# LARGE DEVIATIONS FOR SINGULARLY INTERACTING DIFFUSIONS

JASPER HOEKSEMA*, THOMAS HOLDING, MARIO MAURELLI†, AND OLIVER TSE*

Abstract. In this paper we prove a large deviation principle (LDP) for the empirical measure of a general system of mean-field interacting diffusions with singular drift (as the number of particles tends to infinity) and show convergence to the associated McKean–Vlasov equation. Along the way, we prove an extended version of the Varadhan Integral Lemma for a discontinuous change of measure and subsequently an LDP for Gibbs and Gibbs-like measures with singular potentials.

## Contents

1. Introduction 2
2. An Extension of the Varadhan Integral Lemma 7
2.1. Notations and preliminary results 7
2.2. An extended Varadhan Integral Lemma 9
3. Gibbs measures 13
3.1. Notations and preliminary results 13
3.2. Main results 16
3.3. Extension to Gibbs-like potentials 21
4. Singularly interacting diffusions 24
4.1. Notations and preliminary results 24
4.2. The main result 27
4.3. Applications to concrete examples 32
4.4. Uniqueness for the McKean–Vlasov SDE 38
Appendix A. Limits of convex functions 48
Appendix B. Variational representation of exponential integrals 51
Appendix C. On measurability and exponential approximations 54
C.1. Measurability of integral maps 54
C.2. Exponential approximation theorem in Polish spaces 56
Appendix D. Stochastic estimates and technical parts of Section 4 57
D.1. Basic facts on the Girsanov theorem and the Novikov condition 57
D.2. Definition of the log-densities in Subsection 4.1 58
D.3. Proof of Lemma 4.1 and relative entropy representation 60
D.4. $L^q_t(L^p_x)$-estimates 62
References 65

* Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven,The Netherlands; email addresses: j.hoeksema@tue.nl, o.t.c.tse@tue.nl

† Dipartimento di Matematica 'Federigo Enriques', Università degli Studi di Milano, via Saldini 50, 20133 Milano, Italy; email address: mario.maurelli@unimi.it

## 1. Introduction

In this work we study the limiting behaviour of weakly interacting, or mean field, diffusions, where the interaction depends only on the empirical measures of the particles. For every $N \in \mathbb{N}$ the particle system is defined by the coupled stochastic differential equations (SDEs)

$$\begin{cases} dX_t^{N,i} = b_t\left(X_t^{N,i}, \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{N,i}}\right) dt + dW_t^i, \\ X_0^{N,i} \text{ i.i.d with law } \rho_0. \end{cases} \tag{1.1}$$

Here $W^1, \dots, W^N$ are independent $d$-dimensional Brownian motions, $\rho_0$ is a given initial distribution, and $b$ is a measure-dependent *drift* vector. A common example for $b$ is of the form

$$b_t(x, \mu) = \int_{\mathbb{R}^d} \varphi(t, x-y)\, d\mu(y), \tag{1.2}$$

for some interaction kernel $\varphi$. Such type of drifts commonly appear in models of classical physical systems, biological systems such as the collective motion of micro-organisms (bacteria, cells, etc.), and flocking and swarming behavior of animals, granular media, as well as models in opinion formation.

When $b$ is sufficiently regular, the limiting behaviour of the particle system for a large particle number is well understood. For example, when $b$ is Lipschitz and bounded, the empirical measure

$$z_{\boldsymbol{X}}^N := \frac{1}{N}\sum_{i=1}^N \delta_{X^{N,i}},$$

converges to the law of the McKean-Vlasov equation [Szn91, Tan84]

$$\begin{cases} dX_t = b_t(X_t, \mathrm{Law}(X_t))\, dt + dW_t, \\ X_0 \text{ with law } \rho_0. \end{cases} \tag{1.3}$$

Moreover a large deviation principle (LDP) holds for the empirical measure $z_{\boldsymbol{X}}^N$, as $N \to \infty$, see e.g. [DG87, BDF12, CDFM20].

The case of a *singular* interaction, that is irregular $b$, has been widely studied too. The convergence of the system (1.1) to the corresponding McKean-Vlasov equation has been shown for various examples of singular drifts, most of them of the form (1.2) with singular interaction kernel $\varphi$, e.g. [FHM14, GQ15, BO19, JW18, BJW19, Jab19] (see Section 4.4 for detailed explanations).

However, establishing LDPs for these singular drifts has remained unsolved. Apart from the work by Fontbona in [Fon04], where an LDP for the time-marginals of $(z_{\boldsymbol{X}}^N)$ was shown for a repulsive kernel $\varphi(x) = 1/x$, little is known to our knowledge. We aim to fill this gap, by providing LDP results and new tools for a large class of singular measure-dependent drifts.

As the main example, we consider the following drift

$$b_t(x, \mu) := \psi\left(x, \mu, \int_{\mathbb{R}^d} \varphi(t, x-y)\, d\mu(y)\right), \tag{1.4}$$

where $\psi : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}^d$, with $\mathcal{P}(\mathbb{R}^d)$ the space of probability measures equipped with the bounded Lipschitz metric $d_{BL}$. We will show an LDP when $\psi$ is Lipschitz and

the interaction kernel is in an appropriate $L^p$ space. More precisely, combining several key statements from Section 4 (see Proposition 4.16, Remark 4.18 and Proposition 4.24), we obtain the following result:

**Theorem 1.1.** *Suppose that*

*(*) (Lipschitz property) $\psi : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}^d$ is jointly globally Lipschitz, i.e. there exists some constant $M > 0$, such that for all $w, x, y, z \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$:*

$$|\psi(x, \mu, z) - \psi(y, \nu, w)| \leq M\Big(|x - y| + d_{BL}(\mu, \nu) + |z - w|\Big);$$

*(*) (Linear growth) there exists a constant $L > 0$ such that for all $x, z \in \mathbb{R}^d$, $\mu \in \mathcal{P}(\mathbb{R}^d)$:*

$$|\psi|(x, \mu, z) \leq L(1 + |z|);$$

*(*) (Exponential moment) for all $\beta > 0$, the initial distribution $\rho_0$ satisfies*

$$\int_{\mathbb{R}^d} e^{\beta|x|}\, d\rho_0(x) < \infty;$$

*(*) (Regularity) for $p, q \in [2, \infty]$ with $d/p + 2/q < 1$,*

$$\varphi \in L^q\big((0, T), L^p(\mathbb{R}^d)\big) + L^\infty((0, T) \times \mathbb{R}^d).$$

*Then the family $\{Q^N\}$ of laws of empirical measures $z^N_{\boldsymbol{X}}$ for $\boldsymbol{X} = (X^{N,1}, \ldots, X^{N,N})$ satisfying* (1.1)*, with drift $b$ as in* (1.4)*, has an LDP with rate function*

$$\mathcal{F}(\mu) = \begin{cases} R(\mu \| \mathbb{W}^\mu) & \text{if } R(\mu \| \mathbb{W}) < \infty, \\ +\infty & \text{otherwise.} \end{cases}$$

*Here, $R(\mu \| \nu)$ is the relative entropy of $\mu$ w.r.t. $\nu$ defined by*

$$R(\nu \| \mu) := \begin{cases} \displaystyle\int \log\left(\frac{d\nu}{d\mu}\right) d\nu & \text{if } \nu \ll \mu, \\ +\infty & \text{otherwise,} \end{cases}$$

*$\mathbb{W}^\mu$ is the law of a process $X^\mu_t$ satisfying the SDE*

$$dX^\mu_t = b(X^\mu_t, \mu_t)\, dt + dW_t,$$

*and $\mathbb{W} = \mathrm{Law}(W)$, where $W$ is a Brownian motion with initial law $\rho_0$ (and $\mu_t$ is the time marginal of $\mu$ at time $t$).*

*Furthermore, $z^N_{\boldsymbol{X}}$ converges almost surely to the unique minimizer of $\mathcal{F}(\mu)$, which is the unique law of the solution to the McKean-Vlasov SDE* (1.3)*.*

An application of Theorem 1.1 with $q = \infty$ is the case of a drift $b$ of the form (1.2), with

$$\varphi(t, z) = |z|^\alpha g\left(\frac{z}{|z|}\right) 1_{|z| \leq R} + h(z) 1_{|z| > R},$$

with $g : \mathbb{S}^{d-1} \to \mathbb{S}^{d-1}$ and $h : \mathbb{R}^d \to \mathbb{R}^d$ both Borel bounded, $R > 0$, and exponent $\alpha$ satisfying

$$\alpha > -1 \ \text{ for } \ d \geq 2, \qquad \text{and} \qquad \alpha > -1/2 \ \text{ for } \ d = 1.$$

Here the inequality for $d = 1$ arises from the fact that we require $|z|^\alpha \in L^p_{loc}(\mathbb{R}^d)$ for some $p$ with $p \geq 2$, while for $d > 1$ the constraint $p > d$ comes into play.

In fact, we prove LDP and convergence to the McKean-Vlasov equation for systems with singular drifts under more general assumptions, see Theorem 4.5 and the examples in Sections

4.3, 4.4, where we include many-particle interaction (that is, dependence of $b$ on $\mu^{\otimes k}$) and interaction kernels $\varphi$ merely satisfying

$$\mathbb{E}\left[e^{\beta\int_0^T|\varphi|^2\left(t,W_t^1,W_t^2\right)dt}\right]<\infty,\qquad \forall\beta\in\mathbb{R},$$

where $W^1, W^2$ are independent Brownian motions with common initial law $\rho_0$.

Note that even the convergence result to the McKean-Vlasov SDE in Theorem 1.1 is new: while some works do show convergence for the class of drifts (1.2) with even more singular $\varphi$ [FHM14, JW18, BJW19], we are not aware of a result that covers drifts of the form (1.4) under our assumptions.

Our proof of the LDP relies on using a singular change of measure via Girsanov's theorem and an approximation by regular drifts. Indeed, Girsanov's theorem gives at least formally

$$\frac{dQ^N}{dP^N}(\mu)=e^{-N\mathcal{E}(\mu)},\quad \mu\in\mathcal{P}(C([0,T];\mathbb{R}^d)), \tag{1.5}$$

where $Q^N$ is the law of the empirical measure $z^N_{\boldsymbol{X}}$ of the interacting particle system $\boldsymbol{X}=(X^{N,1},\dots,X^{N,N})$ satisfying (1.1), $P^N$ is the law of the empirical measures $z^N_{\boldsymbol{W}}$ of $N$ independent $d$-dimensional Brownian motions $\boldsymbol{W}=(W^1,\dots W^N)$ and

$$\begin{aligned}\mathcal{E}(\mu)&=\int V(x,\mu)\,\mu(dx),\\ V(x,\mu)&=-\int_0^T b_t(x_t,\mu_t)\cdot\,dx_t+\frac{1}{2}\int_0^T|b_t(x_t,\mu_t)|^2\,dt.\end{aligned} \tag{1.6}$$

By Sanov's theorem, $\{P^N\}$ satisfies an LDP with rate function $R(\cdot\|\mathbb{W})$. We need to transfer this LDP to an LDP for $\{Q^N\}$ and, to deal with this, we extend a classical tool in large deviation theory, Varadhan's Integral Lemma. We believe that this extension and other underlying results are relevant on their own and they might also prove helpful in future works on singular drifts. Let us briefly state our strategy.

In a nutshell, Varadhan's Integral Lemma (or Varadhan's Lemma for short) allows one to transfer the LDP through a continuous change of measure and is in fact a natural extension of Laplace's method to infinite dimensional spaces. To be precise, let $\{z^N\}$ be a family of random variables on a probability space $(\Omega,\mathcal{A},\mathbb{P})$, taking values in a Polish space $\mathcal{X}$ endowed with its Borel $\sigma$-algebra $\mathcal{B}(\mathcal{X})$. The classical Varadhan's Lemma (cf. [Var84, DMZ03, DZ10]) reads as follows.

