibly singular) interaction and confining potential. Under fairly general assumptions on the potentials, we establish LDPs both with speeds $\beta_n/n \to \infty$, in which case the rate function is expressed in terms of a functional involving the potentials, and with the speed $\beta_n = n$, when the rate function contains an additional entropic term. Such LDPs are motivated by questions arising in random matrix theory, sampling and simulated annealing. Our approach, which uses the weak convergence methods developed in [9], establishes large deviation principles with respect to stronger, Wasserstein-type topologies, thus resolving an open question in [5]. It also provides a common framework for the analysis of LDPs with all speeds, and includes cases not covered due to technical reasons in previous works such as [3, 5].

1 Introduction

1.1 Description of problem

We consider configurations of a finite number of $\mathbb{R}^d$-valued particles that are subject to an external force consisting of a confining potential $V : \mathbb{R}^d \to (-\infty, +\infty]$ that acts on each particle and a pairwise interaction potential $W : \mathbb{R}^d \times \mathbb{R}^d \to (-\infty, +\infty]$. For every $n \in \mathbb{N}$, we define a Hamiltonian or energy functional $H_n : \mathbb{R}^{dn} \to (-\infty, +\infty]$ corresponding to any configuration of n particles, which is given by

$$\begin{aligned} H_n(\mathbf{x}^n) &\equiv H_n(\mathbf{x}_1, ..., \mathbf{x}_n) \\ &\doteq \int_{\mathbb{R}^d} V(\mathbf{x})\, L_n(\mathbf{x}^n; d\mathbf{x}) + \frac{1}{2} \int_{\neq} W(\mathbf{x}, \mathbf{y})\, L_n(\mathbf{x}^n; d\mathbf{x})\, L_n(\mathbf{x}^n; d\mathbf{y}). \end{aligned} \tag{1}$$

In (1) the symbol $\neq$ indicates that the integral is over $\left\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d : \mathbf{x} \neq \mathbf{y}\right\}$, and $L_n(\mathbf{x}^n, \cdot)$ is the empirical measure associated with the n-particle configuration $\mathbf{x}^n = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$:

$$L_n(\mathbf{x}^n; \cdot) \doteq \frac{1}{n} \sum_{i=1}^{n} \delta_{\mathbf{x}_i}(\cdot), \tag{2}$$

where $\delta_{\mathbf{y}}$ denotes the Dirac delta mass at $\mathbf{y} \in \mathbb{R}^d$. Given a separable metric space S, let $\mathcal{B}(S)$ denote the collection of Borel subsets of S, and let $\mathcal{P}(S)$ denote the space of probability measures on $(S, \mathcal{B}(S))$. Note that for every $\mathbf{x}^n \in \mathbb{R}^d$, $L_n(\mathbf{x}^n, \cdot)$ lies in $\mathcal{P}(\mathbb{R}^d)$, where $\mathbb{R}^d$ is equipped with the usual Euclidean metric. Let $\{\beta_n\}$ be a sequence of positive numbers diverging to infinity, which can be interpreted as a sequence of inverse temperatures, and for each $n \in \mathbb{N}$, let $P_n \in \mathcal{P}(\mathbb{R}^{dn})$ be the probability measure given by

$$P_n(d\mathbf{x}_1, ..., d\mathbf{x}_n) \doteq \frac{\exp\left(-\beta_n H_n(\mathbf{x}_1, ..., \mathbf{x}_n)\right)}{Z_n} \ell(d\mathbf{x}_1) \cdots \ell(d\mathbf{x}_n), \tag{3}$$

where ℓ is a non-atomic, σ-finite measure on $\mathbb{R}^d$ that acts as a reference measure, and Z_n is the normalization constant (which is also referred to as the partition function) given by

$$Z_n \doteq \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \exp\left(-\beta_n H_n\left(\mathbf{x}_1, ..., \mathbf{x}_n\right)\right) \ell(d\mathbf{x}_1) \cdots \ell(d\mathbf{x}_n). \tag{4}$$

Also, let Q_n be the measure induced on $\mathcal{P}\left(\mathbb{R}^d\right)$ by P_n under the mapping $L_n : \mathbb{R}^{dn} \to \mathcal{P}(\mathbb{R}^d)$ defined in (2).

Measures of the form (3) arise in a variety of contexts. For the case when ℓ is Lebesgue measure on $\mathbb{R}^d$, it is well known that if W and V are smooth enough, then P_n is the invariant distribution of a reversible Markov diffusion on $\mathbb{R}^{dn}$ (with identity diffusion matrix and drift proportional to ∇H_n), which can be viewed as describing the dynamics of n interacting Brownian particles in $\mathbb{R}^d$ [10, Chapter 5]. On the other hand, for particular choices of d, V and W, P_n arises as the law of the spectrum of various random matrix ensembles, including the so-called β-ensemble as well as certain random normal matrices (see Section 1.5.7 of [5] for details).

The aim of this paper is to establish large deviation principles (LDPs) for sequences $\{Q_n\}$ under general conditions on V and W that allow V and W to be not only unbounded, but also highly irregular. We apply the weak convergence methods developed in [9] to provide results for both cases where $\beta_n = n$, and $\lim_{n\to\infty} \frac{\beta_n}{n} = \infty$. We establish these LDPs not only with respect to the weak topology, but also with respect to a family of stronger topologies that include the p-Wasserstein topologies for $p \geq 1$. Our results generalize those obtained in [5] and [3] and additionally, resolve an open question raised in [5, Section 1.5.6]. In contrast to prior works, the LDPs for all speeds and topologies are established using a common methodology. In Section 1.2, we recall basic definitions and notation and in Section 1.3 present the main results. In Section 1.4, we provide a detailed discussion of the assumptions we use, and their relation to those used in prior work on this problem. Section 1.5 contains the outline of the rest of the paper.

1.2 Notation and definitions

We first recall the definition of a rate function on a separable metric space S.

Definition 1. Given a topological space S, a function $\mathcal{H} : S \to [0, \infty]$ is said to be a *rate function* if it is lower semicontinuous (lsc) and each level set $\{x : \mathcal{H}(x) \leq M\}$, $M \in [0, \infty)$, is compact.

Note that a function that satisfies the properties in Definition 1 is sometimes referred to as a good rate function in the literature, as a way to highlight the second property and to distinguish it from functions that are only lower semicontinuous, but which can in some cases provide large deviation rates of decay. When not in the context of LDPs, a function that has the properties stated in Definition 1 is also called a *tightness function*; a term that will be used extensively in the sequel. In contrast to much of the previous application of weak convergence methods in large deviations, here we do not assume S is complete. This will be convenient when dealing with topologies other than the weak topology.

We now recall the definition of an LDP for a sequence of probability measures on $(S, \mathcal{B}(S))$.

Definition 2. *Let* $\{R_n\} \subset \mathcal{P}(S)$, let $\{\alpha_n\}$ be a sequence of positive real numbers such that $\lim_{n\to\infty} \alpha_n = \infty$, and let $\mathcal{H} : S \to [0, \infty]$ be a rate function. The sequence $\{R_n\}$ is said to satisfy a large deviation principle with speed $\{\alpha_n\}$ and rate function $\mathcal{H}$ if for each $E \in \mathcal{B}(S)$,

$$-\inf_{x\in E^\circ} \mathcal{H}(x) \leq \liminf_{n\to\infty} \alpha_n^{-1} \log(R_n(E)) \leq \limsup_{n\to\infty} \alpha_n^{-1} \log(R_n(E)) \leq -\inf_{x\in\bar{E}} \mathcal{H}(x),$$

where E° and $\bar{E}$ denote the interior and closure of E, respectively.

We endow $\mathcal{P}(\mathbb{R}^d)$ with the weak topology and use $\xrightarrow{w}$ to denote convergence with respect to this topology; recall that $\mu_n \xrightarrow{w} \mu$ if and only if

$$\forall f \in C_b(\mathbb{R}^d), \int_{\mathbb{R}^d} f(\mathbf{x})\mu_n(d\mathbf{x}) \to \int_{\mathbb{R}^d} f(\mathbf{x})\mu(d\mathbf{x}),$$

where $C_b(\mathbb{R}^d)$ is the space of bounded continuous functions on $\mathbb{R}^d$. The Lévy-Prohorov metric d_w metrizes the weak topology on $\mathcal{P}(\mathbb{R}^d)$, and the space $(\mathcal{P}(\mathbb{R}^d), d_w)$ is Polish (see [4, Page 72]). We also consider stronger topologies, parameterized by a positive, continuous function $\psi : \mathbb{R}^d \to \mathbb{R}_+$ that satisfies the growth condition

$$\lim_{c\to\infty} \inf_{\mathbf{x}:\|\mathbf{x}\|=c} \psi(\mathbf{x}) = \infty. \tag{5}$$

Given such a function ψ, let

$$\mathcal{P}_\psi(\mathbb{R}^d) \doteq \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \psi(\mathbf{x})\, \mu(d\mathbf{x}) < +\infty \right\}. \tag{6}$$

We endow $\mathcal{P}_\psi\left(\mathbb{R}^d\right)$ with the metric

$$d_\psi(\mu,\nu) \doteq d_w(\mu,\nu) + \left| \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu(d\mathbf{x}) - \int_{\mathbb{R}^d} \psi(\mathbf{x})\nu(d\mathbf{x}) \right|. \tag{7}$$

The space $\mathcal{P}_\psi(\mathbb{R}^d)$ is a separable metric space (see Lemma 28 for a proof).

Remark 1. *When $\psi(\mathbf{x}) = \|\mathbf{x}\|^p$, $\mathbf{x} \in \mathbb{R}^d$, for some $p \in [1,\infty)$, d_ψ induces the p-Wasserstein topology (see [1, Remark 7.1.11]). Another metric that is commonly used to induce the p-Wasserstein topology on $\mathcal{P}(\mathbb{R}^d)$ is*

$$d_p(\mu,\nu) = \inf_{\zeta\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} ||\mathbf{x}-\mathbf{y}||^p \zeta(d\mathbf{x}, d\mathbf{y}),$$

where $\Pi(\mu,\nu)$ is the set of all measures in $\mathbb{R}^{2d}$ with first marginal μ and second marginal ν. Although $\mathcal{P}_\psi(\mathbb{R}^d)$ endowed with d_p is complete and separable, we use the somewhat simpler metric d_ψ defined for any ψ satisfying (5), under which $\mathcal{P}_\psi(\mathbb{R}^d)$ is only separable, and not complete.

1.3 Assumptions and main results

Throughout, we make the following assumptions on the potentials V and W.

Assumption 3. *1 The functions $W : \mathbb{R}^d \times \mathbb{R}^d \to (-\infty,\infty]$ and $V : \mathbb{R}^d \to (-\infty,+\infty]$ are lsc on their respective domains. In addition, there exists a set $A \in \mathcal{B}(\mathbb{R}^d)$ with positive ℓ measure, such that*

$$\sup_{\mathbf{x}\in A} V(\mathbf{x}) < \infty \qquad \textit{and} \qquad \sup_{(\mathbf{x},\mathbf{y})\in A\times A} W(\mathbf{x},\mathbf{y}) < \infty. \tag{8}$$

2 V satisfies

$$\int_{\mathbb{R}^d} \exp\left(-V\left(\mathbf{x}\right)\right) \ell(d\mathbf{x}) = 1.$$

Remark 2. *Note that under Assumption 32, $e^{-V(\mathbf{x})}\ell(d\mathbf{x})$ is a probability measure on $\mathbb{R}^d$. By some abuse of notation, we will use $e^{-V}\ell$ to denote this measure.*

Our first result, which establishes an LDP for the sequence $\{Q_n\}$ with speed $\alpha_n = \beta_n = n$, requires the following additional assumptions. Given $\zeta \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ let

$$\mathfrak{W}(\zeta) \doteq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(\mathbf{x},\mathbf{y})\zeta(d\mathbf{x}d\mathbf{y}). \tag{9}$$

Assumption 4. *1 There exists $c \in \mathbb{R}$ such that*

$$\inf_{(\mathbf{x},\mathbf{y})\in\mathbb{R}^d\times\mathbb{R}^d} W(\mathbf{x},\mathbf{y}) > c. \tag{10}$$

2 There exists a lsc function $\phi : \mathbb{R}_+ \to \mathbb{R}$ with

$$\lim_{s\to+\infty}\frac{\phi(s)}{s} = +\infty,$$

such that for every $\mu \in \mathcal{P}(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d}\phi(\psi(\mathbf{x}))\,\mu(d\mathbf{x}) \le \inf_{\zeta\in\Pi(\mu,\mu)}\left\{\mathfrak{W}(\zeta) + \mathcal{R}\left(\zeta|e^{-V}\ell\otimes e^{-V}\ell\right)\right\}. \tag{11}$$

Assumption 3 and Assumption 41 guarantee that the Gibbs distribution given in (3) is well defined. More precisely, Assumption 31 ensures that the measure

$$\exp\left(-nH_n(\mathbf{x}_1,...,\mathbf{x}_n)\right)\ell(d\mathbf{x}_1)\cdots\ell(d\mathbf{x}_n)$$

is non-trivial, and Assumption 32 and Assumption 41 ensure its finiteness. Assumption 42 is used to establish the LDP with respect to the stronger topology induced by d_ψ.

The rate functions are expressed in terms of the following functionals. For $\mu \in \mathcal{P}(\mathbb{R}^d)$ let

$$\mathcal{W}(\mu) \doteq \mathfrak{W}(\mu\otimes\mu) = \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(\mathbf{x},\mathbf{y})\,\mu(d\mathbf{x})\,\mu(d\mathbf{y}). \tag{12}$$

Given a measure $\nu \in \mathcal{P}(\mathbb{R}^d)$, recall that the relative entropy function $\mathcal{R}(\cdot|\nu) : \mathcal{P}(\mathbb{R}^d) \to [0,\infty]$ is defined by

$$\mathcal{R}(\mu|\nu) \doteq \begin{cases}\displaystyle\int_{\mathbb{R}^d}\frac{d\mu}{d\nu}(\mathbf{x})\log\left(\frac{d\mu}{d\nu}(\mathbf{x})\right)\nu(d\mathbf{x}), & \text{if } \mu \ll \nu,\\ \infty, & \text{otherwise,}\end{cases}$$

where $\mu \ll \nu$ denotes that μ is absolutely continuous with respect to ν. Also, for $\mu \in \mathcal{P}(\mathbb{R}^d)$, define

$$\mathcal{I}(\mu) \doteq \mathcal{R}\left(\mu|e^{-V}\ell\right) + \mathcal{W}(\mu). \tag{13}$$

We now state our first main result, whose proof is given in Section 3.

Theorem 5. *Let V and W satisfy Assumption 3 and Assumption 41, and for $n \in \mathbb{N}$, let $\beta_n = n$, let P_n be defined as in* (3) *and let Q_n be the measure on $\mathcal{P}\left(\mathbb{R}^d\right)$ induced by P_n under the mapping L_n. Then $\{Q_n\}$ satisfies an LDP on $\mathcal{P}\left(\mathbb{R}^d\right)$ with speed $\alpha_n = \beta_n = n$ and rate function*

$$\mathcal{I}_\star(\mu) \doteq \mathcal{I}(\mu) - \inf_{\mu\in\mathcal{P}\left(\mathbb{R}^d\right)}\{\mathcal{I}(\mu)\}, \tag{14}$$

where $\mathcal{I}$ is defined by (13)*. On the other hand, given a positive continuous $\psi : \mathbb{R}^d \mapsto \mathbb{R}_+$ that satisfies* (5)*, suppose Q_n denotes the measure on $\mathcal{P}_\psi\left(\mathbb{R}^d\right)$ induced by P_n under the mapping L_n and Assumption 42 also holds. Then $\{Q_n\}$ satisfies an LDP on $\mathcal{P}_\psi(\mathbb{R}^d)$, equipped with the stronger topology induced by d_ψ, and with rate function*

$$\mathcal{I}_\star^\psi(\mu) \doteq \mathcal{I}(\mu) - \inf_{\mu \in \mathcal{P}_\psi\left(\mathbb{R}^d\right)} \{\mathcal{I}(\mu)\}. \tag{15}$$

When W is identically zero, Theorem 5 recovers the well known Sanov's theorem (see [6, Theorem 6.2.10] or [9, Theorem 2.2.1] for the LDP with respect to the weak topology and [16] for the LDP with respect to the p-Wasserstein topology). Moreover, if W is continuous and satisfies certain growth conditions on $\mathbb{R}^d \times \mathbb{R}^d$, then the result can be obtained from Sanov's theorem by a simple application of Varadhan's lemma (see [6, Theorem 4.3.1] or [9, Theorem 1.2.1]). To the best of our knowledge, there are no general results in the literature that cover the case when W is both unbounded and discontinuous, and therefore Theorem 5 is the first in that direction. Furthermore, Assumption 42, which can be viewed as a generalization of condition (1.3) in [16], provides a sufficient condition for the LDP to hold with respect to a rather large class of stronger topologies.

Motivated by questions arising in random matrix theory, sampling and simulated annealing, several authors [5, 2, 3, 11, 13, 14] have considered LDPs for $\{Q_n\}$ at specific speeds that are faster than n, such as $\beta_n/n\log n \to \infty$ and $\beta_n = n^2$. Our second theorem presents a general result for speeds faster than n, that is, when $\beta_n/n \to \infty$, under Assumption 3 and certain modified assumptions on V and W stated in Assumption 6 below. In what follows, consider the functional $\mathcal{J} : \mathcal{P}(\mathbb{R}^d) \to (-\infty, \infty]$, given by

$$\mathcal{J}(\mu) \doteq \frac{1}{2}\int_{\mathbb{R}^d \times \mathbb{R}^d} (V(\mathbf{x}) + V(\mathbf{y}) + W(\mathbf{x},\mathbf{y}))\,\mu(d\mathbf{x})\,\mu(d\mathbf{y}). \tag{16}$$

Assumption 6. *1 There exist $1 > \epsilon_1 > 0$, and $c, c' \in \mathbb{R}$, such that*

$$\inf_{\mathbf{x}\in\mathbb{R}^d} V(\mathbf{x}) > c', \qquad \inf_{(\mathbf{x},\mathbf{y})\in\mathbb{R}^d\times\mathbb{R}^d} [W(\mathbf{x},\mathbf{y}) + \epsilon_1 (V(\mathbf{x}) + V(\mathbf{y}))] > c. \tag{17}$$

2 There exists a lsc function $\gamma : \mathbb{R}_+ \to \mathbb{R}$ with $\lim_{s\to+\infty}\gamma(s) = +\infty$ such that

$$V(\mathbf{x}) + V(\mathbf{y}) + W(\mathbf{x},\mathbf{y}) \geq \gamma(\|\mathbf{x}\|) + \gamma(\|\mathbf{y}\|). \tag{18}$$

3 For each $\mu \in \mathcal{P}\left(\mathbb{R}^d\right)$ such that $\mathcal{J}(\mu) < +\infty$, there is a sequence μ_n of probability measures, absolutely continuous with respect to the measure ℓ, such that μ_n converges weakly to μ and $\mathcal{J}(\mu_n) \to \mathcal{J}(\mu)$ as $n \to \infty$.