**Proposition 1.2** (Varadhan's Integral Lemma)**.** *Suppose $P^N$ satisfies an LDP with good rate function $I:\mathcal{X}\to[0,\infty]$, and let $\mathcal{E}:\mathcal{X}\to\mathbb{R}$ be any continuous and bounded function. Then the family of measures $\{Q^N\}$ defined by*

$$\frac{dQ^N}{dP^N}(\mu):=\frac{1}{Z_N}e^{-N\mathcal{E}(\mu)}\qquad \textit{for } P^N\textit{-almost every } \mu\in\mathcal{X}, \tag{1.7}$$

*with normalizing constants $Z_N$, satisfies an LDP with good rate function*

$$F(\mu):=(I+\mathcal{E})(\mu)-\inf_{\nu\in\mathcal{X}}(I+\mathcal{E})(\nu). \tag{1.8}$$

Subsequently, various extensions have been developed in the past decades to relax the assumption of continuity; for example to deal with singular functionals or contractions for Gibbs measures on $\mathbb{R}^d$ [BG99, CGZ14, DLR20, HLSS18], or on abstract spaces [Léo87, ES98,

ES02, DMZ03, Ber18, GZ19, LW20]. We refer to the background paragraphs in Sections 2 and 3 for a more detailed discussion, and only highlight a few points here.

A common thread in some of the extensions above are various approximation arguments. For example, as outlined in [DZ10] on *exponential approximations*, the family $\{Q^N\}$ satisfies an LDP if there exists another family $\{Q^N_\lambda\}$ which satisfies an LDP for each $\lambda > 0$ and approximate $Q^N$ in some exponentially good way (as $\lambda \to 0$). Liu and Wu [LW20] make use of techniques involving exponential approximations and prove LDPs for Gibbs measures with singular potential. However, in our setting, we cannot rely on their result since, for general drift as in (1.4), the associated $\mathcal{E}$ is not actually in the form of a Gibbs energy. Hence, we have developed the following extension (see Theorem 2.8 for the precise statement):

**Theorem 1.3** (Extended Varadhan Integral Lemma)**.** *Let $P^N = \mathrm{Law}(z^N)$ be a family satisfying an LDP with rate function $I$, and $\mathcal{E}, \mathcal{E}^N : \mathcal{X} \to [-\infty, \infty]$ measurable functions. Moreover, let $\mathcal{E}_\lambda$, $\mathcal{E}^N_\lambda : \mathcal{X} \to [-\infty, \infty]$, $\lambda > 0$ be measurable functions such that, for each $\lambda > 0$, the family of measures $\{Q^N_\lambda\}_N$ defined by*

$$\frac{dQ^N_\lambda}{dP^N} := \frac{1}{Z^\lambda_N} e^{-N\mathcal{E}^N_\lambda(\mu)} \qquad \text{for } P^N\text{-almost every } \mu \in \mathcal{X}$$

*with normalizing constants $Z^\lambda_N$, satisfy an LDP with rate function*

$$\mathcal{F}_\lambda(\mu) := (I + \mathcal{E}_\lambda)(\mu) - \inf_{\nu \in \mathcal{X}} (I + \mathcal{E}_\lambda)(\nu).$$

*Suppose that for some $\gamma > 1$,*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[ e^{-\gamma N \mathcal{E}^N_\lambda(z^N)} \right] < +\infty,$$
$$\inf_{\mu \in D(I)} (I + \gamma\mathcal{E}_\lambda)(\mu) > -\infty,$$

*for every $\lambda > 0$, and that a constant $K \in \mathbb{R}$ exists, such that for every $\beta \in \mathbb{R}$,*

$$\limsup_{\lambda\to 0} \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[ e^{\beta N(\mathcal{E}^N - \mathcal{E}^N_\lambda)(z^N)} \right] \le K,$$
$$\limsup_{\lambda\to 0} \sup_{\mu \in D(I)} \big(\beta(\mathcal{E}_\lambda - \mathcal{E}) - I\big)(\mu) \le K.$$

*Then the family $\{Q^N\}$ defined by*

$$\frac{dQ^N}{dP^N} := \frac{1}{Z_N} e^{-N\mathcal{E}^N(\mu)} \qquad \text{for } P^N\text{-almost every } \mu \in \mathcal{X}$$

*satisfies an LDP with rate function*

$$\mathcal{F}(\mu) := (I + \mathcal{E})(\mu) - \inf_{\nu \in \mathcal{X}} (I + \mathcal{E})(\nu).$$

We provide a self-contained proof which only relies on basic large deviation theory and elementary convexity estimates. The two main points of this extension are that we do not require $\mathcal{E}$ to be continuous and that we also do not require the approximating energies $\mathcal{E}^N_\lambda$ to be continuous, but merely to induce an LDP in the sense described above. The latter point is an essential tool in proving Theorem 1.1. Namely, even for regular drifts $b$, the change of measure (1.5) provided via Girsanov's theorem is not necessarily continuous, while there are classical results on LDPs for interacting particle systems with regular drifts. Hence we can

prove Theorem 1.1 by approximating the singular drift $b$ in (1.6) with regular drifts $b_\lambda$ and applying then the extended Varadhan's Lemma Theorem 1.3.

Other extensions and applications of Varadhan's Lemma have also been developed to deal with this, for example to prove LDPs for weakly interacting diffusions with regular drifts [PdH96, DMZ03, DFMS17]. Moreover, in [DFMS17], one of the authors of this manuscript developed an enhanced version of Sanovs theorem in the rough path setting, which allows for Varadhans Integral Lemma to be applied. However, to our knowledge, none of them have been used to prove LDPs for weakly interacting diffusions with *singular* interaction.

It should be noted that one cannot expect to establish an LDP via a change of measure for every singular drift. This is indeed the case when the law of the interacting particle system is not absolutely continuous with respect to the law of the non-interacting system. A particular example in which this occurs is the Keller-Segel model with $\varphi(t,z) = -\nabla \log z$, for which a notion of propagation of chaos was established in [BJW19] but an LDP result remains open.

However, even in the large class of drifts for which the system can be described via a change of measure there is still a gap between those for which there is a known LDP and propagation of chaos, and those for which others have merely shown propagation of chaos. We believe that not only are our results such as Theorem 1.1 a sizable step in closing this gap, but that the general tools we provide will help close it even further.

***Organization and highlights of the paper.*** In Section 2 we recall basic definitions and results in large deviation theory, and provide an extension of Varadhan's Lemma in Theorem 2.8. Next, in Section 3, we use this to prove LDPs for empirical measures of mean-field Gibbs systems, where the log-densities $\mathcal{E}_V^N : \mathcal{P}(S) \to [-\infty, \infty]$ are parameterized by a family of Borel functions $V^N : S^k \to [-\infty, \infty]$, $k \in \mathbb{N}$ (see (3.1) for a precise definition). In Theorem 3.4, we provide sufficient conditions, in terms of suitable approximations of $\{V^N, V\}$, under which $\{\mathcal{E}_V^N, \mathcal{E}_V\}$ induces an LDP—this result is in the same spirit as [LW20]. We further prove a result (Theorem 3.10) that generalizes Theorem 3.4 to include 'Gibbs-like' measures—which are not of Gibbs form but such that the error $\mathcal{E}_V^N - \mathcal{E}_{V_\lambda}^N$ can be essentially bounded by Gibbs measures—that simplifies our task in proving LDPs for weakly interacting diffusions in Section 4.

In the latter, with the results of Sections 2 and 3 at hand, establishing an LDP for a system of weakly interacting diffusions amounts to (1) proving a (change-of-measure) representation formula (Girsanov's formula) for the laws $\{Q^N\}$ of the empirical measures $z_{\boldsymbol{X}}^N$ associated to the solution $\boldsymbol{X} = (X^{N,1}, \dots, X^{N,N})$ of (1.1); and (2) proving the existence of a family $\{b_\lambda^N\}$ of "exponentially good" approximations for $b$ (cf. Theorem 4.5 and the concrete examples in Section 4.3), which implies Theorem 1.1 considered above with a drift $b$ specified by (1.2). For drifts $b$ satisfying the assumption of Theorem 1.1, and for all the examples in Subsections 4.3, we further show that the rate function $\mathcal{F}$ associated to $\{Q^N\}$ attains a unique minimizer (see Subsection 4.4 for the general case), which implies almost sure convergence.

Finally, for the sake of completeness and consistency, we have included a relatively large appendix containing technical results and proofs utilized throughout the article, which we believe to be either new, or helpful to the reader.

***Acknowledgements.*** The authors thank M.A. Peletier for proposing the problem and for fruitful discussions at the start of the work. J.H. and O.T. acknowledges support from NWO Vidi grant 016.Vidi.189.102, "Dynamical-Variational Transport Costs and Application

to Variational Evolutions". Parts of this work were undertaken when M.M. was at Weierstrass Institute for Applied Analysis and Stochastics, Germany, at Technische Universität Berlin, Germany, and at University of York, UK. M.M. acknowledges support from the Royal Society via the Newton International Fellowship NF170448 "Stochastic Euler equations and the Kraichnan model", from project PRIN 2015233N54_002 "Deterministic and stochastic evolution equations" from the Italian Ministry of Education, University and Research, and from the Hausdorff Research Institute for Mathematics in Bonn under the Junior Trimester Program "Randomness, PDEs and Nonlinear Fluctuations".

## 2. An Extension of the Varadhan Integral Lemma

In this chapter we extend the classical Varadhan's Integral Lemma (cf. [DZ10, Var84]), to allow for establishing LDPs via a change of measure with possibly discontinuous density.

2.1. **Notations and preliminary results.** We introduce some notation that we will use throughout the manuscript. The space $\mathcal{X}$ is a Polish space endowed with its Borel $\sigma$-algebra $\mathcal{B}(\mathcal{X})$. The symbol $\mathcal{P}(\mathcal{X})$ denotes the set of all probability measures on $\mathcal{X}$ and we use the letters $P$, $Q$, and $P^N$, $Q^N$, ... for probability measures on $\mathcal{X}$. With a little abuse of notation, we use $\mu$ both for a generic element of $\mathcal{X}$ and for the canonical random variable on $\mathcal{X}$ ($\mu(x) = x$ for all $x$ in $\mathcal{X}$); this notation is unusual, but it will be convenient in the next sections, where $\mathcal{X}$ will be itself a space of probability measures. We often consider (without loss of generality) $P^N$ as the law of an $\mathcal{X}$-valued random variable $z^N$, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ (independent of $N$); $\mathbb{E}$ denotes the expectation with respect to $\mathbb{P}$.

We recall a definition of a large deviation principle (LDP).

**Definition 2.1.** A family of measures $\{Q^N\} \subset \mathcal{P}(\mathcal{X})$ satisfies an LDP with *rate function* $\mathcal{F} : \mathcal{X} \to [0, \infty]$ if (1) $\mathcal{F}$ is lower semi-continuous, if (2) for every Borel set $A$,

$$-\inf_{\mu \in A^o} \mathcal{F}(\mu) \leq \liminf_{N\to\infty} \frac{1}{N} \log Q^N(A) \leq \limsup_{N\to\infty} \frac{1}{N} \log Q^N(A) \leq -\inf_{\mu\in\bar{A}} \mathcal{F}(\mu),$$

and if (3) the family $\{Q^N\}$ is *exponentially tight*, i.e. there is a sequence of compact sets $K_M \subset \mathcal{X}$ such that

$$\limsup_{M\to\infty} \limsup_{N\to\infty} \frac{1}{N} \log Q^N(\mathcal{X} \setminus K_M) = -\infty.$$

We denote the domain of $\mathcal{F}$ by $D(\mathcal{F}) := \{\mu \in \mathcal{X} \,|\, \mathcal{F}(\mu) < \infty\}$.

**Remark 2.2.** As shown in [DZ10, p. 8, 120], Definition 2.1 in Polish spaces is equivalent to stating that (2) holds with a *good rate function* $\mathcal{F}$, i.e. $\mathcal{F}$ having compact sub-level sets.

**Remark 2.3.** Let $\{z^N\}$ be a family of $\mathcal{X}$-valued random variables such that $Q^N = \mathrm{Law}(z^N) \in \mathcal{P}(\mathcal{X})$ satisfies an LDP with rate function $\mathcal{F}$. If the minimizer $\mu^* \in \mathcal{X}$ of $\mathcal{F}$ is unique, then the LDP implies the convergence $Q^N \to \delta_{\mu^*}$ weakly. In fact, by a standard argument we obtain a stronger result: almost sure convergence of the random variables $z^N$ to $\mu^*$, as stated below. For a proof, see for example [PS19, Theorem A.2].

**Lemma 2.4.** *Suppose $P^N$ satisfies an LDP with rate function $\mathcal{F}$, and that $\mathcal{F}$ has a unique minimizer $\mu^*$. Then $z^N$ converges $\mathbb{P}$-almost surely to $\mu^*$.*

Now, let $P^N = \mathrm{Law}(z^N)$ be probability measures on $\mathcal{X}$ satisfying a large deviation principle with rate function $I : \mathcal{X} \to [0, \infty]$. We consider pairs $(\{\mathcal{E}^N\}, \mathcal{E})$ (denoted $(\mathcal{E}^N, \mathcal{E})$ for short) of a sequence of Borel functions $\mathcal{E}^N : \mathcal{X} \to \overline{\mathbb{R}}$ and a Borel function $\mathcal{E} : \mathcal{X} \to \overline{\mathbb{R}}$, and study whether an LDP may be established for the induced measures $Q^N$,

$$\frac{dQ^N}{dP^N}(\mu) := \frac{1}{Z_N} e^{-N\mathcal{E}^N(\mu)} \qquad \text{for } P^N\text{-almost every } \mu \in \mathcal{X}, \tag{2.1}$$

where the normalization constants $Z_N$ are assumed to be finite for all $N \in \mathbb{N}$.

Precisely, we define $J$ and $\mathcal{F}$ as follows:

$$J(\mu) := \begin{cases} I(\mu) + \mathcal{E}(\mu) & \mu \in D(I), \\ +\infty & \mu \notin D(I), \end{cases}$$

and, if $\inf_{\mu \in \mathcal{X}} J(\mu)$ is finite, we define

$$\mathcal{F}(\mu) := J(\mu) - \inf_{\mu \in \mathcal{X}} J(\mu).$$

Finally, note that by construction for any Borel set $A$

$$\inf_{\mu \in A \cap D(I)} (\mathcal{E} + I)(\mu) = \inf_{\mu \in A} J(\mu).$$

Then the property we investigate is given in the following definition.

**Definition 2.5.** We say that $(\mathcal{E}^N, \mathcal{E})$ induces an LDP if $\inf_{\mu \in \mathcal{X}} J(\mu)$ is finite, $\{Q^N\}$ (defined as in (2.1)) satisfies an LDP with rate function $\mathcal{F}$ and satisfies the so-called *Laplace principle*,

$$\lim_{N \to \infty} \frac{1}{N} \log \mathbb{E}\left[e^{-N\mathcal{E}^N(z^N)}\right] = -\inf_{\mu \in \mathcal{X}} J(\mu). \tag{2.2}$$

The following lemma provides a characterization in terms of an *unnormalized* LDP.

**Lemma 2.6.** *The following statements are equivalent:*

*(*) The pair $(\mathcal{E}^N, \mathcal{E})$ induces an LDP (according to Definition 2.5);*

*(*) $\inf_{\mu \in \mathcal{X}} J(\mu) \in \mathbb{R}$, $J$ is lower semi-continuous, the family $\{Q^N\}$ is exponentially tight, and for every Borel set $A$,*

$$\begin{aligned} -\inf_{\mu \in A^o} J(\mu) &\le \liminf_{N \to \infty} \frac{1}{N} \log \mathbb{E}\left[e^{-N\mathcal{E}^N(z^N)} 1_A\right] \\ &\le \limsup_{N \to \infty} \frac{1}{N} \log \mathbb{E}\left[e^{-N\mathcal{E}^N(z^N)} 1_A\right] \quad \le -\inf_{\mu \in \bar{A}} J(\mu). \end{aligned} \tag{2.3}$$

*Proof.* Suppose $(\mathcal{E}^N, \mathcal{E})$ induces an LDP. Exponential tightness of $Q^N$ follows from the definition of an LDP, and since $\mathcal{F}$ is lower semi-continuous $J$ is as well. By the Laplace principle (2.2),

$$\lim_{N \to \infty} \frac{1}{N} \log \mathbb{E}\left[e^{-N\mathcal{E}^N(z^N)}\right] = -\inf_{\mu \in \mathcal{X}} J(\mu),$$

where the right-hand side is assumed to be finite. For for any $A \in \mathcal{B}(\mathcal{X})$, we have that

$$\begin{aligned} \frac{1}{N} \log \mathbb{E}\left[e^{-N\mathcal{E}^N(z^N)} 1_A\right] &= \frac{1}{N} \log \mathbb{E}\left[\frac{1}{Z^N} e^{-N\mathcal{E}^N(z^N)} 1_A\right] + \frac{1}{N} \log Z^N \\ &= \frac{1}{N} \log Q^N(A) + \frac{1}{N} \log Z^N. \end{aligned}$$

Therefore, by (2.2) and the LDP of $Q^N$, we then obtain

$$\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-N\mathcal{E}^N(z^N)}1_A\right]=\limsup_{N\to\infty}\frac{1}{N}\log Q^N(A)-\inf_{\mu\in\mathcal{X}}J(\mu)$$
$$\leq-\inf_{\mu\in\bar{A}}\mathcal{F}(\mu)-\inf_{\mu\in\mathcal{X}}J(\mu)=-\inf_{\mu\in\bar{A}}J(\mu).$$

The lower bound follows similarly, and hence (2.3) is satisfied.

Conversely, assume that $J$ is lower semi-continuous, $\inf_{\mu\in\mathcal{X}}J(\mu)$ is finite and that (2.3) holds. Then $\mathcal{F}$ is lower semi-continuous as well, and by (2.3) applied to $A=\mathcal{X}$,

$$\lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-N\mathcal{E}^N(z^N)}\right]=-\inf_{\mu\in\mathcal{X}}J(\mu),$$

where the right-hand side is assumed to be finite. Now normalizing by $Z_N$ and proceeding as above it follows that for any $A\in\mathcal{B}(\mathcal{X})$,

$$-\inf_{\mu\in A^o}\mathcal{F}(\mu)\leq\liminf_{N\to\infty}\frac{1}{N}\log Q^N(A)\leq\limsup_{N\to\infty}\frac{1}{N}\log Q^N(A)\leq-\inf_{\mu\in\bar{A}}\mathcal{F}(\mu).$$

Along with the exponential tightness of $Q^N$, this implies that $(\mathcal{E}^N,\mathcal{E})$ induces an LDP. □

**Remark 2.7.** Notice that the classical Varadhan's Integral Lemma is recovered when $\mathcal{E}$ is continuous and bounded, and $\mathcal{E}^N=\mathcal{E}$ for all $N\in\mathbb{N}$.

2.2. **An extended Varadhan Integral Lemma.** Now we present the main result of this section.

**Theorem 2.8** (Extended Varadhan Integral Lemma)**.** *Let $P^N=\mathrm{Law}(z^N)$ be a family satisfying an LDP with rate function $I$. Moreover, let $(\mathcal{E}^N_\lambda,\mathcal{E}_\lambda)$ be pairs inducing an LDP for all $\lambda>0$. Suppose that the pair $(\mathcal{E}^N,\mathcal{E})$ is such that for some $\gamma>1$,*

$$\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-\gamma N\mathcal{E}^N_\lambda(z^N)}\right]<+\infty, \tag{2.4a}$$

$$\inf_{\mu\in D(I)}(I+\gamma\mathcal{E}_\lambda)(\mu)>-\infty, \tag{2.4b}$$

*for every $\lambda>0$, and that a constant $K\in\mathbb{R}$ exists, such that for every $\beta\in\mathbb{R}$,*

$$\limsup_{\lambda\to0}\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{\beta N(\mathcal{E}^N-\mathcal{E}^N_\lambda)(z^N)}\right]\leq K, \tag{2.5a}$$

$$\limsup_{\lambda\to0}\sup_{\mu\in D(I)}\big(\beta(\mathcal{E}_\lambda-\mathcal{E})-I\big)(\mu)\leq K. \tag{2.5b}$$

*Then the family $\{Q^N\}$ defined by (2.1) satisfies an LDP with rate function $\mathcal{F}$. In particular, the pair $(\mathcal{E}^N,\mathcal{E})$ also induces an LDP.*

**Remark 2.9.** We give some comments on the assumptions:

(*) It was shown in [Hoe17] that under (2.4b), condition (2.5b) is equivalent to the uniform convergence of $\mathcal{E}_\lambda$ to $\mathcal{E}$ on the sub-level sets $\{\mu\,|\,I(\mu)\leq M\}$ of $I$ for any $M\in\mathbb{R}$, and that (2.5a) implies:

$$\text{For any }\delta>0:\quad\lim_{\lambda\to0}\limsup_{N\to\infty}\frac{1}{N}\log P^N\big(|\mathcal{E}-\mathcal{E}_\lambda|(z^N)>\delta\big)=-\infty. \tag{2.6}$$

(*) Condition (2.5b) could appear redundant for those who are familiar with LDPs: if $(\mathcal{E}^N - \mathcal{E}^N_\lambda, \mathcal{E} - \mathcal{E}_\lambda)$ is *a priori* known to induce an LDP, (2.5b) and (2.5a) are equivalent. But in the proof we need to approximate $\mathcal{E}^N$ and $\mathcal{E}$ separately, which requires us to have both conditions. Nevertheless, we do expect from this reasoning that bounds for (2.5a) are also bounds for (2.5b). We will see that this is indeed the case for the interacting particle systems in Section 3.

(*) Note that condition (2.5a) implies (cf. Lemma A.4)

$$\lim_{\lambda\to 0}\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-N\beta\mathcal{E}^N_\lambda(z^N)}\,1_A\right] = \limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-N\beta\mathcal{E}^N(z^N)}\,1_A\right],$$

and

$$\lim_{\lambda\to 0}\liminf_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-N\beta\mathcal{E}^N_\lambda(z^N)}\,1_A\right] = \liminf_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-N\beta\mathcal{E}^N(z^N)}\,1_A\right].$$

for any Borel set $A \in \mathcal{B}(\mathcal{X})$ and $\beta \in \mathbb{R}$, which is considerably stronger than simply inequalities for open and closed sets. An open question is whether this is stronger than, equivalent to or weaker than (or none of the former) the statement that the induced measures $Q^N_\lambda$ *exponentially approximates* $Q^N$ as $\lambda \to 0$ (cf. [DZ10, p. 130]).

(*) An alternative approach to prove the type of results in Theorem 2.8 could be to get an LDP for $z^N$ in a larger space with a stronger topology, where $\mathcal{E}^N$ is a continuous function, and then apply the classical Varadhan Lemma; the LDP in the stronger topology could be obtained by exponential approximation, as in [DZ10]. This strategy is used, for example in [ES02,LW20], in the context of certain singular Gibbs measures.

**Remark 2.10** (**Toward interacting diffusions**)**.** We outline briefly how Theorem 2.8 will be applied in our context of singularly interacting diffusions. Recall that we are interested in an LDP on the space $\mathcal{X} = \mathcal{P}(S)$, with $S = C([0,T];\mathbb{R}^d))$, for the law $Q^N$ of the interacting particle system $\boldsymbol{X} = (X^{N,1},\dots,X^{N,N})$ satisfying (1.1), as $N \to \infty$; the drift $b$ is possibly singular. As we have written in the Introduction, in this context we take $P^N$ as the law of the empirical measures $z^N_{\boldsymbol{W}}$ of $N$ independent $d$-dimensional Brownian motions $W^1,\dots W^N$: by Sanov's theorem, the family $\{P^N\}$ satisfies an LDP with rate function $I = R(\cdot\|\mathbb{W})$ (the relative entropy with respect to the Wiener measure $\mathbb{W}$). As a consequence of Girsanov's theorem, the energy is given formally by

$$\mathcal{E}(\mu) = \int_S V(x,\mu)\,d\mu(x),$$
$$V(x,\mu) = -\int_0^T b_t(x_t,\mu_t)\cdot\,dx_t + \frac{1}{2}\int_0^T |b_t(x_t,\mu_t)|^2\,dt.$$

The approximating energies $\mathcal{E}_\lambda$ and potentials $V_\lambda$ are defined as for $\mathcal{E}$ and $V$ but replacing $b$ with a smooth bounded approximation $b_\lambda$. The fact that, for each $\lambda > 0$, $\mathcal{E}_\lambda$ induces an LDP follows from classical results for interacting diffusions with smooth drift. The assumption (2.5a) reads in this context, for every $\beta \in \mathbb{R}$,

$$\limsup_{\lambda\to 0}\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{\beta\sum_{i=1}^N[V(W^i,z^N_{\boldsymbol{W}})-V_\lambda(W^i,z^N_{\boldsymbol{W}})]}\right] \le K,$$

with $K$ independent of $\beta$. Note that this condition involves deterministic and stochastic integrals of $N$ independent Brownian motions only. In order to check this condition, we need to control the dependence on $N$, which we do in Section 3 by making use of Gibbs-like structures.