4 There exists a lsc function $\phi : \mathbb{R}_+ \to \mathbb{R}$ with

$$\lim_{s\to+\infty} \frac{\phi(s)}{s} = +\infty,$$

such that

$$V(\mathbf{x}) + V(\mathbf{y}) + W(\mathbf{x},\mathbf{y}) \geq \phi(\psi(\mathbf{x})) + \phi(\psi(\mathbf{y})). \tag{19}$$

Similar to the case of Theorem 5, Assumption 3 and Assumption 61 guarantee that the Gibbs distribution given in (3) is well defined. More precisely, Assumption 31 ensures that the measure

$$\exp(-\beta_n H_n(\mathbf{x}_1, ..., \mathbf{x}_n))\,\ell(d\mathbf{x}_1)\cdots\ell(d\mathbf{x}_n)$$

is non-trivial, and Assumption 32 and Assumption 61, together with the fact that $\lim_{n\to\infty} \frac{\beta_n}{n} = \infty$, ensure its finiteness. Assumption 63 is used in Section 4.4 to establish the Laplace principle upper bound. For a class of pairs V, W, where Assumption 63 is satisfied, the reader is directed to [5, Proposition 2.8]. We now state our second main result, whose proof is deferred to Section 4.

Theorem 7. *Let V and W satisfy Assumption 3 and Assumptions 61-63, and consider a sequence $\{\beta_n\}$ such that $\lim_{n\to\infty} \frac{\beta_n}{n} = \infty$. For $n \in \mathbb{N}$, let P_n be as in* (3) *and Q_n be the measure on $\mathcal{P}\left(\mathbb{R}^d\right)$ induced by P_n under the mapping L_n. Then $\{Q_n\}$ satisfies an LDP on $\mathcal{P}(\mathbb{R}^d)$ with speed $\alpha_n = \beta_n$ and rate function*

$$\mathcal{J}_\star(\mu) = \mathcal{J}(\mu) - \inf_{\mu\in\mathcal{P}(\mathbb{R}^d)}\{\mathcal{J}(\mu)\}, \tag{20}$$

where $\mathcal{J}$ is given in (16)*. Furthermore, given a positive continuous function $\psi : \mathbb{R}^d \mapsto \mathbb{R}_+$ that satisfies* (5)*, if Q_n denotes the measure on $\mathcal{P}_\psi(\mathbb{R}^d)$ induced by P_n under the mapping L_n, and we further assume that Assumption 64 holds, then $\{Q_n\}$ satisfies an LDP on $\mathcal{P}_\psi(\mathbb{R}^d)$, with the stronger topology d_ψ, and with the rate function*

$$\mathcal{J}_\star^\psi(\mu) \doteq \mathcal{J}(\mu) - \inf_{\mu\in\mathcal{P}_\psi\left(\mathbb{R}^d\right)}\{\mathcal{J}(\mu)\}. \tag{21}$$

A direct consequence of Theorem 5 and 7 is the following.

Remark 3. *Suppose V and W satisfy Assumption 3, Assumption 41 and Assumption 42 with $\psi(\mathbf{x}) \doteq ||\mathbf{x}||^p$ for some $p \geq 1$. Let $(X_1^n, \ldots, X_n^n)$ be distributed according to P^n and for any $q \leq p$, let $Y_q^n \doteq \frac{1}{n}\sum_{i=1}^n |X_i^n|^q$, $n \in \mathbb{N}$. Then Theorem 5, the continuity of the map $\mu \mapsto \int ||\mathbf{x}||^q \mu(d\mathbf{x})$ in the Wasserstein-p topology and the contraction principle [6, Theorem 4.2.1] together show that $\{Y_q^n\}$ satisfies an LDP with speed $\beta_n = n$ and rate function*

$$H(y) \doteq \inf_{\mu\in\mathcal{P}(\mathbb{R}^d)} \left\{ \mathcal{I}_\star^\psi(\mu) : y = \int_{\mathbb{R}^d} ||\mathbf{x}||^q \mu(dx) \right\}.$$

Likewise, if V and W satisfy Assumption 3, Assumption 61 and Assumption 3 with $\psi(\mathbf{x}) \doteq ||\mathbf{x}||^p$ and $\beta_n/n \to \infty$ as $n \to \infty$, then Theorem 7 shows that $\{Y_q^n\}$ satisfies an LDP with speed β_n and rate function $\tilde{H}(y) = \inf_{\mu\in\mathcal{P}(\mathbb{R}^d)}\{\mathcal{J}_\star^\psi(\mu) : y = \int_{\mathbb{R}^d} ||\mathbf{x}||^q \mu(dx)\}$.

To the best of our knowledge, the most general result in the direction of Theorem 7 is [5, Theorem 1.1]. The latter seems to be the first paper to present a general approach to proving LDPs for empirical measures generated by Gibbs distributions, when the inverse temperatures β_n diverge faster than n, the number of particles (the particular case of $\beta_n = n^2$ was considered earlier in [3]). Our result extends [5, Theorem 1.1] in several ways. First, whereas the paper [5] considers only speeds β_n that satisfy $\lim_{n\to\infty} \frac{\beta_n}{n\log(n)} = \infty$, we allow for any speed diverging faster than n, thus showing that the growth rate condition of [5] is a technical one related to the combinatorial approach used in the proofs therein. Our proof of Theorem 7 also reveals why relative entropy does not appear as a part of the rate function whenever $\lim_{n\to\infty} \frac{\beta_n}{n} = \infty$. Second, for both Theorem 5 and Theorem 7, our results cover cases when the interaction potential is not only unbounded but also discontinuous, which includes several interesting examples, some of which are illustrated in the next section. In contrast, the following assumptions were imposed in Assumptions H1–H3 of [5], which are restated below as Assumption H:

Assumption 8. *$V : \mathbb{R}^d \to (-\infty, \infty)$ and $W : \mathbb{R}^d \times \mathbb{R}^d \to (-\infty, \infty]$ are continuous functions on their respective domains, V satisfies $\lim_{||\mathbf{x}||\to\infty} V(\mathbf{x}) = \infty$ and $\int_{\mathbb{R}^d} e^{-V(\mathbf{x})} d\mathbf{x} < \infty$, and W is*

symmetric, finite on $\mathbb{R}^d \times \mathbb{R}^d \setminus \{(x,x) : x \in \mathbb{R}^d\}$ *and satisfies the following integrability condition: for each compact subset* $K \subset \mathbb{R}^d$*, the function*

$$\mathbf{z} \in \mathbb{R}^d \to \sup\{W(\mathbf{x},\mathbf{y}) : |\mathbf{x}-\mathbf{y}| > |\mathbf{z}|, \mathbf{x},\mathbf{y} \in K\}$$

is locally Lebesgue-integrable on $\mathbb{R}^d$*. Moreover,* V *and* W *satisfy the second inequality in Assumption 61.*

Furthermore, in both cases we cover stronger topologies than the weak topology, including in particular the Wasserstein-p topologies. Finally, we allow a fairly general reference measure, which allows us to consider Gibbs distributions that are defined on sets of Lebesgue measure zero (surfaces, submanifolds, fractal sets).

1.4 Discussion of assumptions and examples

1.4.1 Equivalent Formulations of the Assumptions

It follows from (1) that the representation of H_n in terms of W and V is not unique. Given functions $W, \tilde{W} : \mathbb{R}^d \times \mathbb{R}^d \mapsto (-\infty, \infty]$ and $V, \tilde{V} : \mathbb{R}^d \mapsto (-\infty, \infty]$, we call the pairs (W, V) and $(\tilde{W}, \tilde{V})$ equivalent if the right-hand side of (1) remains unchanged when W and V are replaced by $\tilde{W}$ and $\tilde{V}$, respectively. A first benefit of this observation is that in many cases we can work with alternative, equivalent assumptions that are easier to verify. For example, although the form of the conditions given in Assumptions 32 and 41 is convenient for the proof of Theorem 5, to verify the assumptions, it is often easier to work with the following equivalent set of conditions:

Assumption 9. *For lsc functions* $\tilde{V} : \mathbb{R}^d \to (-\infty, \infty], \tilde{W} : \mathbb{R}^d \times \mathbb{R}^d \to (-\infty, \infty]$*, we have*

1 *there exist lsc functions* $\tilde{V}_1, \tilde{V}_2 : \mathbb{R}^d \to (-\infty, \infty]$*, such that* $\tilde{V} = \tilde{V}_1 + \tilde{V}_2$*, and*

$$\int_{\mathbb{R}^d} \exp\left(-\tilde{V}_2(\mathbf{x})\right) \ell(d\mathbf{x}) < \infty,$$

2 *there exists* $\tilde{c} \in \mathbb{R}$ *such that the function* $\tilde{V}_1$ *in Assumption 91 satisfies*

$$\inf_{(\mathbf{x},\mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d} \left[\tilde{W}(\mathbf{x},\mathbf{y}) + \tilde{V}_1(\mathbf{x}) + \tilde{V}_1(\mathbf{y})\right] > \tilde{c}. \tag{22}$$

Note that the modified conditions, Assumptions 91 and 92, are more akin to Assumptions 32 and 61, in the sense that if V satisfies Assumptions 32 and 61 and there exists $\epsilon > 0$ such that $e^{-(1-\epsilon)V}$ is integrable with respect to the measure ℓ then Assumption 9 is satisfied $\tilde{V}_1 = \epsilon V$ and $\tilde{V}_2 = (1-\epsilon)V$.

Lemma 10. *The pair* (V, W) *satisfies Assumption 32 and Assumption 41 if and only if there exists a pair* $(\tilde{V}, \tilde{W})$ *that is equivalent to* (V, W) *and satisfies Assumption 9.*

Proof. Suppose (V, W) satisfies Assumption 32 and Assumption 41. Then it is clear that $\tilde{V} \doteq V$ satisfies Assumption 91 with $\tilde{V}_1 = 0$ and $\tilde{V}_2 = V$ and, setting $\tilde{W} = W$, $(\tilde{V}, \tilde{W})$ satisfies Assumption 91 and Assumption 92. To prove the converse, suppose that $(\tilde{V}, \tilde{W})$ satisfies Assumptions 91 and 92, for some $\tilde{V}_1$ and $\tilde{V}_2$. It is straightforward to verify that then $V(\mathbf{x}) \doteq \tilde{V}_2(\mathbf{x}) + \log \int_{\mathbb{R}^d} \exp(-\tilde{V}_2(\mathbf{x}))\ell(d\mathbf{x})$ satisfies Assumption 32, $W(\mathbf{x},\mathbf{y}) \doteq \tilde{W}(\mathbf{x},\mathbf{y}) + \tilde{V}_1(\mathbf{x}) + \tilde{V}_1(\mathbf{y})) - \log[\int_{\mathbb{R}^d} \exp(-\tilde{V}_2(\mathbf{x}))\ell(d\mathbf{x})]$ satisfies Assumption 41 with $c = \tilde{c} - \log[\int_{\mathbb{R}^d} \exp(-\tilde{V}_2(\mathbf{x}))\ell(d\mathbf{x})]$, and that (V, W) is equivalent to $(\tilde{V}, \tilde{W})$. □

The observation that the representation of H_n in terms of W and V is not unique, is also useful when Assumption 62 is considered. Given that Assumptions 3 and 61 hold, Assumption 62 is seemingly weaker than $\lim_{\|\mathbf{x}\|\to\infty} V(\mathbf{x}) = +\infty$, posed in [5] as part of Assumption 8. This can be directly seen if one sets $\gamma(t) = \frac{(1-\epsilon_1)}{2} \inf_{\|\mathbf{x}\|=t} V(\mathbf{x}) + C'$, where C' is chosen accordingly. However it is also straightforward to see that if Assumptions 3, 61, and 62 hold, then we can pick a different pair $(\tilde{V}, \tilde{W})$, equivalent to (V, W) such that Assumptions 3 and 61 are satisfied, and also

$$\lim_{\|\mathbf{x}\|\to\infty} \tilde{V}(\mathbf{x}) = +\infty.$$

We also consider the relationship of Assumption 42 to the assumption

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} e^{\lambda\psi(\mathbf{x})-V(\mathbf{x})}\ell(d\mathbf{x}) < +\infty \quad \forall\lambda\in\mathbb{R} \tag{23}$$

appearing in [16], where the case of $W \equiv 0$, and speed $\beta_n = n$ is studied. In [16], the authors prove that when (23) is true, the LDP holds in $\mathcal{P}_\psi(\mathbb{R}^d)$ with the rate function $\mathcal{R}(\mu|e^{-V}\ell)$. As an intermediate step they prove that (23) is true if and only if there exists a lsc, superlinear function $\phi : [0,\infty) \to \mathbb{R}$, such that for every $\mu \in \mathcal{P}(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} \phi(\psi(\mathbf{x}))\,\mu(d\mathbf{x}) \le \mathcal{R}\left(\mu|e^{-V}\ell\right). \tag{24}$$

Assumption 42 can be considered a generalization of (24). In fact, the following analogue holds, whose proof is given in Appendix A.

Lemma 11. *Let V and W satisfy Assumptions 3 and 41, and let $\psi : \mathbb{R}^d \to \mathbb{R}_+$ be a measurable function that satisfies*

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} e^{\lambda(\psi(\mathbf{x})+\psi(\mathbf{y}))} e^{-(V(\mathbf{x})+V(\mathbf{y})+W(\mathbf{x},\mathbf{y}))} d\mathbf{x}d\mathbf{y} < \infty \tag{25}$$

for all $\lambda \in \mathbb{R}$. Then there exists a lsc function $\phi : \mathbb{R}_+ \to \mathbb{R}$ with $\lim_{s\to\infty} \phi(s)/s = \infty$, such that

$$\int_{\mathbb{R}^d} \phi(\psi(\mathbf{x}))\,\mu(d\mathbf{x}) \le \inf_{\zeta\in\Pi(\mu,\mu)} \left\{\mathfrak{W}(\zeta) + \mathcal{R}\left(\zeta|e^{-V}\ell\otimes e^{-V}\ell\right)\right\}, \tag{26}$$

where $\mathfrak{W}$ is defined by (9)*.*

It is worth mentioning that [16] shows the reverse implication, which is that when the LDP holds then (23) is also true. In the case where $W \neq 0$, we have not managed to prove a similar reverse implication.

Considering Assumption 42, one may be tempted to replace Assumption 64 by the condition that there exists a lsc and superlinear function $\phi : \mathbb{R}_+ \to \mathbb{R}$ such that

$$\int_{\mathbb{R}^d} \phi(\psi(\mathbf{x}))\,\mu(d\mathbf{x}) \le \inf_{\zeta\in\Pi(\mu,\mu)} \{\mathfrak{J}(\zeta)\}, \tag{27}$$

where

$$\mathfrak{J}(\zeta) \doteq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} (V(\mathbf{x}) + V(\mathbf{y}) + W(\mathbf{x},\mathbf{y}))\,\zeta(d\mathbf{x}d\mathbf{y}). \tag{28}$$

However, it is easy to see that (27) is equivalent to Assumption 64 by choosing measures of the form $\mu = \frac{1}{2}\delta_{\mathbf{x}} + \frac{1}{2}\delta_{\mathbf{y}}, \delta_{\mathbf{x}}, \delta_{\mathbf{y}}$.

1.4.2 Examples

In the rest of the section we give examples of potentials that satisfy our assumptions. In what follows, let $K_\Delta : \mathbb{R}^d \mapsto \mathbb{R}$ be the Coulomb potential given by $K_\Delta(x) = -|x|$ when $d = 1$, $K_\Delta(x) = -\log(||x||)$ when $d = 2$ and $K_\Delta(x) = 1/||x||^{d-2}$ when $d > 2$.

Lemma 12. *Let ℓ be Lebesgue measure. The pair (V, W) given by $V(\mathbf{x}) = ||\mathbf{x}||^p$ for some $p > 1$ and $W(\mathbf{x}, \mathbf{y}) = K_\Delta(\mathbf{x} - \mathbf{y})$ satisfies Assumptions 3, 41 and 61–63, and also satisfies Assumptions 42 and 64 with $\psi(\mathbf{x}) = ||\mathbf{x}||^q$, $q < p$.*

Proof. Let $\tilde{V}_1 = \tilde{V}_2 = V/2$, and $\tilde{W} = W$. Then it is easy to see that the pair $(\tilde{V}, \tilde{W})$ is equivalent to the pair (V, W) and satisfies Assumptions 31, 91, and 92. Therefore, by Lemma 10, the pair (V, W) satisfies Assumptions 3 and 41. Moreover, it follows from assertion (2) of Theorem 1.2 and the proof of Corollary 1.3 of [5] that for $d \geq 3$, the pair (V, W) satisfies Hypotheses 8. Since these hypotheses are stronger than Assumptions 61–63, it follows that (V, W) also satisfy the latter assumptions. For the case $d = 2$, we just have to observe that $\frac{1}{4}\|\mathbf{x}\|^p + \frac{1}{4}\|\mathbf{y}\|^p - \log \|\mathbf{x} - \mathbf{y}\|$ is bounded from below by a constant c, since $-\log$ is convex and $\lim_{s\to\infty}(s^p - \log s) = \infty$. We recover Assumption 62 by picking $\gamma(s) = \frac{1}{4}s^p + C$, where C is a suitable constant. Finally, it is also easy to see that the pair (V, W) satisfies Assumptions 42 and 64 with $\psi(\mathbf{x}) = ||\mathbf{x}||^q, q < p$, by applying Lemma 11 for the first case and by picking $\phi(s) = \frac{1}{4}s^{p-q} + C$, where C is suitable constant, for the second. Verification of Assumption 63 is a direct application of point (3) in [5, Proposition 2.8]. □

Lemma 12 shows, in particular, that our assumptions are satisfied in the cases covered in [5], including the popular case studied in [5, 3, 11, 13], of $V(\mathbf{x}) = \|\mathbf{x}\|^2$, $W(\mathbf{x}, \mathbf{y}) = -\log(\mathbf{x} - \mathbf{y})$ and with ℓ Lebesgue measure. Our assumptions are also satisfied for discontinuous V and W. To give some illustrative examples, let O be an open convex set of $\mathbb{R}^d$, and K a closed convex subset of $\mathbb{R}^d$, that coincides with the closure of its interior. Furthermore, let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ be a continuous function. Consider the potentials

$$W_1(\mathbf{x}, \mathbf{y}) = \begin{cases} h(\mathbf{x}, \mathbf{y}), & (\mathbf{x}, \mathbf{y}) \in O \times O, \\ 0, & \text{otherwise}, \end{cases}$$

$$W_2(\mathbf{x}, \mathbf{y}) = \begin{cases} h(\mathbf{x}, \mathbf{y}), & (\mathbf{x}, \mathbf{y}) \in (O \times O) \cup ((O_\epsilon^c)^\circ \times (O_\epsilon^c)^\circ), \\ 0, & \text{otherwise}, \end{cases}$$

$$W_3(\mathbf{x}, \mathbf{y}) = \begin{cases} h(\mathbf{x}, \mathbf{y}), & [\mathbf{x}, \mathbf{y}] \cap (K \times K) = \emptyset, \\ 0, & \text{otherwise}, \end{cases}$$

where O_ϵ is an ϵ-neighborhood of O and in the last definition, $[\mathbf{x}, \mathbf{y}]$ stands for the straight line connecting $\mathbf{x}$ and $\mathbf{y}$. These three examples have a nice interpretation from a modeling point of view. W_1 can be interpreted as an interaction that takes place only when both particles are inside a specific area O. W_2 can be interpreted as an interaction that takes place only when both particles are inside the domain O or both outside of it. Finally if we assume that K is a "wall", W_3 can be interpreted as an interaction that takes place only when the particles can "seeëach other.

Also, unlike [5] we do not require local integrability of V and W. Hence, we can work with confined and interacting potentials that are infinite outside a bounded domain, including cases where the particles are confined within such a domain. Finally, the freedom of choice for the reference measure ℓ allows Gibbs distributions defined on sets of $\mathbb{R}^d$-Lebesgue measure zero, for example a non-smooth surface

on $\mathbb{R}^3$ or a fractal set like the Cantor dust in $\mathbb{R}^2$. A more specific example that often appears in complex potential theory is the case where W is the Coulomb potential, and ℓ is Lebesque measure on some 1-dimensional subset of $\mathbb{C}$ such as the unit circle.

1.5 Outline of the paper

The structure of the rest of the article is as follows. In Section 2 we provide definitions and lemmas that are used throughout the paper and then show that the candidate rate functions introduced above are indeed rate functions. In Section 3 we prove results for speeds $\beta_n = n$, and in Section 4 we consider the case of speeds β_n that grow faster than n. Proofs of several lemmas that are needed for the main theorems are collected in the Appendix.

2 Rate Function Property

In what follows, $\psi : \mathbb{R}^d \mapsto \mathbb{R}_+$ is always a positive, continuous function that satisfies (5). In Section 2.2, we show that under various combinations of Assumptions 3-6, the functions $\mathcal{I}_\star$ and $\mathcal{J}_\star$ defined in (14) and (20), and the functions $\mathcal{I}_\star^\psi$ and $\mathcal{J}_\star^\psi$ defined in (15) and (21) are rate functions on the spaces $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}_\psi(\mathbb{R}^d)$, respectively. To begin, in Section 2.1 we first introduce basic notions that will be used in the rest of the paper.

2.1 Basic definitions

Definition 13. Let A be an index set and let $\{\lambda_a,\, a \in A\} \subset \mathcal{P}(S)$. The collection $\{\lambda_a,\, a \in A\}$ is said to be tight if for every $\epsilon > 0$, there is a compact set $K_\epsilon \subset S$, such that $\inf\{\lambda_a(K_\epsilon),\, a \in A\} \geq 1 - \epsilon$.

Further, a sequence of random variables is said to be tight if and only if the corresponding distributions are tight. The proofs of the following three lemmas can be found in [9, 8].

Lemma 14. *A collection $\{\lambda_a, a \in A\} \subset \mathcal{P}(S)$ is tight if and only if there exists a tightness function $g : S \to [0, \infty]$ such that $\sup_{a\in A} \int_S g(x)\lambda_a(dx) < \infty$.*

Lemma 15. *Let g be a tightness function on S. Define $G : \mathcal{P}(S) \to [0, \infty]$ by*

$$G(\mu) = \int_S g(x)\mu(dx).$$

Then for each $M < \infty$ the set $\{\mu \in \mathcal{P}(S) : G(\mu) \leq M\}$ is tight (and hence precompact), and moreover, G is a tightness function on $\mathcal{P}(S)$.

Lemma 16. *Let $\{\Lambda_a, a \in A\}$ be random elements taking values in $\mathcal{P}(S)$ and let $\lambda_a = E\Lambda_a$. Then $\{\Lambda_a, a \in A\}$ is tight if and only if $\{\lambda_a, a \in A\}$ is tight. In other words, a collection of random probability measures is tight if and only if the corresponding collection of "means" is tight in the space of (deterministic) probability measures.*

The next result identifies a convenient tightness function on $\mathcal{P}_\psi\left(\mathbb{R}^d\right)$. For the proof see Appendix B.

Lemma 17. *Let $\psi : \mathbb{R}^d \to \mathbb{R}_+$ be a continuous function that satisfies the growth condition* (5)*, and let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be a lsc function with $\lim_{s\to\infty} \frac{\phi(s)}{s} = \infty$. Then*

$$\Phi(\mu) \doteq \int_{\mathbb{R}^d} \phi\left(\psi\left(\mathbf{x}\right)\right) \mu\left(d\mathbf{x}\right)$$

is a tightness function on $\mathcal{P}_\psi\left(\mathbb{R}^d\right)$ ..

Finally, it will be convenient to introduce the following projection operators to define marginal distributions.

Definition 18. We denote by $\pi^k, k = 1, 2$, the projection operators on a product space $S_1 \times S_2$ defined by

$$\pi^1 : (x_1, x_2) \to x_1 \in S_1, \qquad \pi^2 : (x_1, x_2) \to x_2 \in S_2.$$

2.2 Verification of the rate function property

We first verify that the functions $\mathcal{I}_\star$ and $\mathcal{J}_\star$ are indeed rate functions.

Lemma 19. *Suppose Assumption 3 and Assumption 41 hold. Then $\mathcal{I}_\star$ defined in (14) is a rate function on $\mathcal{P}(\mathbb{R}^d)$. Moreover, if Assumption 3 and Assumption 61 are satisfied then the function $\mathcal{J}_\star$ defined in (20) is lsc on $\mathcal{P}(\mathbb{R}^d)$. If, in addition, Assumption 62 is satisfied, then $\mathcal{J}_\star$ is a rate function on $\mathcal{P}(\mathbb{R}^d)$.*

Proof. We start by showing that the functional $\mathcal{W}$ defined in (12) is lsc. For $\mu \in \mathcal{P}(\mathbb{R}^d)$, let $\mu \otimes \mu$ denote the corresponding product measure on $\mathbb{R}^d \times \mathbb{R}^d$, and recall from (12) that $\mathcal{W}(\mu) = \mathfrak{W}(\mu \otimes \mu)$, with $\mathfrak{W}$ defined as in (9). Now, the map $\mu \to \mu \otimes \mu$ from $\mathcal{P}(\mathbb{R}^d)$ to $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is continuous, and by Fatou's lemma (for weak convergence) the map $\zeta \mapsto \mathfrak{W}(\zeta)$ is lower semicontinuous if W is lower semicontinuous and bounded from below. Since the latter property holds under Assumptions 31 and 41, it follows that $\mathcal{W}$ is lsc. Since $\mathcal{I} = \mathcal{W} + \mathcal{R}(\cdot|e^{-V}\ell)$ and, as is well known, $\mathcal{R}\left(\cdot|e^{-V}\ell\right)$ is lsc on $\mathcal{P}(\mathbb{R}^d)$, this shows that $\mathcal{I}$, and hence $\mathcal{I}_*$, are lsc. By the same argument, the lower semicontinuity of $\mathcal{J}$ can be deduced from the fact that $\mathcal{J} = \mathfrak{J}(\mu \otimes \mu)$ where $\mathfrak{J}$ is given in (28), and the fact that $(\mathbf{x}, \mathbf{y}) \mapsto W(\mathbf{x}, \mathbf{y}) + V(\mathbf{x}) + V(\mathbf{y})$ is lsc and uniformly bounded from below due to Assumptions 31 and 61, and from (20) it follows that $\mathcal{J}_*$ is lsc.

Since $\mathcal{I}_\star$ and $\mathcal{J}_\star$ are lsc, it only remains to show that the level sets of $\mathcal{I}_\star$ and $\mathcal{J}_\star$, or equivalently, $\mathcal{I}$ and $\mathcal{J}$, are (pre)compact. In the case of $\mathcal{I} = \mathcal{W} + \mathcal{R}(\cdot|e^{-V}\ell)$, this holds because $\mathcal{R}(\cdot|e^{-V}\ell)$ is a rate function on $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{W}$ is bounded below due to Assumption 41. Similarly, for $\mathcal{J}$, this holds because Assumption 62 implies that $\mathcal{J}(\mu) \geq \int_{\mathbb{R}^d} \gamma(||x||)\mu(dx)$, where $\gamma : \mathbb{R}_+ \mapsto \mathbb{R}$ is a tightness function in because it is lsc and satisfies $\gamma(s) \to \infty$ as $s \to \infty$, and the fact that the lower level sets of $\mu \mapsto \int_{\mathbb{R}^d} \gamma(||x||)\mu(dx)$ are precompact by Lemma 15. □

We next prove that, under suitable additional assumptions, $\mathcal{I}_\star$ and $\mathcal{J}_\star$ defined in (15) and (21) are rate functions on $\mathcal{P}_\psi$.

Lemma 20. *Suppose Assumptions 3, 41 and 61 are satisfied. If, in addition, there exists $\psi : \mathbb{R}^d \mapsto \mathbb{R}_+$ satisfying the growth condition* (5) *such that Assumption 42 (respectively, Assumption 64) is satisfied, then $\mathcal{I}_\star^\psi$ (respectively, $\mathcal{J}_\star^\psi$) is a rate function on $\mathcal{P}_\psi(\mathbb{R}^d)$.*

Proof. If Assumptions 3 and 41 (respectively, Assumptions 3 and 61) are satisfied, then $\mathcal{I}_\star$ (respectively, $\mathcal{J}_\star$) is lsc on $\mathcal{P}(\mathbb{R}^d)$ by Lemma 19. Since the topology on $\mathcal{P}_\psi(\mathbb{R}^d)$ is stronger than that on $\mathcal{P}(\mathbb{R}^d)$, it follows that both $\mathcal{I}^\psi_\star$ and $\mathcal{J}^\psi_\star$ are also lsc on $\mathcal{P}_\psi(\mathbb{R}^d)$. Thus, to show that $\mathcal{I}^\psi_\star$ and $\mathcal{J}^\psi_\star$ are rate functions, it suffices to show that the lower level sets of $\mathcal{I}$ and $\mathcal{J}$ are compact in $\mathcal{P}_\psi(\mathbb{R}^d)$. For $\mathcal{I}_\star$, this follows from Lemma 14 and the fact that Assumption 42 shows that there exists a superlinear, lsc function $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$ such that if $\mathcal{I} < C$ then $\int_{\mathbb{R}^d} \phi(\psi(\mathbf{x}))\, \mu(d\mathbf{x}) < C$ and, analogously, the result for $\mathcal{J}$ holds due to Lemma 14 and Assumption 64. □

3 The Case $\alpha_n = \beta_n = n$

Throughout this section, we assume that Assumptions 3 and 41 are satisfied. To establish the LDP stated in Theorem 5, by [9, Theorem 1.2.3], we can equivalently verify the Laplace principle. In view of the rate function property of $\mathcal{I}_\star$ and $\mathcal{I}^\psi_\star$ already established in Lemma 19 and Lemma 20, it suffices to show the following: for any bounded and continuous function f on S, as $n \to \infty$, the Laplace principle

$$-\frac{1}{n}\log \mathbb{E}_{Q_n}\left[e^{-nf}\right] \to \inf_{\mu\in S}\left\{f(\mu) + \mathcal{I}_\star(\mu)\right\}, \tag{29}$$

holds both for $S = \mathcal{P}(\mathbb{R}^d)$ and (under the additional condition stated as Assumption 42) with $S = \mathcal{P}_\psi(\mathbb{R}^d)$ and $\mathcal{I}_\star$ replaced by $\mathcal{I}^\psi_\star$.

Remark 4. *While the statement of [9, Theorem 1.2.3] assumes completeness of the space S, a review of the proof shows that this property is not needed (though compactness of the level sets of $\mathcal{I}_\star$ is used).*

To establish the bound (29), we first express $-\frac{1}{n}\log \mathbb{E}_{Q_n}\left[e^{-nf}\right]$ in terms of a variational problem (equivalently, a stochastic control problem). We then prove tightness of nearly minimizing controls, and finally prove convergence of the values of the corresponding controlled problems to the value of the limiting variational problem. The last step is reminiscent of the notion of Γ-convergence that is often used for analyzing variational problems in the analysis community. For a nice exposition of the relationship between LDPs and Γ-convergence, the reader is refered to [12].

In what follows, the push forward operator $\#$ is defined as follows.

Definition 21. Given measurable spaces $(S, \mathcal{F})$ and $(\tilde{S}, \tilde{\mathcal{F}})$, a measurable mapping $f : S \to \tilde{S}$ and a measure $\mu : \mathcal{F} \to [0, \infty]$, the pushforward of μ is the measure induced on $(\tilde{S}, \tilde{\mathcal{F}})$ by μ under f, i.e., the measure $f_\#(\mu) : \tilde{\mathcal{F}} \to [0, \infty]$ is given by

$$(f_\#(\mu))(B) = \mu\left(f^{-1}(B)\right) \text{ for } B \in \tilde{\mathcal{F}}.$$

3.1 Representation formula

Recall that P_n is the probability measure $\mathbb{R}^{nd}$ defined in (3) and Q_n is the push forward of P_n under L_n. Let $P^\star_n$ be the measure on $\mathbb{R}^{nd}$ defined by

$$P^\star_n(d\mathbf{x}_1, \ldots, d\mathbf{x}_n) \doteq e^{-\sum_{i=1}^n V(\mathbf{x}_i)}\ell(d\mathbf{x}_1)\cdots\ell(d\mathbf{x}_n), \tag{30}$$

and note that it is a probability measure due to Assumption 32. Analogous to $\mathcal{W}$ defined in (12), $\mathcal{W}_{\neq}$ is defined as follows: for $\mu \in \mathcal{P}(\mathbb{R}^d)$,

$$\mathcal{W}_{\neq}(\mu) \doteq \frac{1}{2}\int_{\neq} W(\mathbf{x},\mathbf{y})\,\mu(d\mathbf{x})\,\mu(d\mathbf{y}) = \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W_{\neq}(\mathbf{x},\mathbf{y})\,\mu(d\mathbf{x})\,\mu(d\mathbf{y}). \tag{31}$$

with

$$W_{\neq}(\mathbf{x},\mathbf{y}) \doteq \begin{cases} W(\mathbf{x},\mathbf{y}), & \text{when } \mathbf{x} \neq \mathbf{y}, \\ 0, & \text{when } \mathbf{x} = \mathbf{y}. \end{cases}$$

Since $\beta_n = n$, for any measurable function f on $\mathcal{P}(\mathbb{R}^d)$ (or on $\mathcal{P}_\psi(\mathbb{R}^d)$), we have

$$-\frac{1}{n}\log \mathbb{E}_{Q_n}\left[e^{-nf}\right] = -\frac{1}{n}\log \mathbb{E}_{P_n}\left[e^{-nf\circ L_n}\right] = -\frac{1}{n}\log \mathbb{E}_{P_n^\star}\left[\frac{1}{Z_n}e^{-n(f+\mathcal{W}_{\neq})\circ L_n}\right], \tag{32}$$

where Z_n is the normalizing constant defined in (4).

We next state a representation for the quantity on the right-hand side of (32). To avoid confusion with the original distibutions and random variables, we use an overbar (e.g., $\bar{L}_n$) for quantities that will appear in the representation, and refer to them as "controlled" versions.

Given a probability measure $\bar{P}^n \in \mathcal{P}(\mathbb{R}^{nd})$, we can factor it into conditional distributions in the following manner:

$$\bar{P}^n(d\mathbf{x}_1,\ldots,d\mathbf{x}_n) = \bar{P}^n_{\{1\}}(d\mathbf{x}_1)\bar{P}^n_{\{2\}|\{1\}}(d\mathbf{x}_2|\mathbf{x}_1)\cdots\bar{P}^n_{\{n\}|\{1,..,n-1\}}(d\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}),$$

where for $i = 1, ..., n$, $\bar{P}_{i|1,...,i-1}(\cdot|\mathbf{x}_1, ..., \mathbf{x}_{i-1})$ denotes the conditional distribution of the i-th marginal given $\mathbf{x}_1, ..., \mathbf{x}_{i-1}$. Thus, if $\{\bar{\boldsymbol{X}}^n_j\}_{1\le j\le n}$ are random variables with joint distribution $\bar{P}^n(d\mathbf{x}_1\cdots d\mathbf{x}_n)$ on some probability space $(\Omega,\mathcal{F},\mathbb{P})$, then $\bar{\mu}^n_i$, the conditional distribution of $\bar{\mathbf{X}}^n_i$ given $\bar{\mathbf{X}}^n_1,\ldots,\bar{\mathbf{X}}^n_{i-1}$, can be expressed as

$$\bar{\mu}^n_i(d\mathbf{x}_i) \doteq \bar{P}^n_{\{i\}|\{1,...,i-1\}}(d\mathbf{x}_i|\bar{\mathbf{X}}^n_1,\ldots,\bar{\mathbf{X}}^n_{i-1}). \tag{33}$$

Note that $\bar{\mu}^n_i$, $1 \le i \le n$, are random probability measures, and the ith measure is measurable with respect to the σ-algebra generated by $\{\bar{\boldsymbol{X}}^n_j\}_{j<i}$. We refer to the collection $\{\bar{\mu}^n_i, 1 \le i \le n\}$ as a control, and let $\bar{L}_n(\cdot) = L_n(\bar{\mathbf{X}}^n;\cdot)$, with L_n defined by (2), be the (random) empirical measure of $\{\bar{\boldsymbol{X}}^n_j\}_{1\le j\le n}$, which we refer to as the controlled empirical measure. We denote expectation with respect to $\mathbb{P}$ by $\mathbb{E}$, or by $\mathbb{E}_{\mathbb{P}}$, when we want to emphasize the dependence on $\mathbb{P}$.