Note finally that, even when the drift $b$ is smooth and bounded, $V$ and so $\mathcal{E}$ are not continuous in general (in the topology of weak convergence on $\mathcal{P}(S)$ at least), because $V$ involves a stochastic integral. As a technical detail, the precise definitions of $V$ and $\mathcal{E}$ in the particle system differ from the definitions in the McKean-Vlasov SDE, due to both the precise definition of the stochastic integral as Borel map and the issue of self-interaction, see Subsection D.2.

We now briefly explain the strategy to prove Theorem 2.8. For a Borel set $A \in \mathcal{B}(\mathcal{X})$, we define the following functionals on the space of Borel functions $\mathcal{E}$ on $\mathcal{X}$,

$$\begin{cases} \phi_A(\mathcal{E}) := -\inf\limits_{\mu \in A \cap D(I)} (I + \mathcal{E})(\mu), \\ \phi_A^N(\mathcal{E}) := \dfrac{1}{N} \log \mathbb{E}\left[ e^{-N\mathcal{E}(z^N)} 1_A \right], \qquad n \in \mathbb{N}. \end{cases} \tag{2.7}$$

We will show in Lemma 2.11 that $\phi_A^N$ and $\phi_A$ are convex and from above (in $A$) by $\phi_{\mathcal{X}}^N$ and $\phi_{\mathcal{X}}$ respectively. Moreover, the fact that $(\mathcal{E}^N, \mathcal{E})$ induces an LDP can be read as a set of variational inequalities for $\phi_A^N$ and $\phi_A$ for each $A \in \mathcal{B}(\mathcal{X})$, i.e.

$$\phi_{A^o}(\mathcal{E}) \le \liminf_{N\to\infty} \phi_A^N(\mathcal{E}^N) \le \limsup_{N\to\infty} \phi_A^N(\mathcal{E}^N) \le \phi_{\bar{A}}(\mathcal{E}).$$

Finally, the convergence of (2.5) over $\mathcal{X}$ will be seen to imply corresponding statements over every set $A \in \mathcal{B}(\mathcal{X})$, which implies bounds on $\phi_A^N(\mathcal{E}^N - \mathcal{E}_\lambda^N)$ and $\phi_A(\mathcal{E} - \mathcal{E}_\lambda)$. Hence the extended Varadhan Integral Lemma is morally equivalent to a type of stability of variational inequalities for convex functionals, for which we can use Theorem A.1 in Appendix A.

Here are the convexity properties and bounds for $\phi_A$ and $\phi_A^N$:

**Lemma 2.11.** *For any Borel set $A \in \mathcal{B}(\mathcal{X})$ and any $N \in \mathbb{N}$, the functionals $\phi_A$ and $\phi_A^N$ defined in* (2.7) *are convex and bounded from above by $\phi_{\mathcal{X}}$ and $\phi_{\mathcal{X}}^N$ respectively (that is, $\phi_A \le \phi_{\mathcal{X}}$ and $\phi_A^N \le \phi_{\mathcal{X}}^N$ for every $A \in \mathcal{B}(\mathcal{X})$ and any $N \in \mathbb{N}$).*

*Proof.* For any $N \in \mathbb{N}$ and any $\alpha \in (0,1)$, for any Borel functions $\mathcal{E}_1$, $\mathcal{E}_2$, it holds by Hölder's inequality (with exponents $1/\alpha$ and $1/(1-\alpha)$)

$$\begin{aligned} \log \mathbb{E}\left[ e^{-N(\alpha\mathcal{E}_1 + (1-\alpha)\mathcal{E}_2)(z^N)} 1_A \right] &= \log \mathbb{E}\left[ e^{-\alpha N \mathcal{E}_1(z^N)} 1_A \, e^{-(1-\alpha)N\mathcal{E}_2(z^N)} 1_A \right] \\ &\le \alpha \log \mathbb{E}\left[ e^{-N\mathcal{E}_1(z^N)} 1_A \right] + (1-\alpha) \log \mathbb{E}\left[ e^{-N\mathcal{E}_2(z^N)} 1_A \right]. \end{aligned} \tag{2.8}$$

Convexity of $\phi_A^N$ follows by dividing (2.8) by $N$. The bound $\phi_A^N \le \phi_{\mathcal{X}}^N$ follows from the positivity of the exponential and the monotonicity of the logarithm. Finally, for any $\alpha \in (0,1)$,

$$\begin{aligned} \inf_{\mu \in A \cap D(I)} I + (\alpha\mathcal{E}_1 + (1-\alpha)\mathcal{E}_2) &= \inf_{\mu \in A \cap D(I)} [\alpha(I + \mathcal{E}_1) + (1-\alpha)(I + \mathcal{E}_2)] \\ &\ge \alpha \inf_{\mu \in A \cap D(I)} (I + \mathcal{E}_1) + (1-\alpha) \inf_{\mu \in A \cap D(I)} (I + \mathcal{E}_2), \end{aligned}$$

which gives convexity of $\phi_A$. The bound $\phi_A \le \phi_{\mathcal{X}}$ is easily verified. □

We can now prove Theorem 2.8.

*Proof of Theorem 2.8.* Recall, for any Borel set $A \in \mathcal{B}(\mathcal{X})$ and Borel function $\mathcal{G}$, we have by definition

$$\phi_{A^o}(\mathcal{G}) = -\inf_{\mu \in A^o \cap D(I)} (I + \mathcal{G})(\mu), \qquad \phi_{\bar{A}}(\mathcal{G}) = -\inf_{\mu \in \bar{A} \cap D(I)} (I + \mathcal{G})(\mu),$$

and

$$\phi_A^N(\mathcal{G}) = \frac{1}{N}\log \mathbb{E}\left[e^{-N\mathcal{G}(z^N)}1_A\right].$$

Lemma 2.11 gives that $\phi_{A^o}$, $\phi_{\bar{A}}$ and $\phi_A^N$ are convex for every $N \in \mathbb{N}$. Moreover, by the bounds $\phi_A^N \le \phi_{\mathcal{X}}^N$ and $\phi_{\bar{A}} \le \phi_{\mathcal{X}}$, assumptions (2.4) imply, for some $\gamma > 1$ (independent of $\lambda$),

$$\left.\begin{aligned} \limsup_{N\to\infty} \phi_A^N(\gamma\mathcal{E}_\lambda^N) < +\infty \\ \phi_{\bar{A}}(\gamma\mathcal{E}_\lambda) < +\infty \end{aligned}\right\} \quad \text{for every } \lambda > 0,$$

while assumptions (2.5) imply

$$\left.\begin{aligned} \limsup_{\lambda\to 0}\limsup_{N\to\infty} \phi_A^N(\beta(\mathcal{E}^N - \mathcal{E}_\lambda^N)) \le K \\ \limsup_{\lambda\to 0} \phi_{\bar{A}}(\beta(\mathcal{E} - \mathcal{E}_\lambda)) \le K \end{aligned}\right\} \quad \text{for every } \beta \in \mathbb{R}.$$

By Lemma 2.6, the assumption that $(\mathcal{E}_\lambda^N, \mathcal{E}_\lambda)$ induces an LDP is characterized by

$$\phi_{A^o}(\mathcal{E}_\lambda) \le \liminf_{N\to\infty} \phi_A^N\left(\mathcal{E}_\lambda^N\right) \le \limsup_{N\to\infty} \phi_A^N\left(\mathcal{E}_\lambda^N\right) \le \phi_{\bar{A}}(\mathcal{E}_\lambda). \tag{2.9}$$

We are now in a position to apply Theorem A.1, which implies that (2.9) also holds for $(\mathcal{E}^N, \mathcal{E})$ (cf. (A.5)), i.e.

$$\phi_{A^o}(\mathcal{E}) \le \liminf_{N\to\infty} \phi_A^N\left(\mathcal{E}^N\right) \le \limsup_{N\to\infty} \phi_A^N\left(\mathcal{E}^N\right) \le \phi_{\bar{A}}(\mathcal{E}).$$

Moreover, for every $\gamma' \in (0,\gamma)$, both $\limsup_{N\to\infty} \phi_A^N(\gamma'\mathcal{E}^N) < +\infty$ and $\phi_{\bar{A}}(\gamma'\mathcal{E}) < +\infty$ (cf. (A.4a) and (A.4b)). In particular, for $A = \mathcal{X}$, we have that $-\phi_{\mathcal{X}}(\mathcal{E})$ is finite, and

$$\limsup_{N\to\infty} \frac{1}{N}\log \mathbb{E}\left[e^{-\gamma N\mathcal{E}^N(z^N)}\right] < \infty. \tag{2.10}$$

By Lemma 2.6, we can conclude that $(\mathcal{E}^N, \mathcal{E})$ induces an LDP provided we show lower semi-continuity of $J$ and exponential tightness of $Q^N$.

First, from (2.5b) we can conclude that for every $K' > K$ and $\beta \ge 0$ there exists a large enough $\lambda^*(K', \beta)$ such that

$$|\mathcal{E}_\lambda - \mathcal{E}|(\mu) \le \frac{K' + I(\mu)}{\beta}, \quad \forall \mu \in \mathcal{X}, \forall \lambda \ge \lambda^*(K', \beta).$$

In particular we derive that $\mathcal{E}_\lambda$ converges pointwise to $\mathcal{E}$ on $D(I)$ (in fact, the convergence is uniform on sub-level sets of $I$). Next, note that a Borel function $J$ is lower semi-continuous if and only if for every $\mu \in \mathcal{X}$

$$\liminf_{\epsilon\to 0} \inf_{\nu\in B_\epsilon(\mu)} J(\nu) \ge J(\mu),$$

which in the case of

$$J_{\mathcal{G}}(\mu) := \begin{cases} I(\mu) + \mathcal{G}(\mu) & \mu \in D(I), \\ +\infty & \mu \notin D(I), \end{cases}$$

for a Borel function $\mathcal{G}$ can be rewritten as

$$\limsup_{\epsilon\to 0} \phi_{B_\epsilon(\mu)}(\mathcal{G}) \le -J_{\mathcal{G}}(\mu). \tag{2.11}$$

To show this for $\mathcal{G} = \mathcal{E}$, fix $\mu$, and note that by convexity for any $\alpha \in [0,1)$, $\lambda$, $\epsilon$,

$$\begin{aligned}\phi_{B_\epsilon(\mu)}(\alpha\mathcal{E}) &\leq \alpha\phi_{B_\epsilon(\mu)}(\mathcal{E}_\lambda) + (1-\alpha)\phi_{B_\epsilon(\mu)}\left(\alpha(1-\alpha)^{-1}(\mathcal{E}-\mathcal{E}_\lambda)\right)\\ &\leq \alpha\phi_{B_\epsilon(\mu)}(\mathcal{E}_\lambda) + (1-\alpha)\phi_{\mathcal{X}}\left(\alpha(1-\alpha)^{-1}(\mathcal{E}-\mathcal{E}_\lambda)\right).\end{aligned}$$

Since the left-hand side is independent of $\lambda$, taking subsequently limits in $\epsilon$ and $\lambda$ and using the lower semi-continuity of $I + \mathcal{E}_\lambda$ we derive

$$\begin{aligned}\limsup_{\epsilon\to 0}\phi_{B_\epsilon(\mu)}(\alpha\mathcal{E}) &\leq \alpha\liminf_{\lambda\to 0}\limsup_{\epsilon\to 0}\phi_{B_\epsilon(\mu)}(\mathcal{E}_\lambda) + (1-\alpha)\limsup_{\lambda\to 0}\phi_{\mathcal{X}}\left(\alpha(1-\alpha)^{-1}(\mathcal{E}-\mathcal{E}_\lambda)\right)\\ &\leq -\alpha\limsup_{\lambda\to 0} J_{\mathcal{E}_\lambda}(\mu) + (1-\alpha)K.\end{aligned}$$

Since $g(\alpha) := \limsup_{\epsilon\to 0}\phi_{B_\epsilon(\mu)}(\alpha\mathcal{E})$ is convex and bounded from above around $\alpha = 1$ we conclude by Lemma A.3 after letting $\alpha \to 1$,

$$\limsup_{\epsilon\to 0}\phi_{B_\epsilon(\mu)}(\mathcal{E}) \leq -\limsup_{\lambda\to 0} J_{\mathcal{E}_\lambda}(\mu).$$

Now, to establish (2.11) for $\mathcal{G} = \mathcal{E}$, note that either $I(\mu) = +\infty$ in which case the inequality trivially holds, or we have $\mu \in D(I)$ and thus we employ the pointwise convergence of $\mathcal{E}_\lambda$ to $\mathcal{E}$.