Let f belong to the space of functions on $\mathcal{P}(\mathbb{R}^d)$ (or $\mathcal{P}_\psi(\mathbb{R}^d)$) such that the map $\mathbf{x}^n \mapsto f(L_n(\mathbf{x}^n;\cdot))$ from $\mathbb{R}^{nd}$ to $\mathbb{R}$ is measurable and bounded from below. This space clearly includes all bounded continuous functions on $\mathcal{P}(\mathbb{R}^d)$ (respectively, $\mathcal{P}_\psi(\mathbb{R}^d)$). Then, since the functional $\mathcal{W}_{\neq}$ is also measurable and bounded from below (due to Assumption 41), we can apply Proposition 4.5.1 in [9] to the function $\mathbf{x}^n \in \mathbb{R}^d \mapsto f(L_n(\mathbf{x}^n;\cdot)) + \mathcal{W}_{\neq}(L_n(\mathbf{x}^n;\cdot))$, to obtain

$$-\frac{1}{n}\log \mathbb{E}_{P_n^\star}\left[e^{-n(f+\mathcal{W}_{\neq})\circ L_n}\right] = \inf_{\{\bar{\mu}^n\}} \mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{W}_{\neq}\left(\bar{L}_n\right) + \mathcal{R}\left(\bar{P}^n|\otimes_n e^{-V}\ell\right)\right],$$

where $\bar{L}_n$ is the controlled empirical measure associated with $\bar{P}^n$ as defined above, and the infimum is over all controls $\{\bar{\mu}^n_i\}$ defined in terms of some joint distribution $\bar{P}^n \in \mathcal{P}(\mathbb{R}^{nd})$ via (33). Factoring $\bar{P}^n$ as above and using the chain rule for relative entropy (see [9, Theorem B.2.1]), we then have

$$-\frac{1}{n}\log \mathbb{E}_{P_n^\star}\left[e^{-n(f+\mathcal{W}_{\neq})\circ L_n}\right] = \inf_{\{\bar{\mu}^n_i\}} \mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{W}_{\neq}\left(\bar{L}_n\right) + \frac{1}{n}\sum_{i=1}^n \mathcal{R}\left(\bar{\mu}^n_i|e^{-V}\ell\right)\right], \tag{34}$$

where the infimum is over all controls $\{\bar{\mu}^n_i\}$ (equivalently, joint distributions $\bar{P}^n \in \mathcal{P}(\mathbb{R}^{nd})$). Also, setting $f = 0$ in (34) and recalling the definition of Z_n from (4) gives

$$-\frac{1}{n}\log\left(Z_n\right) = -\frac{1}{n}\log \mathbb{E}_{P_n^\star}\left[e^{-n\mathcal{W}_{\neq}(L_n)}\right] = \inf_{\{\bar{\mu}^n_i\}} \mathbb{E}\left[\mathcal{W}_{\neq}\left(\bar{L}_n\right) + \frac{1}{n}\sum_{i=1}^n \mathcal{R}\left(\bar{\mu}^n_i|e^{-V}\ell\right)\right]. \tag{35}$$

We claim that to prove Theorem 5, it suffices to show that for every bounded and continuous (in the respective topology) function f, the lower bound

$$\liminf_{n\to\infty} \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{W}_{\neq}\left(\bar{L}_n\right) + \frac{1}{n}\sum_{i=1}^{n} \mathcal{R}\left(\bar{\mu}_i^n | e^{-V}\ell\right)\right] \geq \inf_{\mu}\left[f(\mu) + \mathcal{I}(\mu)\right] \tag{36}$$

and upper bound

$$\limsup_{n\to\infty} \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{W}_{\neq}\left(\bar{L}_n\right) + \frac{1}{n}\sum_{i=1}^{n} \mathcal{R}\left(\bar{\mu}_i^n | e^{-V}\ell\right)\right] \leq \inf_{\mu}\left[f(\mu) + \mathcal{I}(\mu)\right] \tag{37}$$

hold. Indeed, when combined with (34), (35) and (32), these bounds imply the desired limit (29). The lower and upper bounds are established in Section 3.3 and 3.4, respectively. First, in Section 3.2, we establish some tightness properties of the controls that will be used in the proofs of these bounds.

3.2 Properties of the controls

We continue to use the notation for the controls introduced in the previous section. We start with a simplifying observation.

Remark 5. *In the proof of the lower bound* (36)*, we can assume that there exists $C_0 < \infty$ such that*

$$\sup_{n\in\mathbb{N}} \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[\mathcal{W}_{\neq}\left(\bar{L}_n\right) + \frac{1}{n}\sum_{i=1}^{n} \mathcal{R}\left(\bar{\mu}_i^n | e^{-V}\ell\right)\right] \leq C_0. \tag{38}$$

If this were not true, we could restrict to a subsequence that has such a property, because for any subsequence for which the left-hand side of (38) *is infinite, the lower bound* (36) *is satisfied by default. Furthermore, since under Assumption 41, $\mathcal{W}_{\neq} > \min\{0, c\}$, we can restrict to controls for which the relative entropy cost is bounded by $C_0 + |c|$: that is, for which*

$$\sup_{n} \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathcal{R}\left(\bar{\mu}_i^n | e^{-V}\ell\right)\right] \leq C_0 + |c|. \tag{39}$$

Lemma 22. *Let V satisfy Assumption 32, and let $\{\bar{\mu}_i^n\}, n \in \mathbb{N}$, be a sequence of controls for which* (39) *holds, let $\bar{L}^n$ be the associated sequence of controlled empirical measures and let*

$$\hat{\mu}_n \doteq \frac{1}{n}\sum_{i=1}^{n} \bar{\mu}_i^n. \tag{40}$$

Then $\left\{\left(\bar{L}_n, \hat{\mu}_n\right), n \in \mathbb{N}\right\}$ is tight as a sequence of $\mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)$-valued random elements.

Proof. Let $\{\bar{\mu}_i^n\}, n \in \mathbb{N}$, be a sequence of controls that satisfies (39). By the convexity of relative entropy and Jensen's inequality

$$\sup_{n} \mathbb{E}\left[\mathcal{R}\left(\hat{\mu}^n | e^{-V}\ell\right)\right] < \infty.$$

We know that $\mathcal{R}\left(\cdot | e^{-V}\ell\right)$ is a tightness function on $\mathcal{P}\left(\mathbb{R}^d\right)$ and hence, by Lemma 14, the sequence of random probability measures $\{\hat{\mu}_n, n \in \mathbb{N}\}$ is tight. By Lemma 16, the sequence of probability measures $\{\mathbb{E}[\hat{\mu}_n],\ n \in \mathbb{N}\}$ is tight. Since $\bar{\mu}_i^n$ is the conditional distribution of $\bar{\mathbf{X}}_i^n$ given $(\bar{\mathbf{X}}_1^n, ..., \bar{\mathbf{X}}_{i-1}^n)$,

for any measurable function $g : \mathbb{R}^d \mapsto \mathbb{R}$ that is bounded from below, we have

$$\begin{aligned}\mathbb{E}\left[\int_{\mathbb{R}^d} g(\mathbf{x})\,\bar{L}_n(d\mathbf{x})\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n g\left(\bar{\mathbf{X}}_i^n\right)\right] \\ &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d} g(\mathbf{x})\,\bar{\mu}_i^n(d\mathbf{x})\right] = \mathbb{E}\left[\int_{\mathbb{R}^d} g(\mathbf{x})\,\hat{\mu}_n(d\mathbf{x})\right].\end{aligned} \tag{41}$$

Thus, $\mathbb{E}\left[\bar{L}_n\right] = \mathbb{E}\left[\hat{\mu}_n\right]$, and so $\{\mathbb{E}[\bar{L}_n], n \in \mathbb{N}\}$ is also tight. Another application of Lemma 16 then shows that $\{\bar{L}_n, n \in \mathbb{N}\}$, is tight, which together with the tightness of $\{\hat{\mu}^n\}$ established above, implies $\left\{\left(\hat{\mu}_n, \bar{L}_n\right), n \in \mathbb{N}\right\}$ is tight. □

The following lemma, which uses an elementary martingale argument, appears in [8]. For the reader's convenience the proof is given in Appendix C.

Lemma 23. *Suppose $\bar{L}_n, \hat{\mu}_n, n \in \mathbb{N}$, are as defined in Lemma 22 and further assume that the couple $\left\{\left(\bar{L}_n, \hat{\mu}_n\right), n \in \mathbb{N}\right\}$ converges along a subsequence to $\left(\bar{L}, \hat{\mu}\right)$. Then $\bar{L} = \hat{\mu}$ w.p.1.*

For the next result, it will be convenient to first define a collection of auxiliary random measures that extend the ones that appear in the representation (34). Let $\bar{P}^n$ be a probability measure on $\mathbb{R}^{nd}$, and let $(\bar{\mathbf{X}}_1^n, \ldots, \bar{\mathbf{X}}_n^n)$ be random variables with joint distribution $\bar{P}^n$. For $J \subset \{1, ..., n\}$, let $\bar{P}_J^n$ equal the marginal distribution of $\bar{P}^n$ on $\{\mathbf{x}_j, j \in J\}$, and for disjoint subsets I_1 and I_2 of $\{1, \ldots, n\}$, let $\bar{P}_{I_1|I_2}^n$ denote the stochastic kernel defined as follows:

$$\bar{P}_{I_1|I_2}^n(d\mathbf{x}_i, i \in I_1|\mathbf{x}_k, k \in I_2)\bar{P}_{I_2}^n(d\mathbf{x}_k, k \in I_2) = \bar{P}_{I_1\cup I_2}^n(d\mathbf{x}_j, j \in I_1 \cup I_2).$$

Let $K_k \doteq \{1, \ldots, k-1\}$. In the sequel we fix $i < j$ (the case $j < i$ can be handled in a symmetric way), and define

$$\bar{\mu}_{ij}^n(d\mathbf{x}_i d\mathbf{x}_j) \doteq \bar{P}_{\{i,j\}|K_i}^n(d\mathbf{x}_i d\mathbf{x}_j|\bar{\mathbf{X}}_k^n, k \in K_i). \tag{42}$$

Also, note that with this notation

$$\bar{\mu}_i^n(d\mathbf{x}_i) = \bar{P}_{\{i\}|K_i}^n(d\mathbf{x}_i|\bar{\mathbf{X}}_1^n, \ldots, \bar{\mathbf{X}}_{i-1}^n) \tag{43}$$

are the controls used in the representation (34). We claim that

$$\pi_{\#}^1\bar{\mu}_{ij}^n = \bar{\mu}_i^n \quad \text{and} \quad \pi_{\#}^2\bar{\mu}_{ij}^n = \mathbb{E}[\bar{\mu}_j^n|\bar{\mathbf{X}}_k^n, k \in K_i], \tag{44}$$

where $\pi^k, k = 1, 2$, and $\#$ are the projection and push-forward operators introduced in Definition 18 and Definition 21. The first relation in (44) is an immediate consequence of the definitions of $\bar{\mu}_i^n$ and $\bar{\mu}_{ij}^n$. Due to the asymmetry in the first and second (equivalently, i and j) coordinates in the definition of $\bar{\mu}_{ij}^n$ in (42), the proof of the second equality in (44) is a little more involved. Indeed, note that for every $A \subset \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}&\pi_{\#}^2\bar{\mu}_{ij}^n(A) \\ &\quad= \pi_{\#}^2\bar{P}_{\{i,j\}|K_i}^n(A|\bar{\mathbf{X}}_k^n, k \in K_i) \\ &\quad= \int \bar{P}_{\{j\}|K_{i+1}}^n(A|\bar{\mathbf{X}}_1^n, ..., \bar{\mathbf{X}}_{i-1}^n, \mathbf{x}_i)\bar{P}_{\{i\}|K_i}^n(d\mathbf{x}_i|\bar{\mathbf{X}}_k^n, k \in K_i) \\ &\quad= \int \bar{P}_{\{j\}|K_j}^n(A|\bar{\mathbf{X}}_1^n, ..., \bar{\mathbf{X}}_{i-1}^n, \mathbf{x}_i, ..., \mathbf{x}_{j-1})\bar{P}_{(K_j\setminus K_i)|K_i}^n(d\mathbf{x}_i \cdots d\mathbf{x}_{j-1}|\bar{\mathbf{X}}_1^n, ..., \bar{\mathbf{X}}_{i-1}^n) \\ &\quad= \mathbb{E}[\bar{P}_{\{j\}|K_j}^n(A|\bar{\mathbf{X}}_k^n, k \in K_j)|\bar{\mathbf{X}}_k^n, k \in K_i] \\ &\quad= \mathbb{E}[\bar{\mu}_j^n|\bar{\mathbf{X}}_k^n, k \in K_i](A),\end{aligned}$$

from which the second equality in (44) follows.

Lemma 24. *Let V and W satisfy Assumptions 3 and 4, let $\{\bar{\mu}_i^n\}, n \in \mathbb{N}$, be a sequence of controls for which*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[\mathcal{W}_{\neq}\left(\bar{L}_n\right) + \frac{1}{n}\sum_{i=1}^{n} \mathcal{R}\left(\bar{\mu}_i^n | e^{-V}\ell\right)\right] < \infty, \tag{45}$$

and let $\hat{\mu}_n$ be as defined in (40)*. Then $\left\{\left(\bar{L}_n, \hat{\mu}_n\right), n \in \mathbb{N}\right\}$ is tight in $\mathcal{P}_\psi(\mathbb{R}^d) \times \mathcal{P}_\psi(\mathbb{R}^d)$.*

Proof. Let θ be a probability measure on $\mathbb{R}^d$. By the chain rule for relative entropy, we have

$$\begin{aligned}
&\mathcal{R}(\bar{P}^n_{\{i,j\}|K_i}(d\mathbf{x}_i d\mathbf{x}_j|\mathbf{x}_k, k \in K_i) \,\|\theta(d\mathbf{x}_i)\theta(d\mathbf{x}_j)) \\
&\quad = \int \mathcal{R}(\bar{P}^n_{\{j\}|K_{i+1}}(d\mathbf{x}_j|\mathbf{x}_k, k \in K_{i+1}) \,\|\theta(d\mathbf{x}_j))\bar{P}^n_{\{i\}|K_i}(d\mathbf{x}_i|\mathbf{x}_k, k \in K_i) \\
&\qquad + \mathcal{R}(\bar{P}^n_{\{i\}|K_i}(d\mathbf{x}_i|\mathbf{x}_k, k \in K_i) \,\|\theta(d\mathbf{x}_i)).
\end{aligned}$$

In addition, Jensen's inequality gives

$$\begin{aligned}
&\mathcal{R}(\bar{P}^n_{\{j\}|K_{i+1}}(d\mathbf{x}_j|\mathbf{x}_k, k \in K_{i+1}) \,\|\theta(d\mathbf{x}_j)) \\
&= \mathcal{R}\left(\int \bar{P}^n_{\{j\}|K_j}(d\mathbf{x}_j|\mathbf{x}_k, k \in K_j)\bar{P}^n_{(K_j\setminus K_{i+1})|K_{i+1}}(d\mathbf{x}_{i+1}\cdots d\mathbf{x}_{j-1}|\mathbf{x}_k, k \in K_{i+1})\middle\| \theta(d\mathbf{x}_j)\right) \\
&\leq \int \mathcal{R}\left(\bar{P}^n_{\{j\}|K_j}(d\mathbf{x}_j|\mathbf{x}_k, k \in K_j)\middle\| \theta(d\mathbf{x}_j)\right) \bar{P}^n_{(K_j\setminus K_{i+1})|K_{i+1}}(d\mathbf{x}_{i+1}\cdots d\mathbf{x}_{j-1}|\mathbf{x}_k, k \in K_{i+1}).
\end{aligned}$$

Combining the last two displays with (42) and (43), we obtain

$$\mathbb{E}\left[\mathcal{R}(\bar{\mu}^n_{ij}(d\mathbf{x}_i d\mathbf{x}_j) \,\|\theta(d\mathbf{x}_i)\theta(d\mathbf{x}_j))\right] \leq \mathbb{E}[\mathcal{R}(\bar{\mu}^n_j(d\mathbf{x}_j) \,\|\theta(d\mathbf{x}_j)) + \mathcal{R}(\bar{\mu}^n_i(d\mathbf{x}_i) \,\|\theta(d\mathbf{x}_i))]. \tag{46}$$

Using (46) with $\theta = e^{-V}\ell$, the definition of $\mathcal{W}_{\neq}$ in (31) and the tower property of conditional expecta-

tions to get the first inequality below, we have

$$
\begin{aligned}
&\mathbb{E}\left[\mathcal{W}_{\neq}\left(\bar{L}_n\right)+\frac{1}{n}\sum_{i=1}^{n}\mathcal{R}\left(\bar{\mu}_i^n|e^{-V}\ell\right)\right]\\
&=\mathbb{E}\left[\mathcal{W}_{\neq}\left(\bar{L}_n\right)+\frac{1}{n(n-1)}(n-1)\sum_{i=1}^{n}\mathcal{R}\left(\bar{\mu}_i^n|e^{-V}\ell\right)\right]\\
&\geq\mathbb{E}\Bigg[\frac{1}{2n^2}\sum_{i<j}\int_{\mathbb{R}^d}W\left(\bar{\mathbf{X}}_i^n,\mathbf{x}_j\right)\bar{P}_{\{j\}|K_{i+1}}(d\mathbf{x}_j|\bar{\mathbf{X}}_k^n,k\in K_{i+1})\\
&\qquad+\frac{1}{2n^2}\sum_{j<i}\int_{\mathbb{R}^d}W\left(\mathbf{x}_i,\bar{\mathbf{X}}_j^n\right)\bar{P}_{\{i\}|K_{j+1}}(d\mathbf{x}_i|\bar{\mathbf{X}}_k^n,k\in K_{j+1})\\
&\qquad+\frac{1}{n(n-1)}\sum_{i<j}\mathcal{R}\left(\bar{\mu}_{ij}^n|e^{-V}\ell\otimes e^{-V}\ell\right)+\frac{1}{n(n-1)}\sum_{j<i}\mathcal{R}\left(\bar{\mu}_{ij}^n|e^{-V}\ell\otimes e^{-V}\ell\right)\Bigg] \qquad (47)\\
&=\mathbb{E}\Bigg[\frac{1}{2n^2}\sum_{i<j}\int_{\mathbb{R}^d\times\mathbb{R}^d}W\left(\mathbf{x}_i,\mathbf{x}_j\right)\bar{P}_{\{i,j\}|K_i}(d\mathbf{x}_i d\mathbf{x}_j|\bar{\mathbf{X}}_k^n,k\in K_i)\\
&\qquad+\frac{1}{2n^2}\sum_{j<i}\int_{\mathbb{R}^d\times\mathbb{R}^d}W\left(\mathbf{x}_i,\mathbf{x}_j\right)\bar{P}_{\{i,j\}|K_j}(d\mathbf{x}_j d\mathbf{x}_i|\bar{\mathbf{X}}_k^n,k\in K_j)\\
&\qquad+\frac{1}{n(n-1)}\sum_{i<j}\mathcal{R}\left(\bar{\mu}_{ij}^n|e^{-V}\ell\otimes e^{-V}\ell\right)+\frac{1}{n(n-1)}\sum_{j<i}\mathcal{R}\left(\bar{\mu}_{ij}^n|e^{-V}\ell\otimes e^{-V}\ell\right)\Bigg]\\
&=\mathbb{E}\left[\frac{1}{n^2}\sum_{i\neq j}\mathfrak{W}\left(\bar{\mu}_{ij}^n\right)+\frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{R}\left(\bar{\mu}_{ij}^n|e^{-V}\ell\otimes e^{-V}\ell\right)\right],
\end{aligned}
$$

where $\mathfrak{W}$ is the functional defined in (9). Next, let

$$
\hat{\mu}^{2,n}\doteq\frac{1}{n(n-1)}\sum_{i\neq j}\bar{\mu}_{ij}^n. \qquad (48)
$$

Then combining (47) with the convexity of $\mathcal{R}$ in both arguments (see [9, Lemma 1.4.3]), the linearity of $\mathfrak{W}$, and the definition of $\hat{\mu}^{2,n}$ in (48), we obtain

$$
\mathbb{E}\left[\mathcal{W}_{\neq}\left(\bar{L}_n\right)+\frac{1}{n}\sum_{i=1}^{n}\mathcal{R}\left(\bar{\mu}_i^n|e^{-V}\ell\right)\right]\geq\mathbb{E}\left[\frac{n-1}{n}\mathfrak{W}\left(\hat{\mu}^{2,n}\right)+\mathcal{R}\left(\hat{\mu}^{2,n}|e^{-V}\ell\otimes e^{-V}\ell\right)\right]. \qquad (49)
$$

We now use (49) to establish tightness of both $\{\bar{L}_n\}$ and $\{\hat{\mu}^n\}$ in the d_ψ topology. Note that $\hat{\mu}^{2,n}$ is a random probability measure on $\mathbb{R}^d\times\mathbb{R}^d$ and that it has identical marginals. Since V and W satisfy Assumption 42 and relative entropy is nonnegative, there exists a superlinear function ϕ for which we have the inequalities

$$
\begin{aligned}
&\mathbb{E}\left[\frac{n-1}{n}\mathfrak{W}\left(\hat{\mu}^{2,n}\right)+\mathcal{R}\left(\hat{\mu}^{2,n}|e^{-V}\ell\otimes e^{-V}\ell\right)\right]\\
&\quad\geq\mathbb{E}\left[\frac{n-1}{n}\Big(\mathfrak{W}\left(\hat{\mu}^{2,n}\right)+\mathcal{R}\left(\hat{\mu}^{2,n}|e^{-V}\ell\otimes e^{-V}\ell\right)\Big)\right] \qquad (50)\\
&\quad\geq\frac{n-1}{n}\mathbb{E}\left[\int_{\mathbb{R}^d}\phi\left(\psi\left(\boldsymbol{x}\right)\right)\left(\pi^1_{\#}\hat{\mu}^{2,n}\right)(d\boldsymbol{x})\right].
\end{aligned}
$$

For $n \geq 2$, combining (49) and (50) gives

$$2\mathbb{E}\left[\mathcal{W}_{\neq}\left(\bar{L}_n\right) + \frac{1}{n}\sum_{i=1}^{n}\mathcal{R}\left(\bar{\mu}_i^n | e^{-V}\ell\right)\right] \geq \mathbb{E}\left[\int_{\mathbb{R}^d}\phi\left(\psi\left(\boldsymbol{x}\right)\right)\left(\pi_{\#}^1\hat{\mu}^{2,n}\right)\left(d\boldsymbol{x}\right)\right]. \tag{51}$$

Note that (44) implies $\mathbb{E}[\pi_{\#}^1\bar{\mu}_{ij}^n] = \mathbb{E}[\bar{\mu}_i^n]$ and $\mathbb{E}[\pi_{\#}^2\bar{\mu}_{ij}^n] = \mathbb{E}[\bar{\mu}_j^n]$. Further, recalling the definition of $\hat{\mu}^n$ in (40) and $\hat{\mu}^{2,n}$ in (48), this shows that

$$\mathbb{E}[\pi_{\#}^1\hat{\mu}^{2,n}] = \mathbb{E}[\pi_{\#}^2\hat{\mu}^{2,n}] = \mathbb{E}[\hat{\mu}^n]. \tag{52}$$

Substituting this into the right-hand side of (51) and letting $C_0 < \infty$ denote the left-hand side of (45), we obtain the bound

$$\mathbb{E}\left[\int_{\mathbb{R}^d}\phi\left(\psi\left(\boldsymbol{x}\right)\right)\hat{\mu}^n\left(d\boldsymbol{x}\right)\right] \leq 2C_0.$$

However, since we know from Lemma 17 that $\Phi(\mu) = \int_{\mathbb{R}^d}\phi\left(\psi\left(\mathbf{x}\right)\right)\mu\left(d\mathbf{x}\right)$ is a tightness function on $\mathcal{P}_\psi\left(\mathbb{R}^d\right)$, it follows that $\{\hat{\mu}^n\}$ is tight as a collection of $\mathcal{P}_\psi\left(\mathbb{R}^d\right)$-valued random elements. Finally, note that we have the equality

$$\begin{aligned}\mathbb{E}\left[\int_{\mathbb{R}^d} g\left(\mathbf{x}\right)\bar{L}_n\left(d\mathbf{x}\right)\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} g\left(\bar{\mathbf{X}}_i^n\right)\right] \\ &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d} g\left(\mathbf{x}\right)\bar{\mu}_i^n\left(d\mathbf{x}\right)\right] = \mathbb{E}\left[\int_{\mathbb{R}^d} g\left(\mathbf{x}\right)\hat{\mu}_n\left(d\mathbf{x}\right)\right].\end{aligned} \tag{53}$$

Setting $g(\mathbf{x}) = \phi(\psi(\mathbf{x}))$, and again invoking Lemma 17, we see that $\{\bar{L}_n\}$ is also tight. □

Remark 6. *In the remainder of the proof, which is carried out in Sections 3.3 and 3.4, the arguments for both $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}_\psi(\mathbb{R}^d)$ are similar, and so we will treat both cases simultaneously. The functions f used will be considered continuous in the respective topology and any infimum taken should be with respect to the corresponding set $\mathcal{P}(\mathbb{R}^d)$ or $\mathcal{P}_\psi(\mathbb{R}^d)$.*

Remark 7. *Due to Remark 5 and Lemma 22 and Lemma 24, it is without loss of generality, for the lower bound* (36)*, to restrict to controls for which $\left\{\left(\bar{L}_n, \hat{\mu}_n\right), n \in \mathbb{N}\right\}$ is tight in $\mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)$, or (with the additional Assumption 42) in $\mathcal{P}_\psi(\mathbb{R}^d) \times \mathcal{P}_\psi(\mathbb{R}^d)$.*

3.3 Proof of the lower bound

For the proof of the lower bound (36) we will use some auxiliary functionals. For $d' \in \mathbb{N}$, an arbitrary function $F : \mathbb{R}^{d'} \to (-\infty, \infty]$ and $M \in [0, \infty)$, let $F^M(\mathbf{z}) \doteq \min\{F(\mathbf{z}), M\}$. For $\mu \in \mathcal{P}(\mathbb{R}^d)$, let

$$\mathcal{W}^M\left(\mu\right) \doteq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W^M\left(\mathbf{x},\mathbf{y}\right)\mu\left(d\mathbf{x}\right)\mu\left(d\mathbf{y}\right),$$

$$\mathcal{W}_{\neq}^M\left(\mu\right) \doteq \frac{1}{2}\int_{\neq} W^M\left(\mathbf{x},\mathbf{y}\right)\mu\left(d\mathbf{x}\right)\mu\left(d\mathbf{y}\right) = \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W_{\neq}^M\left(\mathbf{x},\mathbf{y}\right)\mu\left(d\mathbf{x}\right)\mu\left(d\mathbf{y}\right),$$

and note that for every $\mu \in \mathcal{P}(\mathbb{R}^d)$,

$$\mathcal{W}^M\left(\mu\right) \leq \mathcal{W}_{\neq}^M\left(\mu\right) + \frac{M}{2}(\mu\otimes\mu)\{(x,x) : x \in \mathbb{R}^d\}. \tag{54}$$

Let $\epsilon > 0$ be given. Then by (38) and the boundedness of f, there exist $C' \in \mathbb{R}$ and a sequence of controls $\{\tilde{\mu}_i^n\}$ with associated sequence of controlled empirical measures $\{\tilde{L}_n\}$, such that

$$\begin{aligned} C' &> \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{W}_{\neq}\left(\bar{L}_n\right) + \tfrac{1}{n}\sum_{i=1}^n \mathcal{R}\left(\bar{\mu}_i^n | e^{-V}\ell\right)\right] + \epsilon \\ &\geq \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{W}_{\neq}\left(\tilde{L}_n\right) + \tfrac{1}{n}\sum_{i=1}^n \mathcal{R}\left(\tilde{\mu}_i^n | e^{-V}\ell\right)\right] \\ &\geq \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{W}_{\neq}^M\left(\tilde{L}_n\right) + \tfrac{1}{n}\sum_{i=1}^n \mathcal{R}\left(\tilde{\mu}_i^n | e^{-V}\ell\right)\right] \\ &\geq \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{W}^M\left(\tilde{L}_n\right) - \tfrac{M}{n} + \tfrac{1}{n}\sum_{i=1}^n \mathcal{R}\left(\tilde{\mu}_i^n | e^{-V}\ell\right)\right], \end{aligned} \tag{55}$$

where $\mathcal{W}_{\neq} \geq \mathcal{W}_{\neq}^M$ is used for the third inequality and the last inequality uses (54) and the fact that $\bar{L}^n \otimes \bar{L}^n$ put mass at most $1/n$ on the diagonal of $\mathbb{R}^d \times \mathbb{R}^d$.

Let $\hat{\mu}_n \doteq \frac{1}{n}\sum_{i=1}^n \tilde{\mu}_i^n$. Since Lemma 24 implies $\{(\tilde{L}_n, \hat{\mu}_n)\}$ is tight, we can extract a further subsequence, which we denote again by $\{(\tilde{L}_n, \hat{\mu}_n)\}$, which converges in distribution to some limit $(\tilde{L}, \hat{\mu})$. If the lower bound is demonstrated for this subsequence, the standard argument by contradiction establishes the lower bound for the original sequence. Let $\{M_n\}$ be an increasing sequence such that $\lim_{n\to\infty} M_n = \infty$ and $\lim_{n\to\infty} \frac{M_n}{n} = 0$, and let $m \in \mathbb{N}$. By the monotonicity of $n \mapsto \mathcal{W}^{M_n}$, Jensen's inequality, the definition of $\hat{\mu}_n$, and Fatou's lemma we have

$$\begin{aligned} &\liminf_{n\to\infty} \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{W}^{M_n}\left(\tilde{L}_n\right) - \frac{M_n}{n} + \frac{1}{n}\sum_{i=1}^n \mathcal{R}\left(\tilde{\mu}_i^n | e^{-V}\ell\right)\right] \\ &\geq \liminf_{n\to\infty} \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{W}^{M_m}\left(\tilde{L}_n\right) - \frac{M_n}{n} + \mathcal{R}\left(\hat{\mu}_n | e^{-V}\ell\right)\right] \\ &\geq \mathbb{E}\left[f\left(\tilde{L}\right) + \mathcal{W}^{M_m}\left(\tilde{L}\right) + \mathcal{R}\left(\hat{\mu} | e^{-V}\ell\right)\right], \end{aligned} \tag{56}$$

where the continuity of f and lower semicontinuity of $\mathcal{W}^{M_m}$ and $\mathcal{R}(\cdot|e^{-V}\ell)$ are also used in the last inequality. Since this inequality holds for arbitrary $m \in \mathbb{N}$, the monotone convergence theorem, the property that $\tilde{L} = \hat{\mu}$ almost surely (due to Lemma 23) and the definition of $\mathcal{I}$ in (14), together imply

$$\begin{aligned} \lim_{m\to\infty} \mathbb{E}\left[f\left(\tilde{L}\right) + \mathcal{W}^{M_m}\left(\tilde{L}\right) + \mathcal{R}\left(\hat{\mu} | e^{-V}\ell\right)\right] &\geq \mathbb{E}\left[f\left(\hat{\mu}\right) + \mathcal{W}\left(\hat{\mu}\right) + \mathcal{R}(\hat{\mu} | e^{-V}\ell)\right] \\ &\geq \inf_{\mu}\left[f\left(\mu\right) + \mathcal{I}\left(\mu\right)\right]. \end{aligned} \tag{57}$$

Since $\epsilon > 0$ is arbitrary, (55), (56) and (57) together imply the lower bound (36).

3.4 Proof of the upper bound

Again, fix f to be a bounded continuous function on $\mathcal{P}(\mathbb{R}^d)$, let $\epsilon > 0$ and let $\mu^* \in \mathcal{P}(\mathbb{R}^d)$ (respectively, $\mathcal{P}_\psi(\mathbb{R}^d)$) be such that

$$f\left(\mu^*\right) + \mathcal{W}\left(\mu^*\right) + \mathcal{R}\left(\mu^* | e^{-V}\ell\right) \leq \inf_{\mu}\left[f\left(\mu\right) + \mathcal{I}(\mu)\right] + \epsilon. \tag{58}$$

For $n \in \mathbb{N}$, let $\{\tilde{\mu}_i^n, 1 \leq i \leq n\}$ denote the particular control defined by $\tilde{\mu}_i^n \doteq \mu^*$ for all $n \in \mathbb{N}$ and $i \in \{1, ..., n\}$, and let $\tilde{\mathbf{X}}_i^n, i = 1, \ldots, n$, and $\tilde{L}_n$ denote the associated controlled objects. Recall that ℓ and hence μ^* are non-atomic. From the definition of $\mathcal{W}$ and $\mathcal{W}_{\neq}$ in (12) and (31), respectively,

we have

$$\mathbb{E}\left[\mathcal{W}_{\neq}\left(\tilde{L}_n\right)\right] = \frac{1}{2}\mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} W\left(\tilde{\mathbf{X}}_i^n, \tilde{\mathbf{X}}_j^n\right)\right]$$
$$= \frac{n-1}{2n}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(\mathbf{x},\mathbf{y})\,\mu^*(d\mathbf{x})\,\mu^*(d\mathbf{y})$$
$$= \frac{n-1}{n}\mathcal{W}(\mu^*). \tag{59}$$

Define $\check{\mu}_n \doteq \frac{1}{n}\sum_{i=1}^n \tilde{\mu}_i^n = \mu^*$. Then, due to (58), the conditions of Lemma 24 hold for $\{(\tilde{L}_n, \check{\mu}_n)\}$. Together with Lemma 22, this shows that $\{\tilde{L}_n\}$ is tight in $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}_\psi(\mathbb{R}^d)$. When combined with the almost sure convergence $\tilde{L}_n \to \mu^*$, which holds due to Lemma 23 (or the Glivenko-Cantelli lemma), this implies convergence of $\tilde{L}_n$ to μ^* with respect to both d_w and d_ψ, as appropriate. Since f is bounded and continuous, $\lim_{n\to\infty}\mathbb{E}[f(\tilde{L}_n)] = f(\mu^*)$ by the dominated convergence theorem. The above observations, together with (59), the nonnegativity of $\mathcal{W}$ and (58) show that

$$\limsup_{n\to\infty}\inf_{\{\bar{\mu}_i^n\}}\mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{W}_{\neq}\left(\bar{L}_n\right) + \frac{1}{n}\sum_{i=1}^{n}\mathcal{R}\left(\bar{\mu}_i^n|e^{-V}\ell\right)\right]$$
$$\leq \limsup_{n\to\infty}\mathbb{E}\left[f\left(\tilde{L}_n\right) + \frac{n-1}{n}\mathcal{W}(\mu^*) + \frac{1}{n}\sum_{i=1}^{n}\mathcal{R}\left(\tilde{\mu}_i^n|e^{-V}\ell\right)\right]$$
$$\leq f(\mu^*) + \mathcal{W}(\mu^*) + \mathcal{R}\left(\mu^*|e^{-V}\ell\right)$$
$$\leq \inf_{\mu}\left[f(\mu) + \mathcal{I}(\mu)\right] + \epsilon.$$

Since ϵ is arbitrary, this implies the upper bound (37), which together with (36) and the discussion at the end of Section 3.1, completes the proof of Theorem 5.

4 The Case $\lim_{n\to\infty}\frac{\beta_n}{n} = \infty$

This section is devoted to the proof of Theorem 7. The structure of the proof is similar to that of the case with speed $\beta_n = n$. In view of Lemmas 19 and 20 and Theorem 1.2.3 in [9], it suffices to prove that for any bounded and continuous function f on S (where $S = \mathcal{P}(\mathbb{R}^d)$ or $S = \mathcal{P}_\psi(\mathbb{R}^d)$, as appropriate), as $n\to\infty$,

$$-\frac{1}{n}\log\mathbb{E}_{Q_n}\left[e^{-\beta_n f}\right] \to \inf_{\mu\in S}\left\{f(\mu) + \mathcal{J}_\star(\mu)\right\}. \tag{60}$$

4.1 Representation formula

As before, let $P_n^\star$ be the measure on $\mathbb{R}^{nd}$ defined by

$$P_n^\star(d\mathbf{x}_1,\ldots,d\mathbf{x}_n) \doteq e^{-\sum_{i=1}^n V(\mathbf{x}_i)}\ell(d\mathbf{x}_1)\cdots\ell(d\mathbf{x}_n),$$

which is a probability measure due to Assumption 32. We now introduce the functional $\mathcal{J}_{n,\neq} : \mathcal{P}(\mathbb{R}^d)\to(-\infty,\infty]$ given by

$$\mathcal{J}_{n,\neq}(\mu) \doteq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\left(\left(1-\frac{n}{\beta_n}\right)V(\mathbf{x}) + \left(1-\frac{n}{\beta_n}\right)V(\mathbf{y}) + W_{\neq}(\mathbf{x},\mathbf{y})\right)\mu(d\mathbf{x})\,\mu(d\mathbf{y}). \tag{61}$$

Note that $\mathcal{J}_{n,\neq}(\mu)$ is bounded below for all sufficiently large n due to Assumption 61 and the fact that $\beta_n/n \to \infty$. Also, recalling the definitions of L_n and H_n from (2) and (1), observe that

$$\mathcal{J}_{n,\neq} \circ L_n = \left(\frac{1}{n} - \frac{1}{\beta_n}\right) \sum_{i=1}^{n} V(\mathbf{x}_i) + \frac{1}{2n^2} \sum_{i,j=1}^{n} W_{\neq}(\mathbf{x}_i, \mathbf{x}_j) \tag{62}$$

Let f be a measurable function on $\mathcal{P}(\mathbb{R}^d)$ (or on $\mathcal{P}_\psi(\mathbb{R}^d)$) that is bounded below (in particular f could be bounded and continuous). Then by (62) and the definition of P_n^*, we have

$$-\frac{1}{\beta_n} \log \mathbb{E}_{Q_n}\left[e^{-\beta_n f}\right] = -\frac{1}{\beta_n} \log \mathbb{E}_{P_n}\left[e^{-\beta_n f \circ L_n}\right] = -\frac{1}{\beta_n} \log \mathbb{E}_{P_n^\star}\left[\frac{1}{Z_n} e^{-\beta_n \left(f + \mathcal{J}_{n,\neq}\right) \circ L_n}\right], \tag{63}$$

where Z_n is the normalization constant defined in (4).