Finally, to prove exponential tightness, fix an arbitrary $M \geq 1$ and let $K_M$ be a compact set in $\mathcal{X}$ such that

$$\limsup_{N\to\infty}\frac{1}{N}\log P^N(\mathcal{X}\setminus K_M) < -M.$$

By Hölder inequality we derive, for every $\alpha$ in $(0,1)$,

$$\frac{1}{N}\log\mathbb{E}\left[e^{-N\mathcal{E}^N(z^N)}1_{\mathcal{X}\setminus K_M}\right] \leq \frac{\alpha}{N}\log\mathbb{E}\left[e^{\alpha^{-1}N\mathcal{E}^N(z^N)}\right] + \frac{1-\alpha}{N}\log P^N\big(\mathcal{X}\setminus K_M\big).$$

After taking the limit supremum in $N$, the first term on the right-hand side is independent of $M$ and finite by (2.10), provided $\alpha^{-1} < \gamma$. Therefore,

$$\limsup_{M\to\infty}\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-N\mathcal{E}^N(z^N)}1_{\mathcal{X}\setminus K_M}\right] = -\infty,$$

which proves the exponential tightness of $Q^N$ and concludes the proof. □

***Background.*** We are only aware of one paper extending Varadhan lemma for discontinuous log-densities in a general framework, namely [DMZ03], which however uses different assumptions. Instead, extensions of Varadhan's lemma for particular contexts have been proven, often involving conditions like (2.6). For example, [ES98] proves an extended contraction principle which is closely related to exponential approximations (cf. [DZ10]); a localized version of (2.6) is used in [BG99] for their concept of quasi-continuity to prove LDPs for vortex systems; the papers [ES02, LW20] use an alternative strategy to get an extension of the classical Varadhan Lemma in the context of singular Gibbs measures, see Remark 2.9(iv).

## 3. Gibbs measures

3.1. **Notations and preliminary results.** In this section we consider the setting for Gibbs measures over weakly interacting particle systems.

Let $S$ be a Polish space, endowed with its Borel $\sigma$-algebra $\mathcal{B}(S)$. Let $\mu_0 \in \mathcal{P}(S)$ be a given reference measure. For simplicity, we will suppose in the following that $\mu_0$ is *non-atomic*, i.e. it has no atoms (cf. Remark 3.1). We are given i.i.d. random variables $\omega_i$, $i \in \mathbb{N}$, defined on

some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with values in $S$ and common law $\mu_0$; we can think of $\omega_i$ as non-interacting particles. We denote by $\mathbb{E}$ the expectation with respect to $\mathbb{P}$.

To describe our interacting particle system, we fix $k$ in $\mathbb{N}$, $k \geq 2$, and take a $k$-particle interaction potential $V^N$ on $S$, i.e. a Borel function $V^N : S^k \to \overline{\mathbb{R}}$. We define the energy $E_V^N : S^N \to \overline{\mathbb{R}}$ of an $N$-particle configuration ($N \gg k$) by

$$E_V^N(x_1, \dots, x_N) := \frac{1}{N^k} \sum_{i_1, \dots i_k \text{ distinct}} V^N(x_{i_1}, \dots, x_{i_k}),$$

(when the $N$-dependence is made explicit in the superscript, with a little abuse of notation we use $E_V^N$ instead of $E_{V^N}^N$). Then the interacting particle system is described by the probability measure $\mathbb{Q}_V^N$ on $(\Omega, \mathcal{A})$ defined by

$$\mathbb{Q}_V^N = \frac{1}{Z_V^N} e^{-N E_V^N(\omega_1, \dots \omega_N)} \mathbb{P},$$

where $Z_V^N$ is the normalizing constant, assumed to be finite. Under $\mathbb{Q}_V^N$, the particles $\omega_i$ are subject to interaction via the potential $V^N$. Note that the energy $E_V^N$ is invariant under permutation, or, in other words, depends only on the positions of the particle $\omega_i$ and not on their label $i$, via a fixed interaction potential $V^N$—this is a *mean-field* interaction. Note also that particle configurations $(x_1, \dots, x_N)$ are more likely, according to $\mathbb{Q}_V^N$, if $V^N$ assumes lower values in these configurations.

Our main interest is in large deviations for the empirical measures associated with $\omega_i$ under the probability measure $\mathbb{Q}_V^N$. For this reason, we consider the state space $\mathcal{X} = \mathcal{P}(S)$, equipped with the weak topology (w.r.t. continuous and bounded functions on $S$), which turns $\mathcal{P}(S)$ into a Polish space [DZ10, Theorem D.8]. For each $N \in \mathbb{N}$, we denote by $z_\bullet^N : S^N \to \mathcal{P}(S)$ the continuous map

$$S^N \ni (x_1, \dots, x_N) =: \boldsymbol{x} \mapsto z_{\boldsymbol{x}}^N := \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \in \mathcal{P}(S).$$

We denote by $P^N \in \mathcal{P}(\mathcal{P}(S))$, resp. $Q_V^N \in \mathcal{P}(\mathcal{P}(S))$ the law of the empirical measure $z_{\boldsymbol{\omega}}^N$ for $\boldsymbol{\omega} = (\omega_1, \dots, \omega_N)$ under $\mathbb{P}$ (where $\omega_i$ $i = 1, \dots N$ are i.i.d. with common law $\mu_0$), resp. under $\mathbb{Q}_V^N$. We aim at giving an LDP for $Q_V^N$.

We further define the function $\mathcal{E}_V^N : \mathcal{P}(S) \to \overline{\mathbb{R}}$ by

$$\mathcal{E}_V^N(\mu) := \begin{cases} \int_{(S^k)'} V^N \, d\mu^{\otimes k} & \text{if } V^N \in L^1(\mu^{\otimes k}), \\ +\infty & \text{otherwise,} \end{cases} \tag{3.1}$$

where $(S^k)'$ is $S^k$ but with the diagonals removed, i.e.

$$(S^k)' := \Big\{ (x_1, \dots, x_k) \in S^k \,\Big|\, x_i \neq x_j, \, \forall i, j \in \{1, \dots, k\} \text{ with } i \neq j \Big\}.$$

Notice that when $\omega_i$, $i = 1, \dots, N$ are i.i.d. random variables with common law $\mu_0$, then

$$E_V^N(\omega_1, \dots, \omega_N) = \int_{(S^k)'} V^N d(z_{\boldsymbol{\omega}}^N)^{\otimes k} = \mathcal{E}_V^N(z_{\boldsymbol{\omega}}^N) \qquad \mathbb{P}^{\otimes N}\text{-a.e.}$$

due to the non-atomic property of the reference measure $\mu_0$. Hence, by construction and the mean-field interaction property, the interacting particle system may then be recast as a

change of measure in $\mathcal{P}(\mathcal{P}(S))$, namely

$$Q_V^N := \frac{1}{Z_V^N} e^{-N\mathcal{E}_V^N(\mu)} P^N. \tag{3.2}$$

Given a Borel function $V : S^k \to \overline{\mathbb{R}}$, we define similarly the function $\mathcal{E}_V : \mathcal{P}(S) \to \overline{\mathbb{R}}$ by

$$\mathcal{E}_V(\mu) := \begin{cases} \int_{S^k} V \, d\mu^{\otimes k} & \text{if } V \in L^1(\mu^{\otimes k}), \\ +\infty & \text{otherwise,} \end{cases} \tag{3.3}$$

The functions $\mathcal{E}_V^N$ and $\mathcal{E}_V$ are Borel maps on $\mathcal{P}(S)$ (cf. Appendix C, in particular Lemma C.2 and Corollary C.4) and are defined identically except for the $N$ dependence in $V$ and the domain of integration, i.e., $(S^k)'$ instead of $S^k$.

**Remark 3.1.** A few comments on the assumption that $\mu_0$ is non-atomic:

(i) The energy $E_V^N$ as defined above does not rule out self-interaction, i.e., two particles occupying the same position ($x_i = x_j$ for $i \neq j$). While $E_V^N$ is meaningful for bounded potentials $V^N$, this may cause an issue when $V^N$ is singular on the diagonal. The non-atomicity of $\mu_0$ resolves this issue: indeed, if $\omega_i$ are i.i.d. random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ with common law $\mu_0$, then $\mathbb{P}(\omega_i = \omega_j,\ i \neq j) = 0$, which then allows $E_V^N(\omega_1, \dots, \omega_N)$ to be defined $\mathbb{P}$-a.s..

(ii) However, if the reference measure $\mu_0$ is atomic but the energy for the $N$-particle system is still given by $E_V^N$ there is also an alternative method. Namely, fix $N$ and note that $E_V^N(\boldsymbol{x}) = E_V^N(\mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in S^N$ with $z_{\boldsymbol{x}}^N = z_{\mathbf{y}}^N$. Hence the function $\mathcal{E}_V^N : \mathcal{P}(S) \to \overline{\mathbb{R}}$

$$\mathcal{E}_V^N(\mu) := \begin{cases} E^N((z_\bullet^N)^{-1}(\mu)) & \mu \in z_\bullet^N(S^N), \\ +\infty & \text{otherwise,} \end{cases}$$

is well defined. Moreover, it is easy to verify that $z_\bullet^N(S^N)$ is closed and for $V^N \in C_b(S^N)$ the map $E^N((z_\bullet^N)^{-1}(\mu))$ is continuous on $z_\bullet^N(S^N)$, hence $\mathcal{E}_V^N(\mu)$ is Borel. By a monotone class argument similar as in Appendix C one can extend this to all Borel $V^N$, and we can then proceed as in the rest of this section. However, note that $\mathcal{E}_V^N$ might no longer be of integral form as in (3.1).

Now we give the classical results concerning LDP. We start with the non-interacting case, namely Sanov's theorem:

**Theorem 3.2** (Sanov's theorem)**.** *The family $\{P^N\}$ of laws of the non-interacting particle system satisfies an LDP with rate function $I : \mathcal{X} \to \overline{\mathbb{R}}$, where*

$$I(\mu) := R(\mu \| \mu_0)$$

*is the relative entropy of $\mu \in \mathcal{X}$ with respect to $\mu_0$.*

We recall that the relative entropy is defined as

$$R(\nu \| \mu) := \begin{cases} \displaystyle\int_S \log\left(\frac{d\nu}{d\mu}\right) d\nu & \text{if } \nu \ll \mu \\ +\infty & \text{otherwise.} \end{cases}$$

For the LDP for the interacting particle systems, we introduce the following notation:

$$J_V(\mu) := \begin{cases} \mathcal{E}_V(\mu) + I(\mu) & \text{for } \mu \in D(I), \\ +\infty & \text{otherwise,} \end{cases}$$

with $D(I) := \{\mu \,|\, R(\mu\|\mu_0) < \infty\}$, and, if $\inf_{\mathcal{X}} J_V > -\infty$,

$$\mathcal{F}_V(\mu) := J_V(\mu) - \inf_{\nu\in\mathcal{X}} J_V(\nu). \tag{3.4}$$

We now give an LDP in the case when $V^N = V$ is in $C_b(S^k)$. In this case, $\mathcal{E}_V$ is also continuous and bounded, and so the LDP for $Q_V^N$ is essentially a consequence of the classical Varadhan Lemma. The only (and technical) difference with the classical Varadhan Lemma comes from the missing diagonal in $(S^k)'$, which in general causes $\mathcal{E}_V^N$ to not be continuous.

**Lemma 3.3.** *Suppose $V : S^k \to \mathbb{R}$ is continuous and bounded. Then $(\mathcal{E}_V^N, \mathcal{E}_V)$ induces an LDP (in the sense of Definition 2.5). In particular, the family $\{Q_V^N\}$ given by* (3.2) *satisfies an LDP with rate function $\mathcal{F}_V$.*

*Proof.* By applying Lemma C.2 $k$-times, we get that, for any continuous and bounded $V$, the function $\mathcal{E}_V$ is continuous and bounded on $\mathcal{P}(S)$. Hence, by the classical Varadhan Lemma (cf. Proposition 1.2), the couple $(\mathcal{E}_V, \mathcal{E}_V)$ induces an LDP in the sense of Definition 2.1.

Then, approximating $(\mathcal{E}_V^N, \mathcal{E}_V)$ with $(\mathcal{E}_V, \mathcal{E}_V)$, we have that $(\mathcal{E}_V^N, \mathcal{E}_V)$ induces an LDP by Theorem 2.8, provided we show that, for some $\gamma > 1$ and $K \in \mathbb{R}$,

$$\begin{aligned} &\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[ e^{-\gamma N \mathcal{E}_V^N(z_{\boldsymbol{\omega}}^N)} \right] < +\infty, \\ &\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[ e^{\beta N |\mathcal{E}_V^N - \mathcal{E}_V|(z_{\boldsymbol{\omega}}^N)} \right] \le K \qquad \text{for every } \beta \ge 0. \end{aligned} \tag{3.5}$$

The first limit follows easily from the boundedness of $V$. For the second limit, we can bound away all the self-interactions to obtain

$$\begin{aligned} \left|\mathcal{E}_V^N(z_{\boldsymbol{\omega}}^N) - \mathcal{E}_V(z_{\boldsymbol{\omega}}^N)\right| &= \frac{1}{N^k} \left| \sum_{i_1,\dots,i_k \text{ all distinct}} V(\omega_{i_1},\dots,\omega_{i_k}) - \sum_{i_1,\dots,i_k} V(\omega_{i_1},\dots,\omega_{i_k}) \right| \\ &\le \frac{1}{N^k} \frac{k(k-1)}{2} N^{k-1} \|V\|_\infty = \frac{k(k-1)}{2N} \|V\|_\infty. \end{aligned} \tag{3.6}$$

In the inequality above, we used that the number of $k$-tuples $(i_1, \dots i_k)$ with at least two equal indices is bounded by $N^{k-1}k(k-1)/2$. The second limit in (3.5) follows easily. □

Thus, we are in the same setting as in the previous section, i.e., we have created a large class of family of functions $(V^N, V)$ such that their induced interacting particle systems satisfy a certain LDP. Hence, next we will show how to extend this class by approximation.

3.2. **Main results.** We give our main result for this section, which serves as a tool for LDPs for Gibbs measures with a possibly discontinuous interaction potential. The result brings the general LDP result of Theorem 2.8 into the Gibbs measure context. In the following, $(f)^-$ denotes the negative part of the function $f$.

**Theorem 3.4.** *Let $(V_\lambda^N, V_\lambda)$ be a family of Borel functions on $S^k$ such that $(\mathcal{E}_{V_\lambda}^N, \mathcal{E}_{V_\lambda})$ induces an LDP. Let $(V^N, V)$ be a family of Borel functions on $S^k$ and assume that, for some $\gamma > 1$,*

$$\left.\begin{aligned} \limsup_{N\to\infty} \log \int_{S^k} e^{\gamma k |(V_\lambda^N)^-|}\, d\mu_0^{\otimes k} < +\infty \\ \log \int_{S^k} e^{\gamma k |(V_\lambda)^-|}\, d\mu_0^{\otimes k} < +\infty \end{aligned}\right\} \quad \text{for every } \lambda > 0, \tag{3.7}$$

*and that, for some $K \in \mathbb{R}$,*

$$\left.\begin{aligned} \limsup_{\lambda\to 0} \limsup_{N\to\infty} \log \int_{S^k} e^{\beta |V^N - V_\lambda^N|}\, d\mu_0^{\otimes k} \le K, \\ \limsup_{\lambda\to 0} \log \int_{S^k} e^{\beta |V - V_\lambda|}\, d\mu_0^{\otimes k} \le K. \end{aligned}\right\} \quad \text{for every } \beta \ge 0. \tag{3.8}$$

*Then $(\mathcal{E}_V^N, \mathcal{E}_V)$ induces an LDP. In particular, the family of induced interacting particle systems $Q_V^N$ satisfies an LDP with normalized rate function $\mathcal{F}_V$ given in* (3.4).

Similarly to Theorem 2.8 for general LDPs, informally this results states that $Q_V^N$ satisfy an LDP if there exists interacting potentials $V_\lambda^N$ and $V_\lambda$ which approximate $V^N$ and $V$ in an exponentially good way and whose corresponding interacting systems $Q_{V_\lambda}^N$ satisfy an LDP, for each $\lambda$. Again, this allows the following generalization:

(1) It allows $V$ to be discontinuous.
(2) The only requirement on LDPs for approximants is that $(\mathcal{E}_{V_\lambda}^N, \mathcal{E}_{V_\lambda})$ induces an LDP, not that $V_\lambda$ are continuous.
(3) We can allow for the potential $V^N$ in the interacting particle system to depend on $N \in \mathbb{N}$ (cf. Remark 3.7 below).

In the case where the sequence of functions $V^N$ is constant and equal to $V$, an LDP follows whenever $V$ satisfies the appropriate exponential moment condition.

**Corollary 3.5.** *Suppose that $V^N = V$ for all $N \in \mathbb{N}$, and such that*

$$\int_{S^k} e^{\beta |V|}\, d\mu_0^{\otimes k} < \infty \qquad \text{for all } \beta \ge 0. \tag{3.9}$$

*Then $(\mathcal{E}_V^N, \mathcal{E}_V)$ induces an LDP.*

*Proof.* The result follows from Lemma 3.3 and Theorem C.5. Indeed, by Theorem C.5, with the choice $\mu = \mu_0^{\otimes k}$, $X = S^k$, there exists a sequence $(V_\lambda) \subset C_b(S^k)$ such that

$$\lim_{\lambda\to 0} \log \int_{S^k} e^{\beta |V - V_\lambda|}\, d\mu_0^{\otimes k} = 0 \qquad \text{for any } \beta \ge 0.$$

Since the family $V_\lambda$ induces an LDP by Lemma 3.3, both (3.7) and (3.8) are satisfied. □

**Remark 3.6.** Recall that in the definition of the energy $E_V^N$ for the $N$-particle configuration we have excluded self-interaction, due to possible singularities in $V$. However, if $V$ is bounded (but not necessarily continuous), Corollary 3.5 remains valid when the energy does include self-interactions, i.e.,

$$E_V^N(x_1, \dots, x_N) := \frac{1}{N^k} \sum_{i_1,\dots,i_k} V(x_{i_1}, \dots, x_{i_k}).$$

In this case, $(\mathcal{E}_V^N, \mathcal{E}_V)$ induces an LDP with

$$\mathcal{E}_V^N(\mu) := \mathcal{E}_V(\mu) = \int_{S^k} V \, d\mu^{\otimes k} \qquad \text{for all } \mu \in \mathcal{P}(S^k).$$

Indeed, this holds simply due to the estimate (3.6).

**Remark 3.7.** The setting of Theorem 3.4 includes also the case of interactions among different number of particles. For example, let ($N$-independent) interaction potentials $U_k : S^\ell \to \mathbb{R}$, $k = 1, 2, 3$, be given and assume that the energy function $E_U^N$ is the sum of these three interactions, i.e.

$$E_U^N(x_1, \dots x_N) = \frac{1}{N^3} \sum_{i_1, i_2, i_3 \text{ distinct}} U_3(x_{i_1}, x_{i_2}, x_{i_3}) + \frac{1}{N^2} \sum_{i_1 \neq i_2} U_2(x_{i_1}, x_{i_2}) + \frac{1}{N} \sum_{i_1} U_1(x_{i_1}).$$

Therefore, by taking

$$V^N(x_1, x_2, x_3) = U_3(x_1, x_2, x_3) + \frac{N}{N-2} U_2(x_1, x_2) + \frac{N^2}{(N-1)(N-2)} U_1(x_1),$$

we see that $E_U^N = E_V^N$ (for $x_1, \dots x_N$ all distinct) and we are in the previous setting.

**Remark 3.8** (**Toward interacting diffusions**)**.** As for Theorem 2.8, we outline briefly how Theorem 3.4 will be applied in our context of singularly interacting diffusions, in the case of two-point interaction drift (1.2). Recall that we are interested in an LDP on the space $\mathcal{X} = \mathcal{P}(S)$, with $S = C([0, T]; \mathbb{R}^d))$, for the law $Q^N$ of the interacting particle system $\boldsymbol{X} = (X^{N,1}, \dots, X^{N,N})$ satisfying (1.1), as $N \to \infty$; the drift $b$ is of the form (1.2), the interaction kernel $\varphi$ being possibly singular. As we have written in the Introduction and in Remark 2.10, here we take $P^N$ as the law of the empirical measures $z_{\boldsymbol{W}}^N$ of $N$ independent $d$-dimensional Brownian motions $W^1, \dots W^N$. In the case of two-point interaction, the energy given formally by Girsanov's theorem takes the form (3.1) of a Gibbs potential:

$$\mathcal{E}(\mu) = \int_{S^3} V(x, y, z) \, d\mu(x) d\mu(y) d\mu(z),$$
$$V(x, y, z) = -\int_0^T \varphi(t, x_t - y_t) \cdot \, dx_t + \frac{1}{2} \int_0^T \varphi(t, x_t - y_t) \cdot \varphi(t, x_t - z_t) \, dt.$$

The approximating potentials $V_\lambda$ are defined as above for $V$ but replacing $\varphi$ with a smooth bounded approximation $\varphi_\lambda$; for each $\lambda > 0$, the corresponding energy $\mathcal{E}_{V_\lambda}$ induces an LDP by classical results for interacting diffusions with smooth drift. The first assumption in (3.8) reads in this context simply as

$$\limsup_{\lambda \to 0} \log \mathbb{E} \left[ e^{\beta |V - V_\lambda| (W^1, W^2, W^3)} \right] \leq K,$$

for some $K$ independent of $\beta$. This condition involves deterministic and stochastic integrals of three independent Brownian motions $W^1, W^2, W^3$ only, and, as we will see in Section 4 and Appendix D, can be checked using tools like Itô calculus for exponential martingales (Lemma D.2) and Khasminskii's Lemma D.5.

We come now to the proof of Theorem 3.4, which relies heavily on Theorem 2.8 and the following bounds. These bounds convert the approximation properties for $V$ into those for $\mathcal{E}_V$, needed to apply Theorem 2.8.