Using the same notation and arguments as in Section 3.1, the following representations are valid. Since the function $(\mathbf{x}, \mathbf{y}) \mapsto \left(1 - \frac{n}{\beta_n}\right) V(\mathbf{x}) + \left(1 - \frac{n}{\beta_n}\right) V(\mathbf{y}) + W_{\neq}(\mathbf{x}, \mathbf{y})$ is measurable and bounded from below, we can apply Proposition 4.5.1 in [9] to $f(L_n(\boldsymbol{x}^n; \cdot)) + \mathcal{J}_{n,\neq}(L_n(\boldsymbol{x}^n; \cdot))$, for any function f on $\mathcal{P}(\mathbb{R}^d)$ (or $\mathcal{P}_\psi(\mathbb{R}^d)$), such that $f \circ L_n$ is measurable in $\mathbb{R}^{nd}$ and bounded from below. This includes all continuous and bounded functions on $\mathcal{P}(\mathbb{R}^d)$ or $\mathcal{P}_\psi(\mathbb{R}^d)$, and we obtain

$$-\frac{1}{\beta_n} \log \mathbb{E}_{P_n^\star}\left[e^{-\beta_n \left(f + \mathcal{J}_{n,\neq}\right) \circ L_n)}\right] = \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{J}_{n,\neq}\left(\bar{L}_n\right) + \frac{1}{\beta_n} \sum_{i=1}^{n} \mathcal{R}\left(\bar{\mu}_i^n | e^{-V} \ell\right)\right]. \tag{64}$$

Setting $f = 0$ in the last display, we have

$$\begin{aligned} -\frac{1}{\beta_n} \log\left(Z_n\right) &= -\frac{1}{\beta_n} \log \mathbb{E}_{P_n^\star}\left[e^{-\beta_n \mathcal{J}_{n,\neq} \circ L_n}\right] \\ &= \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[\mathcal{J}_{n,\neq}\left(\bar{L}_n\right) + \frac{1}{\beta_n} \sum_{i=1}^{n} \mathcal{R}\left(\bar{\mu}_i^n | e^{-V} \ell\right)\right]. \end{aligned} \tag{65}$$

As before, to establish Theorem 7, in view of (64), (65) and (63), it suffices to establish the lower bound

$$\liminf_{n\to\infty} \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{J}_{n,\neq}\left(\bar{L}_n\right) + \frac{1}{\beta_n} \sum_{i=1}^{n} \mathcal{R}\left(\bar{\mu}_i^n | e^{-V} \ell\right)\right] \geq \inf_{\mu} \left[f(\mu) + \mathcal{J}(\mu)\right], \tag{66}$$

and the upper bound

$$\limsup_{n\to\infty} \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{J}_{n,\neq}\left(\bar{L}_n\right) + \frac{1}{\beta_n} \sum_{i=1}^{n} \mathcal{R}\left(\bar{\mu}_i^n | e^{-V} \ell\right)\right] \leq \inf_{\mu} \left[f(\mu) + \mathcal{J}(\mu)\right], \tag{67}$$

for all bounded and continuous functions f (with respect to the corresponding topologies). The lower and upper bounds are established in Section 4.3 and Section 4.4, respectively, with preliminary results on the controls first established in Section 4.2.

4.2 Tightness of controls

We first make an observation that simplifies the proof of the lower bound.

Remark 8. *In proving the lower bound* (66)*, without loss of generality we can assume there exists* $C_1 \in \mathbb{R}$ *such that*

$$\sup_{n\in\mathbb{N}} \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[\mathcal{J}_{n,\neq}\left(\bar{L}_n\right) + \frac{1}{\beta_n}\sum_{i=1}^{n} \mathcal{R}\left(\bar{\mu}_i^n | e^{-V}\ell\right)\right] < C_1. \tag{68}$$

If this were not true, we could restrict to a subsequence that has such a property. For any subsequence for which the left-hand side of (68) *is infinite, the lower bound* (66) *is satisfied by default. Since* $\mathcal{R} \geq 0$ *it is therefore possible to restrict to controls such that the associated sequence of controlled empirical measures satisfies*

$$\sup_n \mathbb{E}\left[\mathcal{J}_{n,\neq}\left(\bar{L}_n\right)\right] < \infty. \tag{69}$$

Let $\mathcal{V} : \mathcal{P}(\mathbb{R}^d) \to (-\infty, \infty]$, be given by

$$\mathcal{V}(\mu) \doteq \int_{\mathbb{R}^d} V(\mathbf{x})\, \mu(d\mathbf{x}), \tag{70}$$

and note that $\mathcal{V}$ is well defined due to Assumption 3 and Assumption 61.

Lemma 25. *Let* $\{\bar{\mu}_i^n\}$ *be a sequence of controls such that the associated controlled empirical measures* $\{\bar{L}_n\}$ *satisfy* (69)*. Assume also that* V *and* W *satisfy Assumptions 3,61 and 62. Then* $\{\bar{L}_n\}$ *is tight in* $\mathcal{P}(\mathbb{R}^d)$*. If Assumption* 64 *is also satisfied with respect to some* ψ *then* $\{\bar{L}_n\}$ *is tight on* $\mathcal{P}_\psi(\mathbb{R}^d)$*.*

Proof. Let $\epsilon_1 > 0$ be as in Assumption 61. We observe that $\mathbb{E}\left[\mathcal{J}_{n,\neq}(\bar{L}_n)\right]$ is equal to

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{i,j=1,j\neq i}^{n} W\left(\bar{\mathbf{X}}_i^n, \bar{\mathbf{X}}_j^n\right) + 2\left(1 - \frac{n}{\beta_n}\right)\frac{1}{n}\sum_{i=1}^{n} V(\bar{\mathbf{X}}_i^n)\right] \\
&= \frac{1}{2}\mathbb{E}\Bigg[\frac{1-\epsilon_1}{2n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left(W\left(\bar{\mathbf{X}}_i^n, \bar{\mathbf{X}}_j^n\right) + (V(\bar{\mathbf{X}}_i^n) + V(\bar{\mathbf{X}}_j^n))\right) \\
&\qquad + \frac{1+\epsilon_1}{2n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left(W\left(\bar{\mathbf{X}}_i^n, \bar{\mathbf{X}}_j^n\right) + \epsilon_1(V(\bar{\mathbf{X}}_i^n) + V(\bar{\mathbf{X}}_j^n))\right) \\
&\qquad + 2\left(\frac{1}{n} - \frac{(1+\epsilon_1^2)(n-1)}{4n^2} - \frac{1}{\beta_n}\right)\sum_{i=1}^{n} V(\bar{\mathbf{X}}_i^n)\Bigg].
\end{aligned}$$

For large enough n, Assumption 61 and the fact that $n/\beta_n \to 0$ as $n \to \infty$ imply that the last two summands are bounded from below by a fixed constant $C' \in \mathbb{R}$. Therefore, we have

$$\mathbb{E}\left[\mathcal{J}_{n,\neq}\left(\bar{L}_n\right)\right] \geq \frac{1}{2}\mathbb{E}\left[\frac{1-\epsilon_1}{4n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left(W\left(\bar{\mathbf{X}}_i^n, \bar{\mathbf{X}}_j^n\right) + (V(\bar{\mathbf{X}}_i^n) + V(\bar{\mathbf{X}}_j^n))\right)\right] + C'.$$

Recalling the definition of the random probability measure $\bar{\mu}_{ij}^n$ in (42) and using the same argument as the one that led to (47), we obtain

$$\begin{aligned}
\mathbb{E}\left[\mathcal{J}_{n,\neq}\left(\bar{L}_n\right)\right] &\geq \mathbb{E}\left[\mathcal{J}_{n,\neq}\left(\bar{L}_n\right)\right] \\
&\geq \frac{1}{2}\mathbb{E}\left[\frac{1-\epsilon_1}{2n^2}\sum_{i=1,i\neq j}^{n}\int_{\mathbb{R}^d\times\mathbb{R}^d}\left(W(\mathbf{x}_i,\mathbf{x}_j) + (V(\mathbf{x}_i) + V(\mathbf{x}_j))\bar{\mu}_{ij}^n(d\mathbf{x}_i d\mathbf{x}_j)\right)\right] + C', \\
&= \frac{1}{2}\mathbb{E}\left[\frac{1-\epsilon_1}{2n^2}\sum_{i=1,i\neq j}^{n}\mathfrak{J}(\bar{\mu}_{ij}^n)\right],
\end{aligned}$$

where $\mathfrak{J}$ is the functional defined in (28). Now, let $\hat{\mu}^n$ and $\hat{\mu}^{2,n}$ be defined as in (40) and (48), respectively. By the linearity of $\mathfrak{J}$, we obtain

$$\mathbb{E}[\mathfrak{J}(\hat{\mu}^{2,n})] \leq \frac{(1-\epsilon_1)(n-1)}{2n}\mathbb{E}[\mathcal{J}_{\neq,n}(\bar{L}_n)] - C' \tag{71}$$

When Assumption 62 is satisfied, using (52), it follows that there exists a function γ with $\lim_{s\to\infty}\gamma(s) = \infty$ such that

$$\mathbb{E}\left[\int_{\mathbb{R}^d}\gamma\left(||\mathbf{x}||\right)\hat{\mu}^n\left(d\mathbf{x}\right)\right] \leq \mathbb{E}\left[\mathfrak{J}\left(\hat{\mu}^{2,n}\right)\right] ..$$

Similarly, under Assumption 64, again using (52), there exists a superlinear function ϕ such that

$$\mathbb{E}\left[\int_{\mathbb{R}^d}\phi\left(\psi(\mathbf{x})\right)\hat{\mu}^n\left(d\mathbf{x}\right)\right] \leq \mathbb{E}\left[\mathfrak{J}\left(\hat{\mu}^{2,n}\right)\right]$$

However, Lemma 15 and Lemma 17 show that both $\int_{\mathbb{R}^d}\gamma\left(\|\mathbf{x}\|\right)\mu\left(d\mathbf{x}\right)$, and $\Phi(\mu) = \int_{\mathbb{R}^d}\phi\left(\psi\left(\mathbf{x}\right)\right)\mu\left(d\mathbf{x}\right)$, are tightness functions on $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}_\psi\left(\mathbb{R}^d\right)$, respectively. Therefore, the last three displays and the uniform bound (69) on $\mathbb{E}\left[\mathcal{J}_{n,\neq}\left(\bar{L}_n\right)\right]$ imply that $\{\hat{\mu}^n\}$ is tight. Finally, by establishing (41) for the weak topology and (53) for the topology generated by ψ, we see that $\{\bar{L}_n\}$ is also tight. □

Remark 9. *Using Remark 8 and Lemma 25, while proving the lower bound we can restrict to controls such that* $\left\{\left(\bar{L}_n, \hat{\mu}_n\right), n \in \mathbb{N}\right\}$ *is tight on* $\mathcal{P}(\mathbb{R}^d)$ *or* $\mathcal{P}_\psi(\mathbb{R}^d)$ *as appropriate.*

4.3 Proof of the lower bound

For the proof of the lower bound we use some auxiliary functionals on $\mathcal{P}(\mathbb{R}^d)$:

$$\mathcal{J}^M\left(\mu\right) \doteq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\left(V\left(\mathbf{x}\right)+V\left(\mathbf{y}\right)+W^M\left(\mathbf{x},\mathbf{y}\right)\right)\mu\left(d\mathbf{x}\right)\mu\left(d\mathbf{y}\right),$$

$$\mathcal{J}_n^M\left(\mu\right) \doteq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\left(\left(1-\frac{n}{\beta_n}\right)V\left(\mathbf{x}\right)+\left(1-\frac{n}{\beta_n}\right)V\left(\mathbf{y}\right)+W^M\left(\mathbf{x},\mathbf{y}\right)\right)\mu\left(d\mathbf{x}\right)\mu\left(d\mathbf{y}\right),$$

$$\mathcal{J}_{n,\neq}^M\left(\mu\right) \doteq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\left(\left(1-\frac{n}{\beta_n}\right)V\left(\mathbf{x}\right)+\left(1-\frac{n}{\beta_n}\right)V\left(\mathbf{y}\right)+W_{\neq}^M\left(\mathbf{x},\mathbf{y}\right)\right)\mu\left(d\mathbf{x}\right)\mu\left(d\mathbf{y}\right),$$

where for a function F on $\mathbb{R}^{d'}$ and $M < \infty$ we define $F^M(\mathbf{z}) \doteq \min\{F(\mathbf{z}), M\}$. These integrals are well defined for sufficiently large n because of Assumption 3 and Assumptions 61-62.

Since $W \geq W^M$, for every M and $n \in \mathbb{N}$,

$$\begin{aligned}\inf_{\{\bar{\mu}_i^n\}}\mathbb{E}\left[f\left(\bar{L}_n\right)+\mathcal{J}_{n,\neq}\left(\bar{L}_n\right)+\frac{1}{\beta_n}\sum_{i=1}^n\mathcal{R}\left(\bar{\mu}_i^n|e^{-V}\ell\right)\right] \\ \geq \inf_{\{\bar{\mu}_i^n\}}\mathbb{E}\left[f\left(\bar{L}_n\right)+\mathcal{J}_{n,\neq}^M\left(\bar{L}_n\right)+\frac{1}{\beta_n}\sum_{i=1}^n\mathcal{R}\left(\bar{\mu}_i^n|e^{-V}\ell\right)\right] ..\end{aligned} \tag{72}$$

Let $\epsilon > 0$ and $\{\tilde{\mu}_i^n\}$ be such that

$$\begin{aligned} C' &> \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{J}_{n,\neq}\left(\bar{L}_n\right) + \frac{1}{\beta_n}\sum_{i=1}^n \mathcal{R}\left(\bar{\mu}_i^n | e^{-V}\ell\right)\right] + \epsilon \\ &\geq \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{J}_{n,\neq}\left(\tilde{L}_n\right) + \frac{1}{\beta_n}\sum_{i=1}^n \mathcal{R}\left(\tilde{\mu}_i^n | e^{-V}\ell\right)\right] \\ &\geq \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{J}_n^M\left(\tilde{L}_n\right) - \frac{M}{n} + \frac{1}{\beta_n}\sum_{i=1}^n \mathcal{R}\left(\tilde{\mu}_i^n | e^{-V}\ell\right)\right], \end{aligned}$$

where C' is a finite upper bound, whose existence is a result of Remark 8 and the boundedness of f, and the last inequality follows from (72) and the fact $\tilde{L}_n\left(d\mathbf{x}\right)\tilde{L}_n\left(d\mathbf{y}\right)$ puts mass $1/n$ on the diagonal $\mathbf{x} = \mathbf{y}$.

Owing to tightness (see Lemma 25) we can extract a further subsequence of $\{(\tilde{L}_n, \hat{\mu}_n)\}$, which (with some abuse of notation) we denote again by $\{(\tilde{L}_n, \hat{\mu}_n)\}$, for which $\hat{\mu}_n \doteq \frac{1}{n}\sum_{i=1}^n \tilde{\mu}_i^n$, that converges weakly to some limit $(\tilde{L}, \hat{\mu})$. Let M_n be a sequence that goes to infinity such that $\lim_{n\to\infty}\frac{M_n}{n} = 0$ and let $m \in \mathbb{N}$. Also, recall the constant ϵ_1 from Assumption 61. By Fatou's lemma, the nonnegativity of $\mathcal{R}(\cdot|e^{-V})$, the definition of $\mathcal{V}$ in (70), and the fact that $n/\beta_n \to 0$, we have

$$\begin{aligned} &\liminf_{n\to\infty} \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{J}_n^{M_n}\left(\tilde{L}_n\right) - \frac{M_n}{n} + \frac{1}{\beta_n}\sum_{i=1}^n \mathcal{R}\left(\tilde{\mu}_i^n | e^{-V}\ell\right)\right] \\ &\quad\geq \liminf_{n\to\infty} \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{J}_n^{M_m}\left(\tilde{L}_n\right)\right] \\ &\quad\geq \liminf_{n\to\infty} \mathbb{E}\Bigg[f\left(\tilde{L}_n\right) + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\left(\epsilon_1\left(V\left(\mathbf{x}\right) + V\left(\mathbf{y}\right)\right) + W^{M_m}\left(\mathbf{x},\mathbf{y}\right)\right)\tilde{L}_n\left(d\mathbf{x}\right)\tilde{L}_n\left(d\mathbf{y}\right) \\ &\qquad\qquad + \left(1 - \frac{n}{\beta_n} - \epsilon_1\right)\mathcal{V}\left(\tilde{L}_n\right)\Bigg] \\ &\quad\geq \mathbb{E}\Bigg[f(\tilde{L}) + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\left(\epsilon_1\left(V\left(\mathbf{x}\right) + V\left(\mathbf{y}\right)\right) + W^{M_m}\left(\mathbf{x},\mathbf{y}\right)\right)\tilde{L}\left(d\mathbf{x}\right)\tilde{L}\left(d\mathbf{y}\right) \\ &\qquad\qquad + \left(1 - \epsilon_1\right)\mathcal{V}\left(\tilde{L}\right)\Bigg] \\ &\quad= \mathbb{E}\left[f\left(\tilde{L}\right) + \mathcal{J}^{M_m}\left(\tilde{L}\right)\right]. \end{aligned}$$

Since the above inequality holds for arbitrary m, using the monotone convergence theorem

$$\begin{aligned} &\liminf_{n\to\infty} \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{J}_n^{M_n}\left(\tilde{L}_n\right) - \frac{M_n}{n} + \frac{1}{\beta_n}\sum_{i=1}^n \mathcal{R}\left(\tilde{\mu}_i^n | e^{-V}\ell\right)\right] \\ &\qquad\geq \mathbb{E}\left[f\left(\tilde{L}\right) + \mathcal{J}\left(\tilde{L}\right)\right] \geq \inf_{\mu}\left[f\left(\mu\right) + \mathcal{J}\left(\mu\right)\right]. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, this establishes (66).

4.4 Proof of the upper bound

We start by making an observation.

Remark 10. *Under Assumption 31, there exists at least one probability measure μ such that $J(\mu) < \infty$ (e.g., $\mu \doteq \frac{\ell_{|B}}{\ell(B)}$ where B is a bounded subset of A, defined in Assumption 31, with $0 < \ell(B) < \infty$). Also, in Assumption 63, it follows without any loss of generality, that each μ_n can be assumed to have a bounded density with respect to $e^{-V}\ell$. The argument is as follows. First, by direct application of Assumption 63, we can assume that μ_n has a density with respect to ℓ. We can even assume that it has a density with respect to $e^{-V}\ell$, because otherwise $J(\mu_n) = \infty$. Let ρ_n be the density of μ_n with respect to $e^{-V}\ell$. We let*

$$\mu_n^M(A) \doteq \frac{\int_A \rho_n^M(\mathbf{x}) e^{-V}\ell(d\mathbf{x})}{\int_{\mathbb{R}} \rho_n^M(\mathbf{x}) e^{-V}\ell(d\mathbf{x})}.$$

Since ρ_n^M is increasing with respect to M, and the map $(\mathbf{x},\mathbf{y}) \mapsto W(\mathbf{x},\mathbf{y}) + V(\mathbf{x}) + V(\mathbf{y})$ is bounded from below, by an application of the monotone convergence theorem $\mathcal{J}(\mu_n^M) \to \mathcal{J}(\mu_n)$, and therefore we get the desired result.

Now, when a measure μ is an element of $\mathcal{P}_\psi(\mathbb{R}^d)$, the approximating sequence $\{\mu_n\}$ can be taken to satisfy $d_\psi(\mu_n, \mu) \to 0$. To see this, first assume that μ has compact support K. By the previous part of this remark, we can find a sequence $\{\mu_n\}$, with bounded densities with respect to $e^{-V}\ell$, that converges weakly to μ and satisfies $\mathcal{J}(\mu_n) \to \mathcal{J}(\mu)$. Exploiting the fact that the map $(\mathbf{x},\mathbf{y}) \mapsto W(\mathbf{x},\mathbf{y}) + V(\mathbf{x}) + V(\mathbf{y})$ is bounded from below, with $\tilde{\mu}_n(\cdot) \doteq \frac{\mu_n(\cdot \cap K)}{\mu_n(K)}$ an application of Fatou's lemma gives

$$\begin{aligned}
\mathcal{J}(\mu) &= \lim_{n\to\infty} \frac{1}{2} \int_{\mathbb{R}^d\times\mathbb{R}^d} \left(V(\mathbf{x}) + V(\mathbf{y}) + W(\mathbf{x},\mathbf{y})\right) \mu_n(d\mathbf{x})\, \mu_n(d\mathbf{y}) \\
&\geq \lim_{n\to\infty} \mu_n^2(K) \lim_{n\to\infty} \frac{1}{2} \int_{\mathbb{R}^d\times\mathbb{R}^d} \left(V(\mathbf{x}) + V(\mathbf{y}) + W(\mathbf{x},\mathbf{y})\right) \tilde{\mu}_n(d\mathbf{x})\, \tilde{\mu}_n(d\mathbf{y}) \\
&= \mathcal{J}(\mu).
\end{aligned}$$

Thus $\tilde{\mu}_n$ satisfies $\mathcal{J}(\tilde{\mu}_n) \to \mathcal{J}(\mu)$. Moreover, since all $\tilde{\mu}_n$ have the same compact support and converge to μ, and since ψ is continuous, $\tilde{\mu}_n$ converges to μ in the topology generated by ψ. Next consider an arbitrary $\mu \in \mathcal{P}_\psi(\mathbb{R}^d)$ and let $\nu_n(\cdot) \doteq \frac{\mu(\cdot \cap B(0,n))}{\mu(B(0,n))}$. By dominated convergence $\int_{B(0,n)} \psi(\mathbf{x})\, \nu_n(d\mathbf{x}) \to \int_{\mathbb{R}^d} \psi(\mathbf{x})\, \mu(d\mathbf{x})$, which shows that $d_\psi(\nu_n, \mu) \to 0$. Since we also have $\mathcal{J}(\nu_n) \to \mathcal{J}(\mu)$, the desired approximating measures can be found by combining the two approximations and using a diagonal argument.

Let $\epsilon > 0$ and let μ^* be such that

$$f(\mu^*) + \mathcal{J}(\mu^*) \leq \inf_\mu \left[f(\mu) + \mathcal{J}(\mu)\right] + \epsilon.$$

We will also assume that for μ^* we have $\mathcal{R}\left(\mu^* | e^{-V}\ell\right) < \infty$. We can make this claim because of Assumption 63 and Remark 10. Then let $\tilde{\mu}_i^n = \mu^*$ for all $n \in \mathbb{N}$ and $i \in \{1, ..., n\}$, and let the random variables $\tilde{\mathbf{X}}_i^n, 1 \leq i \leq n, n \in \mathbb{N}$, be iid with distribution μ^*. By Lemma 23, the weak limit of $\tilde{L}_n$ equals μ^*. Calculations very similar to those of (59) give

$$\begin{aligned}
&\mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{J}_{n,\neq}\left(\tilde{L}_n\right) + \frac{1}{\beta_n}\sum_{i=1}^n \mathcal{R}\left(\mu^* | e^{-V}\ell\right)\right] \\
&\quad = \mathbb{E}\left[f\left(\tilde{L}_n\right)\right] + \frac{n-1}{n}\mathcal{J}(\mu^*) + 2\left(\frac{1}{n} - \frac{n}{\beta_n}\right)\mathcal{V}(\mu^*) + \frac{n}{\beta_n}\mathcal{R}\left(\mu^* | e^{-V}\ell\right).
\end{aligned}$$

Thus, by the dominated convergence theorem, the quantity

$$\limsup_{n\to\infty}\left(\mathbb{E}\left[f\left(\tilde{L}_n\right)\right] + \frac{n}{n-1}\mathcal{J}(\mu^*) + 2\left(\frac{1}{n} - \frac{n}{\beta_n}\right)\mathcal{V}(\mu^*) + \frac{n}{\beta_n}\mathcal{R}\left(\mu^* | e^{-V}\ell\right)\right)$$

is equal to $f(\mu^*) + \mathcal{J}(\mu^*)$. Combining these inequalities we have

$$\begin{aligned}
&\limsup_{n\to\infty} \inf_{\{\bar{\mu}_i^n\}} \mathbb{E}\left[f\left(\bar{L}_n\right) + \mathcal{J}_{n,\neq}\left(\bar{L}_n\right) + \frac{1}{\beta_n}\sum_{i=1}^n \mathcal{R}\left(\bar{\mu}_i^n | e^{-V}\ell\right)\right] \\
&\quad \leq \limsup_{n\to\infty} \mathbb{E}\left[f\left(\tilde{L}_n\right) + \mathcal{J}_{n,\neq}\left(\tilde{L}_n\right) + \frac{1}{\beta_n}\sum_{i=1}^n \mathcal{R}\left(\mu^* | e^{-V}\ell\right)\right] \\
&\quad \leq \inf_{\mu}\left[f(\mu) + \mathcal{J}(\mu)\right] + \epsilon,
\end{aligned}$$

and since $\epsilon > 0$ is arbitrary, we obtain the upper bound (67), and thus the proof of Theorem 7 is complete.

A Proof of Lemma 11

The proof of Lemma 11 is based on two preliminary results, established in Lemma 26 and Lemma 27 below.

Lemma 26. *Let $\nu \in \mathcal{P}(\mathbb{R}^m)$ and let $\bar{\psi} : \mathbb{R}^m \to \mathbb{R}_+$ be measurable. Then*

$$\int_{\mathbb{R}^m} e^{\lambda\bar{\psi}(\mathbf{z})}\nu(d\mathbf{z}) < \infty \tag{73}$$

for all $\lambda < \infty$ if and only if there exists a convex, increasing and superlinear function $\bar{\phi} : \mathbb{R}_+ \to \mathbb{R}$ such that

$$\int_{\mathbb{R}^m} e^{\bar{\phi}\left(\bar{\psi}(\mathbf{z})\right)}\nu(d\mathbf{z}) < \infty. \tag{74}$$

Proof. ($\Rightarrow$) If (73) holds, for every $k \in \mathbb{N}$ we can find $M_k \in (0,\infty)$ such that

$$\int_{\{\mathbf{z}:\,\bar{\psi}(\mathbf{z})\geq M_k\}} e^{k\bar{\psi}(\mathbf{z})}\nu(d\mathbf{z}) < \frac{1}{2^k}.$$

Without loss of generality, we can assume $M_{k+1} \geq M_k$, and $\lim_{k\to\infty} M_k = \infty$. We then define $\bar{\phi}(s) = ks, s \in [M_k, M_{k+1}]$, and $\bar{\phi}(s) = M_1, s \in [0, M_1]$, which implies $\lim_{s\to\infty}\frac{\bar{\phi}(s)}{s} = \infty$ and also that $\bar{\phi}$ is convex and increasing. Finally, we have

$$\int_{\mathbb{R}^m} e^{\bar{\phi}\left(\bar{\psi}(\mathbf{z})\right)}\nu(d\mathbf{z}) \leq e^{M_1} + \sum_{k=1}^{\infty}\int_{\{\mathbf{z}:\,\bar{\psi}(\mathbf{z})\geq M_k\}} e^{k\bar{\psi}(\mathbf{z})}\nu(d\mathbf{z}) \leq e^{M_1} + \sum_{k=1}^{\infty}\frac{1}{2^k} < \infty.$$

($\Leftarrow$) Let $\bar{\phi}$ be as in the statement of the lemma. Since $\bar{\phi}$ satisfies $\lim_{s\to\infty}\frac{\bar{\phi}(s)}{s} = \infty$, for every $\lambda < \infty$ there exists $M_\lambda < \infty$ such that $\bar{\phi}(s) \geq \lambda s$ if $s \geq M_\lambda$. Then we have

$$\begin{aligned}
\int_{\mathbb{R}^m} e^{\lambda\bar{\psi}(\mathbf{z})}\nu(d\mathbf{z}) &= \int_{\mathbb{R}^m} 1_{\{\bar{\psi}(\mathbf{z})<M_\lambda\}} e^{\lambda\bar{\psi}(\mathbf{z})}\nu(d\mathbf{z}) + \int_{\mathbb{R}^m} 1_{\{\bar{\psi}(\mathbf{z})\geq M_\lambda\}} e^{\lambda\bar{\psi}(\mathbf{z})}\nu(d\mathbf{z}) \\
&\leq e^{\lambda M_\lambda} + \int_{\mathbb{R}^m} e^{\bar{\phi}\left(\bar{\psi}(\mathbf{z})\right)}\nu(d\mathbf{z}) \\
&< \infty.
\end{aligned}$$

□

Lemma 27. *Let $\nu \in \mathcal{P}(\mathbb{R}^m)$ and let $\bar{\psi} : \mathbb{R}^m \to \mathbb{R}_+$ be measurable. Then*

$$\int_{\mathbb{R}^m} e^{\lambda \bar{\psi}(\mathbf{z})} \nu(d\mathbf{z}) < \infty \tag{75}$$

for all $\lambda < \infty$ if and only if there exists a convex, increasing and superlinear function $\bar{\phi} : \mathbb{R}_+ \to \mathbb{R}$ and a constant $C < \infty$ such that for any $\mu \in \mathcal{P}(\mathbb{R}^m)$,

$$\int_{\mathbb{R}^m} \bar{\phi}\left(\bar{\psi}(\mathbf{z})\right) \mu(d\mathbf{z}) \leq \mathcal{R}(\mu|\nu) + C. \tag{76}$$

Proof. ($\Rightarrow$) First assume that (75) holds. Then by the previous lemma there exists a positive convex function $\bar{\phi} : \mathbb{R} \to \mathbb{R}$, with $\lim_{s\to\infty} \frac{\bar{\phi}(s)}{s} = \infty$ such that (74) holds. Since $-\bar{\phi} \leq 0$, by using Proposition 4.5.1 in [9] with $k = -\bar{\phi}$, we get

$$\sup_{\mu\in\mathcal{P}(\mathbb{R}^m):\mathcal{R}(\mu|\nu)<\infty} \left\{ \int_{\mathbb{R}^m} \bar{\phi}\left(\bar{\psi}(\mathbf{z})\right) \mu(d\mathbf{z}) - \mathcal{R}(\mu|\nu) \right\} = \log \int_{\mathbb{R}^m} e^{\bar{\phi}\left(\bar{\psi}(\mathbf{z})\right)} \nu(d\mathbf{z}) < \infty, \tag{77}$$

from which we obtain

$$\int_{\mathbb{R}^m} \bar{\phi}\left(\bar{\psi}(\mathbf{z})\right) \mu(d\mathbf{z}) \leq \mathcal{R}(\mu|\nu) + \log \int_{\mathbb{R}^m} e^{\bar{\phi}\left(\bar{\psi}(\mathbf{z})\right)} \nu(d\mathbf{z})$$

for all $\mu \in \mathcal{P}(\mathbb{R}^m)$ with $\mathcal{R}(\mu|\nu) < \infty$. Thus, (76) follows.

($\Leftarrow$) For the converse, if we assume that (76) is true, then we have

$$\sup_{\mu\in\mathcal{P}(\mathbb{R}^m)} \left\{ \int_{\mathbb{R}^m} \bar{\phi}\left(\bar{\psi}(\mathbf{z})\right) \mu(d\mathbf{z}) - \mathcal{R}(\mu|\nu) \right\} \leq C,$$

and (77) implies that $\log \int_{\mathbb{R}^m} e^{\bar{\phi}\left(\bar{\psi}(\mathbf{z})\right)} \nu(d\mathbf{z})$ is bounded, which proves (75). □

Proof of Lemma 11. Consider the probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ defined by

$$\nu(d\mathbf{x}d\mathbf{y}) = \frac{1}{Z} e^{-(V(\mathbf{x})+V(\mathbf{y})+W(\mathbf{x},\mathbf{y}))} \ell(d\mathbf{x})\ell(d\mathbf{y}),$$

where Z is the normalization constant that makes ν a probability measure; the finiteness of Z follows on setting $\lambda = 0$ in (25). Since ψ satisfies (25), we can apply Lemma 27 with $\bar{\psi}(\mathbf{x},\mathbf{y}) = \psi(\mathbf{x}) + \psi(\mathbf{y})$ to conclude that there exists a convex and increasing function $\bar{\phi} : \mathbb{R}_+ \mapsto \mathbb{R}$ with $\lim_{s\to\infty} \bar{\phi}(s)/s = \infty$ such that for any $\zeta \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \bar{\phi}\left(\psi(\mathbf{x}) + \psi(\mathbf{y})\right) \zeta(d\mathbf{x}d\mathbf{y}) \leq \mathcal{R}\left(\zeta | e^{-(V(\mathbf{x})+V(\mathbf{y})+W(\mathbf{x},\mathbf{y}))} \ell(d\mathbf{x})\ell(d\mathbf{y})/Z\right) + C. \tag{78}$$

We claim, and prove below, that for every ζ

$$\int_{\mathbb{R}^d} \bar{\phi}\left(\psi(\mathbf{x})\right) (\pi^i_{\#}\zeta)(d\mathbf{x}) \leq 2\mathfrak{W}(\zeta) + 2\mathcal{R}\left(\zeta | e^{-V}\ell \otimes e^{-V}\ell\right) + C + \log Z, \tag{79}$$

where recall from Definition 18 and Definition 21, that $\pi^i_{\#}\zeta$ represents the ith marginal of ζ. If the claim holds, then since $\bar{\phi}$ is increasing and since ψ and $\mathcal{R}$ are positive, for $i = 1, 2$, we have (79). Adding the inequality (79) for $i = 1$ and $i = 2$ we have

$$\int_{\mathbb{R}^d} \bar{\phi}\left(\psi(\mathbf{x})\right) (\pi^1_{\#}\zeta)(d\mathbf{x}) + \int_{\mathbb{R}^d} \bar{\phi}\left(\psi(\mathbf{x})\right) (\pi^2_{\#}\zeta)(d\mathbf{x}) \leq 4\mathfrak{W}(\zeta) + 4\mathcal{R}\left(\zeta | e^{-V}\ell \otimes e^{-V}\ell\right) + 2(C + \log Z).$$

If $\zeta \in \Pi(\mu,\mu)$ then $\pi^1_\# \zeta = \pi^2_\# \zeta = \mu$. Dividing both sides by 2, the assertion (26) of the lemma holds with $\phi \doteq [\bar{\phi} - C - \log Z]/2$.

We now turn to the proof of the claim (79). We can assume without loss of generality that $\zeta(d\mathbf{x}d\mathbf{y})$ has a density with respect to the measure $e^{-V}\ell \otimes e^{-V}\ell$, because otherwise (79) holds trivially, since W is bounded from below. Denoting this density (with some abuse of notation) by $\zeta(\mathbf{x},\mathbf{y})$, (78) then gives

$$\begin{aligned}
&\int_{\mathbb{R}^d\times\mathbb{R}^d} \bar{\phi}\left(\psi\left(\mathbf{x}\right)+\psi\left(\mathbf{y}\right)\right)\zeta\left(d\mathbf{x}d\mathbf{y}\right) \\
&\quad\leq \int_{\mathbb{R}^d\times\mathbb{R}^d} \zeta(\mathbf{x},\mathbf{y}) \log \frac{\zeta(\mathbf{x},\mathbf{y})}{e^{-W(\mathbf{x},\mathbf{y})}/Z} e^{-(V(\mathbf{x})+V(\mathbf{y}))}\ell(d\mathbf{x})\ell(d\mathbf{y}) + C \\
&\quad\leq \int_{\mathbb{R}^d\times\mathbb{R}^d} W(\mathbf{x},\mathbf{y})\zeta(\mathbf{x},\mathbf{y}) e^{-(V(\mathbf{x})+V(\mathbf{y}))}\ell(d\mathbf{x})\ell(d\mathbf{y}) \\
&\qquad + \int_{\mathbb{R}^d\times\mathbb{R}^d} \zeta(\mathbf{x},\mathbf{y}) \log \zeta(\mathbf{x},\mathbf{y}) e^{-(V(\mathbf{x})+V(\mathbf{y}))}\ell(d\mathbf{x})\ell(d\mathbf{y}) + \log Z + C.
\end{aligned}$$

Therefore, recalling the definition of $\mathfrak{W}$ in (9), we have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \bar{\phi}\left(\psi\left(\mathbf{x}\right)+\psi\left(\mathbf{y}\right)\right)\zeta\left(d\mathbf{x}d\mathbf{y}\right) \leq 2\mathfrak{W}\left(\zeta\right) + \mathcal{R}\left(\zeta | e^{-V}\ell \otimes e^{-V}\ell\right) + \log Z + C,$$

which completes the proof of the claim, and therefore the lemma. □

B Proof of Lemma 17

We first establish a preliminary result in Lemma 28 below. Let $B(0,r)$ denote the closed ball about 0 of radius r, and let $B^c(0,r)$ denote its complement.