**Lemma 3.9.** *For any $k, N \in \mathbb{N}$ with $N > 1$, and nonnegative Borel functions $V$, $V^N : S^k \to \overline{\mathbb{R}}$, the following inequalities hold true:*

$$\frac{1}{N} \log \mathbb{E}\left[e^{N\mathcal{E}_V^N(z_{\boldsymbol{\omega}}^N)}\right] \leq \frac{1}{k} \log \int_{S^k} e^{k\frac{N}{N-1}V^N} d\mu_0^{\otimes k}, \tag{3.10a}$$

$$\sup_{\mu \in D} \big(\mathcal{E}_V(\mu) - R(\mu\|\mu_0)\big) \leq \frac{1}{k} \log \int_{S^k} e^{kV} d\mu_0^{\otimes k}, \tag{3.10b}$$

*where $D = \{\nu \in \mathcal{P}(S) \mid R(\nu\|\mu_0) < +\infty\}$.*

*Proof.* For the proof of (3.10a), we use the Hoeffding decomposition [Hoe63] for $\mathcal{E}_V^N$, which reads

$$\mathcal{E}_V^N(z_{\boldsymbol{\omega}}^N) = \frac{N!}{N^k(N-k)!} \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \frac{1}{[N/k]} \sum_{j=1}^{[N/k]} V^N(\omega_{\sigma(jk-k+1)}, \dots \omega_{\sigma(jk)}),$$

where $\mathcal{S}_N$ is the group of permutations of $\{1, \dots N\}$ and $[N/k]$ denotes the integer part of $N/k$. The point of this decomposition is to group the elements $V^N(\omega_{i_1}, \dots \omega_{i_k})$ into an average (over possible permutations $\sigma$) of averages (over $j$) of $O(N)$ independent elements (that is, for fixed $\sigma$, the elements within the nested average are independent).

By Jensen's inequality, applied to the exponential function and the average over $\sigma$,

$$\begin{aligned}
\mathbb{E}\left[e^{N\mathcal{E}_V^N(z_{\boldsymbol{\omega}}^N)}\right] &= \mathbb{E}\left[\exp\left[\frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \frac{N!}{N^k(N-k)!} \frac{N}{[N/k]} \sum_{j=1}^{[N/k]} V^N(\omega_{(\sigma(jk-k+1))}, \dots \omega_{\sigma(jk)})\right]\right] \\
&\leq \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \mathbb{E}\left[\exp\left[\frac{N!}{N^k(N-k)!} \frac{N}{[N/k]} \sum_{j=1}^{[N/k]} V^N(\omega_{(\sigma(jk-k+1))}, \dots \omega_{\sigma(jk)})\right]\right] \\
&= \mathbb{E}\left[\exp\left[\frac{N!}{N^k(N-k)!} \frac{N}{[N/k]} \sum_{j=1}^{[N/k]} V^N(\omega_{jk-k+1}, \dots \omega_{jk})\right]\right],
\end{aligned}$$

where we used the fact that that $\omega_i$ are exchangeable (as i.i.d.) in the last line. Since the random variables $(\omega_{jk-k+1}, \dots \omega_{jk})$ are independent in $j$, we have therefore

$$\begin{aligned}
\mathbb{E}\left[e^{N\mathcal{E}_V^N(z_{\boldsymbol{\omega}}^N)}\right] &\leq \mathbb{E}\left[\prod_{j=1}^{[N/k]} \exp\left[\frac{N!}{N^k(N-k)!} \frac{N}{[N/k]} V^N(\omega_{jk-k+1}, \dots \omega_{jk})\right]\right] \\
&= \prod_{j=1}^{[N/k]} \mathbb{E}\left[\exp\left[\frac{N!}{N^k(N-k)!} \frac{N}{[N/k]} V^N(\omega_{jk-k+1}, \dots \omega_{jk})\right]\right] \\
&= \mathbb{E}\left[\exp\left[\frac{N!}{N^k(N-k)!} \frac{N}{[N/k]} V^N(\omega_1, \dots \omega_k)\right]\right]^{[N/k]},
\end{aligned}$$

where we used again that $\omega_i$ are exchangeable in the second equality. Since $N! \leq N^k(N-k)!$ and $N/[N/k] \leq Nk/(N-1)$ for every $k, N \in \mathbb{N}$, $N > 1$, we then obtain

$$\mathbb{E}\left[e^{N\mathcal{E}_V^N(z_{\boldsymbol{\omega}}^N)}\right] \leq \mathbb{E}\left[\exp\left[\frac{N}{N-1} k\, V^N(\omega_1, \dots \omega_k)\right]\right]^{[N/k]}.$$

Taking the logarithm and noting that $[N/k] \leq N/k$ for every $k, N \in \mathbb{N}$ yields

$$\frac{1}{N}\log \mathbb{E}\left[e^{N\mathcal{E}_V^N(z_{\boldsymbol{\omega}}^N)}\right] \leq \frac{1}{k}\log \mathbb{E}\left[\exp\left[\frac{N}{N-1}k\, V^N(\omega_1,\dots\omega_k)\right]\right],$$

which concludes the proof of (3.10a).

To prove (3.10b), we first recall the additivity property of the relative entropy, i.e.

$$R(\mu^{\otimes k}\|\mu_0^{\otimes k}) = kR(\mu\|\mu_0).$$

Hence, for any $\mu \in D$, we have that

$$k\Big(\mathcal{E}_V(\mu) - R(\mu\|\mu_0)\Big) = \int_{S^k} kV\, d\mu^{\otimes k} - R(\mu^{\otimes k}\|\mu_0^{\otimes k}) \leq \sup_{\nu\in D^k}\left\{\int_{S^k} kV\, d\nu - R(\nu\|\mu_0^{\otimes k})\right\},$$

where $D^k = \{\nu \in \mathcal{P}(S^k) \mid R(\nu\|\mu_0^{\otimes k}) < +\infty\}$. By Lemma B.1 we have that

$$\sup_{\nu\in D^k}\left\{\int_{S^k} kV d\nu - R(\nu\|\mu_0^{\otimes k})\right\} = \log\int_{S^k} e^{kV}\, d\mu_0^{\otimes k}.$$

Therefore, if the right-hand side is finite, the desired estimate (3.10b) follows. □

*Proof of Theorem 3.4.* In order to apply Theorem 2.8, we have to show that there exists some $\gamma > 1$, such that for every $\lambda > 0$,

$$\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-\gamma N\mathcal{E}_{V_\lambda}^N(z_{\boldsymbol{\omega}}^N)}\right] < +\infty, \tag{3.11a}$$

$$\inf_{\mu\in D}\big(R(\mu\|\mu_0) + \gamma\mathcal{E}_{V_\lambda}(\mu)\big) > -\infty. \tag{3.11b}$$

and that, for some $K \in \mathbb{R}$, the following holds true for all $\beta \geq 0$:

$$\limsup_{\lambda\to 0}\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{\beta N|\mathcal{E}_V^N - \mathcal{E}_{V_\lambda}^N|(z_{\boldsymbol{\omega}}^N)}\right] \leq K, \tag{3.12a}$$

$$\limsup_{\lambda\to 0}\sup_{\mu\in D}\big(\beta|\mathcal{E}_V - \mathcal{E}_{V_\lambda}|(\mu) - R(\mu\|\mu_0)\big) \leq K. \tag{3.12b}$$

Due to linearity and (3.10a) of Lemma 3.9, we have for some $\gamma' \in (1,\gamma)$:

$$\frac{1}{N}\log\mathbb{E}\left[e^{-\gamma' N\mathcal{E}_{V_\lambda}^N(z_{\boldsymbol{\omega}}^N)}\right] \leq \frac{1}{N}\log\mathbb{E}\left[e^{\gamma' N\mathcal{E}^N_{\gamma'|(V_\lambda)^-|}(z_{\boldsymbol{\omega}}^N)}\right] \leq \frac{1}{k}\log\int_{S^k} e^{k\frac{N}{N-1}\gamma'|(V_\lambda^N)^-|}\, d\mu_0^{\otimes k}.$$

Since $\gamma' N/(N-1) \leq \gamma$ for all $N \geq N_\gamma := \gamma/(\gamma-\gamma')$, we obtain from assumption (3.7) the finiteness of the right-hand side uniformly in $N$ for $N \geq N_\gamma$. Hence, taking the lim sup yields (3.11a).

As for (3.11b), we apply (3.10b) of Lemma 3.9 to any $\mu \in \mathcal{P}(S)$ with $R(\mu\|\mu_0) < +\infty$, to obtain

$$R(\mu\|\mu_0) + \gamma'\mathcal{E}_{V_\lambda}(\mu) \geq R(\mu\|\mu_0) - \mathcal{E}_{\gamma'|(V_\lambda)^-|}(\mu) \geq -\frac{1}{k}\log\int_{S^k} e^{\gamma' k|(V_\lambda)^-|}\, d\mu_0^{\otimes k} > -\infty.$$

Similarly, we apply Lemma 3.9 to obtain

$$\frac{1}{N}\log\mathbb{E}\left[e^{\beta N|\mathcal{E}_V^N - \mathcal{E}_{V_\lambda}^N|(z_{\boldsymbol{\omega}}^N)}\right] \leq \frac{1}{k}\log\int_{S^k} e^{k\frac{N}{N-1}\beta|V^N - V_\lambda^N|}\, d\mu_0^{\otimes k},$$

$$\sup_{\mu\in D}\big(\beta|\mathcal{E}_V - \mathcal{E}_{V_\lambda}| - I\big)(\mu) \leq \frac{1}{k}\log\int_{S^k} e^{k\beta|V - V_\lambda|}\, d\mu_0^{\otimes k},$$

which by assumption (3.8) yields (3.12a) and (3.12b). The conclusion follows applying Theorem 2.8. □

***Background.*** As mentioned in Section 2, various extension principles to singular functionals or contractions have arisen. See for example [ES98] on various generalizations of Sanov's theorem, in which a stronger topology is used—defined by the property that all functions $\mathcal{E}_V$ for which $k = 1$ and $V$ satisfies (3.9) are continuous with respect to this topology. This was subsequently generalized in [ES02] to the case $k \geq 2$. Note that an application of the classical Varadhan Lemma to the result in [ES02] should yield a very similar result to our Corollary 3.5; this type of argument has been used in [EZ03, Lemma 2.4] in the proof of a moderate deviation principle for a bounded interaction kernel.

More recently, in [LW20], a similar setting as this section was studied, i.e., large deviations for mean field Gibbs measures on Polish spaces, involving singular potentials. In this work they assume so-called *strong exponential integrability* of the negative part $V^-$ of the interaction potential $V$, and require a entropic inequality involving the positive part $V^+$ of the interaction potential. It is straightforward to verify that strong exponential integrability of $|V|$ is in fact a stronger assumption and is equivalent to (3.9), and in particular their results include the case of Corollary 3.5. In a sense this allows them to treat stronger repulsive singularities than the ones considered here. It should be noted that while both their work and Corollary 3.5 are not sufficient for the general setting of interacting diffusions as in Section 4, a similar assymetric generalization involving variational or weak convergence techniques for Theorem 4.5 might be possible, which we leave for future work.

Other LDP results for Gibbs measures with singular potentials have been proven, see e.g. [Ber18,CGZ14,DLR20,Rey18], with locally compact base space $S$, [GZ19] with potentials satisfying a bound from below and a lower semi-continuity assumption. In particular in [Ber18] the case where (3.10) only holds for some $\beta$ instead of all—which, in $\mathbb{R}^d$, allows for potentials $V$ with a logarithmic singularity—is considered, on compact Polish spaces and assuming $V$ lower semi-continuous.

It should be noted that inequalities related to (3.10a), which in our case is derived from the Hoeffding decomposition has also arisen in different context and different names. For example, in [ES02] a similar inequality stems from the existence of regular partitions of complete hypergraphs due to Baranyai, and in [LW20] modifications of decoupling inequalities of de la Peña were used. Hoeffding decomposition has been used directly in some generalizations to [ES02], for example [Eic04].

3.3. **Extension to Gibbs-like potentials.** The previous results use the Gibbs structure of the potential $\mathcal{E}_V$ to reduce the LDP problem to the context of Section 2. However, for this reduction to hold, only some Gibbs-like bounds are needed. This allows to prove LDPs for empirical measures not coming from Gibbs laws, as soon as Gibbs-type bounds are possible. As we will see in the example of Subsection 4.3.3, this is the case of interacting diffusions where the drift depends non-linearly on the empirical measure and/or on its $k$-times tensor product.