Lemma 28. *Let $\psi, \mathcal{P}_\psi(\mathbb{R}^d)$, and d_ψ be defined as in (5)-(7). Then $d_\psi(\mu_n,\mu) \to 0$ as $n \to \infty$ if and only if*

$$d_w(\mu_n,\mu) \to 0 \text{ and } \lim_{r\to\infty} \sup_n \left\{ \int_{B^c(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) \right\} = 0. \tag{80}$$

Furthermore, the metric space $(\mathcal{P}_\psi(\mathbb{R}^d), d_\psi)$ is separable.

Proof. ($\Rightarrow$). Let $\mu_n, n \in \mathbb{N}$, $\mu \in \mathcal{P}_\psi(\mathbb{R}^d)$ be such that $d_\psi(\mu_n,\mu) \to 0$. Since $d_w(\mu_n,\mu) \leq d_\psi(\mu_n,\mu)$, this implies $d_w(\mu_n,\mu) \to 0$. Let $\epsilon > 0$. By the integrability of ψ there exists $r < \infty$ such that $\int_{B^c(0,r)} \psi(\mathbf{x})\mu(d\mathbf{x}) < \frac{\epsilon}{3}$, and also $\mu(\partial B(0,r)) = 0$. Hence, we have

$$\begin{aligned}
\int_{B^c(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) &= \int_{B^c(0,r)} \psi(\mathbf{x})(\mu_n(d\mathbf{x}) - \mu(d\mathbf{x})) + \int_{B^c(0,r)} \psi(\mathbf{x})\mu(d\mathbf{x}) \\
&\leq \left| \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu_n(d\mathbf{x}) - \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu(d\mathbf{x}) \right| \\
&\quad + \left| \int_{B(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) - \int_{B(0,r)} \psi(\mathbf{x})\mu(d\mathbf{x}) \right| + \frac{\epsilon}{3}.
\end{aligned} \tag{81}$$

From the definition of d_ψ in (7) and the nonnegativity of d_w, we can find $n_0 \in \mathbb{N}$ such that $\forall n > n_0$, we have

$$\left| \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu_n(d\mathbf{x}) - \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu(d\mathbf{x}) \right| < \frac{\epsilon}{3}.$$

Since $\mu(\partial B(0,r)) = 0$, the μ-measure of the discontinuity points of $\mathbf{x} \to \psi(\mathbf{x})1_{B(0,r)}(\mathbf{x})$ is zero. Since $\psi(\mathbf{x})$ can be extended outside of $B(0,r)$ to obtain a bounded and continuous function on $\mathbb{R}^d$, the fact that $d_w(\mu_n, \mu) \to 0$ implies that there exists $n_0' < \infty$ such that $\forall n \geq n_0'$,

$$\left| \int_{B(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) - \int_{B(0,r)} \psi(\mathbf{x})\mu(d\mathbf{x}) \right| < \frac{\epsilon}{3}. \tag{82}$$

Combining the above estimates for all terms in (81) we obtain

$$\sup_{n \geq \max\{n_0, n_0'\}} \left\{ \int_{B^c(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) \right\} < \epsilon.$$

Since ψ is integrable with respect to each μ_n, for all $n \leq \max\{n_0, n_0'\}$ we can find an $r_n < \infty$ such that $\int_{B^c(0,r_n)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) < \epsilon$. Taking $r' = \max\{r_1, ..., r_{\max\{n_0,n_0'\}}, r\}$ yields

$$\sup_n \left\{ \int_{B^c(0,r')} \psi(\mathbf{x})\mu_n(d\mathbf{x}) \right\} < \epsilon.$$

Since ϵ is arbitrary, the conclusion follows.

($\Leftarrow$) To prove the converse, let $\mu_n, n \in \mathbb{N}$, $\mu \in \mathcal{P}_\psi(\mathbb{R}^d)$, be such that (80) holds. For $\epsilon > 0$ there exists $r < \infty$ such that $\mu(\partial B(0,r)) = 0$ and

$$\sup_n \left\{ \int_{B^c(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) \right\} < \frac{\epsilon}{3} \quad \text{and} \quad \int_{B^c(0,r)} \psi(\mathbf{x})\mu(d\mathbf{x}) < \frac{\epsilon}{3},$$

where the latter inequality holds because $\mu \in \mathcal{P}_\psi$ implies that ψ is μ-integrable. Thus, we have

$$\begin{aligned} \left| \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu_n(d\mathbf{x}) - \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu(d\mathbf{x}) \right| &\leq \left| \int_{B(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) - \int_{B(0,r)} \psi(\mathbf{x})\mu(d\mathbf{x}) \right| \\ &\quad + \left| \int_{B^c(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) - \int_{B^c(0,r)} \psi(\mathbf{x})\mu(d\mathbf{x}) \right| \\ &\leq \left| \int_{B(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) - \int_{B(0,r)} \psi(\mathbf{x})\mu(d\mathbf{x}) \right| + \frac{2\epsilon}{3}. \end{aligned} \tag{83}$$

Since $d_w(\mu_n, \mu) \to 0$ and μ puts no mass on the set of discontinuities of the bounded function $\psi(\mathbf{x})1_{B(0,r)}(\mathbf{x})$, there exists $n_0' < \infty$ such that

$$\left| \int_{B(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x}) - \int_{B(0,r)} \psi(\mathbf{x})\mu(d\mathbf{x}) \right| < \frac{\epsilon}{3}, \qquad \forall n \geq n_0'.$$

Since ϵ is arbitrary, when substituted back into (83), this shows that

$$\lim_{n\to\infty} \left| \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu_n(d\mathbf{x}) - \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu(d\mathbf{x}) \right| = 0.$$

We now turn to the proof that $\mathcal{P}_\psi(\mathbb{R}^d)$ is separable. Let $\{\mathbf{x_n}\}$ be a countable dense subset of $\mathbb{R}^d$, and define

$$\mathcal{A} \doteq \left\{ \sum_{i=1}^N c_n \delta_{\mathbf{x_n}} : c_n \in \mathbb{Q}_+, n = 1, \ldots, \mathbb{N}, \sum_{n=1}^N c_n = 1, \sum_{n=1}^N c_n \psi(\mathbf{x_n}) < \infty, N \in \mathbb{N} \right\},$$

where $\mathbb{Q}_+$ is the set of nonnegative rational numbers, and observe that $\mathcal{A}$ is a countable subset of $\mathcal{P}_\psi$. We nows show that $\mathcal{A}$ is dense in $\mathcal{P}_\psi$. Fix $\mu \in \mathcal{P}_\psi$ and $\varepsilon > 0$. Also, consider the space $\mathbb{F}$ of bounded, Lipschitz continuous functions on $\mathbb{R}^d$, equipped with the norm

$$||f||_{BL} \doteq \max\left(\sup_{\mathbf{x},\mathbf{y}\in\mathbb{R}^d,\mathbf{x}\neq\mathbf{y}} \frac{|f(\mathbf{x}) - f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|}, 2\sup_{x\in\mathbb{R}^d}|f(\mathbf{x})|\right),$$

and let $\mathbb{F}_1$ be the subspace of functions with $||f||_{BL} \leq 1$. Then consider the metric on $\mathcal{P}(\mathbb{R}^d)$ given by

$$d_{BL}(\mu,\nu) \doteq \sup_{f\in\mathbb{F}_1}\left|\int_{\mathbb{R}^d} f(\mathbf{x})\mu(\mathbf{dx}) - \int_{\mathbb{R}^d} f(\mathbf{x})\nu(\mathbf{dx})\right|.$$

In view of the definition of d_ψ in (7) and the fact that there exists a constant $C < \infty$ such that $d_w(\mu,\nu) \leq 3\sqrt{d_{BL}(\mu,\nu)}$ (see [7, p. 396]), it suffices to show that there exists $\nu \in \mathcal{A}$ such that

$$\sup_{f\in\mathbb{F}_1\cup\{\psi\}}\left|\int_{\mathbb{R}^d} f(x)\mu(dx) - \int_{\mathbb{R}^d} f(x)\nu(dx)\right| \leq \varepsilon.. \tag{84}$$

Recalling that ψ is continuous, for each $n \in \mathbb{N}$, choose $r_n \in (0, \varepsilon/2)$ such that

$$\sup_{\mathbf{x}\in B_{r_n}(\mathbf{x_n})} |\psi(\mathbf{x}) - \psi(\mathbf{x_n})| \leq \frac{\varepsilon}{2}, \tag{85}$$

and note that then we also have

$$\sup_{\mathbf{x}\in B_{r_n}(\mathbf{x_n})} |f(\mathbf{x}) - f(\mathbf{x_n})| \leq r_n \leq \frac{\varepsilon}{2}, \quad f \in \mathbb{F}_1.. \tag{86}$$

Now, define $\tilde{B}_n \doteq B_{r_n}(x_n) \setminus \cup_{k=1}^{n-1} B_{r_k}(x_k)$ and $b_n \doteq \mu(\tilde{B}_n)$. Clearly, $\{\tilde{B}_n\}_{n\in\mathbb{N}}$ forms a disjoint partition of $\mathbb{R}^d$ and hence, $\sum_{n=1}^\infty b_n = 1$. Moreover, by (85) and (86) we have for all $f \in \mathbb{F}_1 \cup \{\psi\}$,

$$\left|\sum_{n=1}^\infty b_n f(\mathbf{x_n}) - \int_{\mathbb{R}^d} f(\mathbf{x})\mu(d\mathbf{x})\right| \leq \sum_{n=1}^\infty b_n \sup_{x\in\tilde{B}_n} |f(\mathbf{x_n}) - f(\mathbf{x})| \leq \frac{\varepsilon}{2}. \tag{87}$$

We can assume without loss of generality that ψ is uniformly bounded from below away from zero. Since $\int_{\mathbb{R}^d} \psi(\mathbf{x})\mu(\mathbf{dx})$ is finite, this implies $\sum_{n=1}^\infty b_n\psi(\mathbf{x_n}) < \infty$, and hence there exists $N \in \mathbb{N}$ such that

$$\sum_{n=N+1}^\infty b_n \leq \frac{\varepsilon}{8(\psi(\mathbf{x_1}) \vee 1)} \quad \text{and} \quad \sum_{n=N+1}^\infty b_n\psi(\mathbf{x_n}) \leq \frac{\varepsilon}{8}. \tag{88}$$

Now, for $n = 2, \ldots, N$, choose $c_n \in \mathbb{Q}_+$ such that

$$0 \leq b_n - c_n \leq \left(\frac{b_n}{\max(|\psi(\mathbf{x_1}) + \psi(\mathbf{x_n})|, |\mathbf{x_n} - \mathbf{x_1}|)}\right)\frac{\varepsilon}{4}, \tag{89}$$

and set

$$c_1 \doteq b_1 + \sum_{n=2}^N (b_n - c_n) + \sum_{n=N+1}^\infty b_n.$$

Observe that $\sum_{n=1}^{N} c_n = \sum_{n=1}^{\infty} b_n = 1$, and hence, c_1 also lies in $\mathbb{Q}_+$. Set $\nu \doteq \sum_{n=1}^{N} c_n \delta_{x_n}$. Then, for $f \in \mathbb{F}_1 \cup \{\psi\}$, using (89) and (88), we have

$$\begin{aligned}\left|\int_{\mathbb{R}^d} f(x)\nu(dx) - \sum_{n=1}^{\infty} b_n f(x_n)\right| &= \left|\sum_{n=1}^{N} c_n f(x_n) - \sum_{n=1}^{\infty} b_n f(x_n)\right| \\ &\leq \sum_{n=2}^{N}(b_n - c_n)|f(x_n) - f(x_1)| + \sum_{n=N+1}^{\infty} b_n |f(x_1) - f(x_n)| \\ &\leq \frac{\varepsilon}{4} + |f(x_1)| \sum_{n=N+1}^{\infty} b_n + \sum_{n=N+1}^{\infty} b_n |f(x_n)| \\ &\leq \frac{\varepsilon}{2}.\end{aligned}$$

When combined with (87) this establishes the desired inequality (84). □

Proof of Lemma 17. Let $C < \infty$ and let $\{\mu_n\} \subset \mathcal{P}_\psi(\mathbb{R}^d)$ be a sequence such that $\Phi(\mu_n) \leq C$ for all n. Now $\lim_{c\to\infty} \inf_{\mathbf{x}:\|\mathbf{x}\|=c} \phi(\psi(\mathbf{x})) = \infty$ because $\lim_{c\to\infty} \inf_{\mathbf{x}:\|\mathbf{x}\|=c} \psi(\mathbf{x}) = \infty$ and $\lim_{s\to\infty} \frac{\phi(s)}{s} = \infty$. Hence, by Lemma 14 with $g = \phi \circ \psi$, the sequence $\{\mu_n\}$ is tight in the weak topology, and we have

$$\begin{aligned}&\lim_{r\to\infty} \sup_n \left\{\int_{B^c(0,r)} \psi(\mathbf{x})\mu_n(d\mathbf{x})\right\} \\ &= \lim_{r\to\infty} \sup_n \left\{\int_{B^c(0,r)} \phi(\psi(\mathbf{x}))\frac{\psi(\mathbf{x})}{\phi(\psi(\mathbf{x}))}\mu_n(d\mathbf{x})\right\} \\ &\leq \lim_{r\to\infty} \sup_n \left\{\left(\sup_{x\in B^c(0,r)} \frac{\psi(\mathbf{x})}{\phi(\psi(\mathbf{x}))}\right)\int_{B^c(0,r)} \phi(\psi(\mathbf{x}))\mu_n(d\mathbf{x})\right\} \\ &\leq C \lim_{r\to\infty} \sup_{\mathbf{x}\in B^c(0,r)} \frac{\psi(\mathbf{x})}{\phi(\psi(\mathbf{x}))} \\ &= 0.\end{aligned}$$

Thus, by the first assertion of Lemma 28, $\{\mu_n\}$ is tight in $\mathcal{P}_\psi(\mathbb{R}^d)$. □

C Tightness Results

Proof of Lemma 23. Since $\mathbb{R}^d$ is a Polish space, to verify weak convergence of a sequence of measures in $\mathcal{P}(\mathbb{R}^d)$ it suffices to consider convergence of integrals with respect to the measures of functions f that are uniformly continuous. We use the fact [15, Lemma 3.1.4] that there is an equivalent metric m on $\mathbb{R}^d$, such that if $\mathcal{U}_b(\mathbb{R}^d, m)$ is the space of bounded uniformly continuous functions with respect to this metric, then there is a countable dense subset $\{f_m\}_{m\in\mathbb{N}} \subset \mathcal{U}_b(\mathbb{R}^d, m)$. Define $K_m \doteq \sup_{\mathbf{x}\in\mathbb{R}^d} |f_m(\mathbf{x})|$ and $\Delta_{m,i}^n \doteq f_m\left(\bar{\mathbf{X}}_i^n\right) - \int_{\mathbb{R}^d} f_m(\mathbf{x})\,\bar{\mu}_i^n(d\mathbf{x})$. For any $\varepsilon > 0$, Chebyshev's inequality shows that

$$\begin{aligned}&\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d} f_m(\mathbf{x})\,\delta_{\bar{\mathbf{X}}_i^n}(d\mathbf{x}) - \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d} f_m(x)\,\bar{\mu}_i^n(dx)\right| > \varepsilon\right] \\ &\qquad \leq \frac{1}{\varepsilon^2}\mathbb{E}\left[\frac{1}{n^2}\sum_{i,j=1}^{n}\Delta_{m,i}^n\Delta_{m,j}^n\right].\end{aligned}$$

Let $\mathcal{F}_j^n = \sigma(\bar{\mathbf{X}}_i^n, i = 1, \ldots, j)$. As we show below, by a standard conditioning argument, the off-diagonal terms vanish: for $i > j$,

$$\mathbb{E}\left[\Delta_{m,i}^n \Delta_{m,j}^n\right] = \mathbb{E}\left[\mathbb{E}\left[\Delta_{m,i}^n \Delta_{m,j}^n \middle| \mathcal{F}_i^n\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\Delta_{m,i}^n \middle| \mathcal{F}_i^n\right] \Delta_{m,j}^n\right] = 0.$$

Since $|\Delta_{m,i}^n| \leq 2K_m$,

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d} f_m(x)\, \delta_{\bar{X}_i^n}(dx) - \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d} f_m(\mathbf{x})\, \bar{\mu}_i^n(d\mathbf{x})\right| > \varepsilon\right] \leq \frac{4K_m^2}{n\varepsilon^2}.$$

Since $(\bar{L}^n, \hat{\mu}^n) \Rightarrow \left(\bar{L}, \hat{\mu}\right)$ and $\varepsilon > 0$ is arbitrary, by Fatou's lemma,

$$\mathbb{P}\left[\int_{\mathbb{R}^d} f_m(x)\, \bar{L}(dx) = \int_{\mathbb{R}^d} f_m(x)\, \hat{\mu}(dx)\right] = 1.$$

Now use the property that $\{f_m, m \in \mathbb{N}\}$ is countable and dense to conclude that $\bar{L} = \hat{\mu}$ a.s. □

References

[1] Luigi Ambrosio, Nicola Gigli, Giuseppe Savare, and G. Savaré. Gradient Flows in Metric Spaces and in the Spaces of Probability Measures. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, Lectures in Mathematics. ETH Zürich, page 334, Basel, 2008. Birkhauser.

[2] G Ben Arous and A Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probability Theory and Related Fields*, 108(4):517–542, aug 1997.

[3] Gérard Ben Arous and Ofer Zeitouni. Large deviations from the circular law. *ESAIM: Probability and Statistics*, 2:123–134, 1998.

[4] Patrick Billingsley. Probability and Measure (2nd Edition). page 636, 1986.

[5] Djalil Chafaï, Nathael Gozlan, and Pierre-André Zitt. First-order global asymptotics for confined particles with singular pair repulsion. *The Annals of Applied Probability*, 24(6):2371–2413, dec 2014.

[6] A Dembo and O Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2nd edition, 1987.

[7] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2002.

[8] P Dupuis. Representations and weak convergence methods for the analysis and approximation of rare events, may 2013.

[9] P Dupuis and R S Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, 1997.

[10] Crispin Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences*, volume 2010. Springer, 2010.

[11] Adrien Hardy. A note on large deviations for 2D Coulomb gas with weakly confining potential. *Electronic Communications in Probability*, 17:1–12, may 2012.

[12] Mauro Mariani. A Gamma-convergence approach to large deviations. *arXiv preprint arXiv:1204.0640*, apr 2012.

[13] D Petz and F Hiai. Logarithmic energy as an entropy functional. *Contemporary Mathematics*, 1998.

[14] Sylvia Serfaty. *Coulomb Gases and Ginzburg-Landau Vortices*. European Mathematical Society Publishing House, Zuerich, Switzerland, mar 2015.

[15] D W Stroock. *Probability Theory, An Analytic View*. Cambridge University Press, Cambridge, 1993.

[16] Ran Wang, Xinyu Wang, and Liming Wu. Sanov's theorem in the Wasserstein distance: A necessary and sufficient condition. *Statistics & Probability Letters*, 80(5-6):505–512, mar 2010.