As before, let $S$ be a Polish space with its Borel $\sigma$-algebra $\mathcal{B}(S)$, $\mu_0 \in \mathcal{P}(S)$ be a given reference measure. The state space for the LDP is $\mathcal{X} = \mathcal{P}(S)$, equipped with the weak topology. As before, $P^N$ denotes the law of the empirical measure $z^N_{\boldsymbol{\omega}}$, where $\omega_i$, $i = 1, \dots N$ are i.i.d. random variables defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with values in $S$ and common law $\mu_0$; $\mathbb{E}$ denotes the expectation with respect to $\mathbb{P}$. Now let $\mathcal{E}^N : \mathcal{P}(S) \to \overline{\mathbb{R}}$,

$N \in \mathbb{N}$ and $\mathcal{E} : \mathcal{P}(S) \to \overline{\mathbb{R}}$ be Borel functions, let $Q^N \ll P^N$ be the probability measure on $\mathcal{P}(S)$ given by

$$Q^N = \frac{1}{Z^N} e^{-N\mathcal{E}^N(\mu)} P^N,$$

where $Z^N$ is the renormalization constant, assumed to be finite. We further recall the notations $\mathcal{E}_V^N(\mu)$ and $\mathcal{E}_V^N(\mu)$ given in (3.1) and (3.3) for any Borel function $V : S^k \to \overline{\mathbb{R}}$. As before, we denote $D = \{\nu \in \mathcal{P}(S) \mid R(\nu\|\mu_0) < \infty\}$.

Our main result is the following:

**Theorem 3.10.** *Let $(\mathcal{E}_\lambda^N, \mathcal{E}_\lambda)$ be a family of LDP inducing pairs for all $\lambda > 0$ such that (2.4) holds for some $\gamma > 1$ (independent of $\lambda$). Assume that, for every $\lambda > 0$ and every $\beta \in \mathbb{R}$, for every $\mu$ in $D$, there holds*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{\beta N(\mathcal{E}^N - \mathcal{E}_\lambda^N)(z_{\boldsymbol{\omega}}^N)}\right] \le C + C \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{c_\beta N \mathcal{E}_{G_\lambda^N}^N(z_{\boldsymbol{\omega}}^N)}\right], \tag{3.13a}$$

(3.13b)

$$|\beta(\mathcal{E} - \mathcal{E}_\lambda)(\mu)| - R(\mu\|\mu_0) \le C + C \log \int_S \exp\left(c_\beta \int_{S^{k-1}} G_\lambda(x,y)\, d\mu^{\otimes(k-1)}(y)\right) d\mu_0(x),$$

*for some constant $C > 0$ independent of $\beta$, $\lambda$ and $\mu$, some $c_\beta \ge 0$ independent of $\lambda$ and $\mu$, and some nonnegative Borel functions $G_\lambda^N, G_\lambda : S^k \to \overline{\mathbb{R}}$ (independent of $\beta$ and $\mu$). Assume also that $G_\lambda^N$ and $G_\lambda$ satisfy, for some $K \in \mathbb{R}$ independent of $\beta$, for every $\beta \ge 0$,*

$$\limsup_{\lambda\to 0} \limsup_{N\to\infty} \int_{S^k} e^{\beta G_\lambda^N} d\mu_0^{\otimes k} \le K, \tag{3.14a}$$

$$\limsup_{\lambda\to 0} \int_{S^k} e^{\beta G_\lambda} d\mu_0^{\otimes k} \le K. \tag{3.14b}$$

*Then the pair $(\mathcal{E}^N, \mathcal{E})$ induces an LDP with the normalized rate function $\mathcal{F}_V$ in (3.4).*

**Remark 3.11.** A simple condition for the inequalities (3.13a) and (3.13b) to hold is if

$$\left|\mathcal{E}^N - \mathcal{E}_\lambda^N\right|(\mu) \le \mathcal{E}_{G_\lambda^N}^N(\mu), \qquad |\mathcal{E} - \mathcal{E}_\lambda|(\mu) \le \mathcal{E}_{G_\lambda}(\mu) \qquad \text{for all } \mu \in \mathcal{P}(\mathcal{X}).$$

**Remark 3.12 (Toward interacting diffusions).** As for Theorem 2.8, we outline briefly how Theorem 3.10 will be applied in our context of singularly interacting diffusions, here in the case of a quite generic interaction drift, as in (1.4) with $\varphi$ possibly singular. Recall that we are interested in an LDP on the space $\mathcal{X} = \mathcal{P}(S)$, with $S = C([0,T];\mathbb{R}^d))$, for the law $Q^N$ of the interacting particle system $\boldsymbol{X} = (X^{N,1}, \dots, X^{N,N})$ satisfying (1.1), as $N \to \infty$. As we have written in the Introduction and in Remark 2.10, here we take $P^N$ as the law of the empirical measures $z_{\boldsymbol{W}}^N$ of $N$ independent $d$-dimensional Brownian motions $W^1, \dots W^N$; the energy given formally by Girsanov's theorem is

$$\mathcal{E}(\mu) = \int_S V(x,\mu)\, d\mu(x),$$
$$V(x,\mu) = -\int_0^T b_t(x_t,\mu_t) \cdot dx_t + \frac{1}{2}\int_0^T |b_t(x_t,\mu_t)|^2\, dt.$$

The approximating energies $\mathcal{E}_\lambda$ and potentials $V_\lambda$ are defined as above but replacing $b$ with a smooth bounded approximation $b_\lambda$; for each $\lambda > 0$, the corresponding energy $\mathcal{E}_\lambda$ induces

an LDP by classical results for interacting diffusions with smooth drift. The Gibbs-like assumption (3.13a) reads in this context: for every $\beta \in \mathbb{R}$,

$$\begin{aligned}&\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{\beta \sum_{i=1}^N [V(W^i, z^N_{\boldsymbol{W}}) - V_\lambda(W^i, z^N_{\boldsymbol{W}})]}\right]\\ &\leq C + C \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{c_\beta N^{-k+1} \sum_{1\leq j_1,\dots j_k \leq N} G^N_\lambda (W^{j_1},\dots W^{j_k})},\right]\end{aligned}$$

with $C$ independent of $\beta$ and $\lambda$ and $c_\beta$ independent of $\lambda$. The assumptions (3.14a) and (3.14b) read in this context, taking $G^N_\lambda = G_\lambda$ for simplicity: for every $\beta \geq 0$,

$$\limsup_{\lambda\to 0} \log \mathbb{E}\left[e^{\beta G_\lambda (W^1,\dots W^k)}\right] \leq K,$$

with $K$ independent of $\beta$.

As we will see in Section 4 and Appendix D, by using Itô calculus for exponential martingales (Lemma D.2), Theorem 4.5 shows the Gibbs-like assumption (3.13a) for a convenient function $G_\lambda$; for such $G_\lambda$, the conditions (3.14a) and (3.14b) can be checked by tools like Khasminskii's Lemma D.5.

*Proof of Theorem 3.10.* The proof is similar to that of Theorem 3.4. The result follows from Theorem 2.8 provided we verify condition (2.5). For condition (2.5a), by assumption (3.13a) and Lemma 3.9, we have

$$\begin{aligned}\limsup_{\lambda\to 0} \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{\beta N(\mathcal{E}^N - \mathcal{E}_\lambda)(z^N_{\boldsymbol{\omega}})}\right] &\leq C + C \limsup_{\lambda\to 0} \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{c_\beta N \mathcal{E}^N_{G_\lambda}(z^N_{\boldsymbol{\omega}})}\right]\\ &\leq C + C \limsup_{\lambda\to 0} \limsup_{N\to\infty} \frac{1}{k} \log \int_{S^k} e^{k \frac{N}{N-1} c_\beta G^N_\lambda} d\mu_0^{\otimes k},\end{aligned}$$

for all $\beta \in \mathbb{R}$. Hence, by assumption (3.14a), we get

$$\limsup_{\lambda\to 0} \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{\beta N(\mathcal{E}^N - \mathcal{E}^N_\lambda)(z^N_{\boldsymbol{\omega}})}\right] \leq C + C\frac{\log K}{k}.$$

For condition (2.5b), by assumption (3.13b) and Lemma B.4, we have, for every $\mu$ in $D$, for every $\lambda > 0$ and every $\beta \in \mathbb{R}$,

$$\begin{aligned}&|\beta(\mathcal{E} - \mathcal{E}_\lambda)(\mu)| - (1 + C(k-1))R(\mu\|\mu_0)\\ &\leq -C(k-1)R(\mu\|\mu_0) + C + C \log \int_S \exp\left(c_\beta \int_{S^{k-1}} G_\lambda(x,y)\, d\mu^{\otimes(k-1)}(y)\right) d\mu_0(x)\\ &\leq -C(k-1)R(\mu\|\mu_0) + C + C(k-1)R(\mu\|\mu_0) + C \log \int_{S^k} e^{c_\beta G_\lambda} d\mu_0^{\otimes k}\end{aligned}$$

Hence, taking the sup over $\mu$ in $D$ and then the lim sup over $\lambda$, by assumption (3.14b) we get

$$\limsup_{\lambda\to 0} \sup_{\mu\in D} |\beta(\mathcal{E} - \mathcal{E}_\lambda)(\mu)| - (1 + C(k-1))R(\mu\|\mu_0) \leq C + C\log K.$$

Condition (2.5b) follows by simply dividing by $1 + C(k-1)$. The proof is complete. □

**Remark 3.13.** As the above proof shows (and as consequence of Lemmas 3.9 and B.4), the assumptions (3.14a) and (3.14b) imply, respectively, the following two bounds:

$$\limsup_{\lambda\to 0}\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{c_\beta N\mathcal{E}^N_{G_\lambda}(z^N_{\boldsymbol{\omega}})}\right]\le\frac{\log K}{k}<\infty,$$

$$\limsup_{\lambda\to 0}\log\int_S\exp\left(c_\beta\int_{S^{k-1}}G_\lambda(x,y)\,d\mu^{\otimes(k-1)}(y)\right)d\mu_0(x)\le(k-1)R(\mu\|\mu_0)+\log K<\infty.$$

## 4. Singularly interacting diffusions

4.1. **Notations and preliminary results.** In this section we study LDPs for a system of mean field interacting diffusions, defined by a singular drift. We will put this problem in the framework of Gibbs-like structure of the previous section (cf. Section 3.3).

For $N\in\mathbb{N}$, we consider the following system of interacting SDEs

$$\begin{cases} dX^{N,i}_t=b^N_t\left(X^{N,i}_t,\dfrac{1}{N}\displaystyle\sum_{j=1}^N\delta_{X^{N,j}_t}\right)dt+dW^i_t,\quad i=1,\dots N,\\ X^{N,i}_0 \text{ i.i.d with law } \rho_0.\end{cases}\tag{4.1}$$

Here the *drift* $b^N:[0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)\to\mathbb{R}^d$ is a Borel map, where $\mathcal{P}(\mathbb{R}^d)$ is endowed with the (metrizable, complete and separable) topology of weak convergence. The processes $W^i$, $i\in\mathbb{N}$, are independent $d$-dimensional Brownian motions on a standard filtered probability space $(\Omega,\mathcal{A},(\mathcal{F}_t)_t,\mathbb{P})$, $\rho_0\in\mathcal{P}(\mathbb{R}^d)$ is the law of the i.i.d. initial data $X^{N,i}_0$.

We now set $S:=C([0,T];\mathbb{R}^d)$, the space of continuous paths in $\mathbb{R}^d$, and note that $X^{N,i}\in S$ for each $i=1,\dots,N$. Moreover, we set $\tilde{Q}^N_{b^N}\in\mathcal{P}(S^N)$ to be the law of $\boldsymbol{X}^N:=(X^{N,1},\dots,X^{N,N})$ defined by the system (4.1). We set $\mathbb{W}\in\mathcal{P}(S)$ to be the Wiener measure with $\rho_0$ as marginal at time 0 and $\tilde{P}^N\in\mathcal{P}(S^N)$ the law of $N\in\mathbb{N}$ independent Brownian motions, i.e. $\tilde{P}^N:=\mathbb{W}^{\otimes N}$. With a little abuse of notation, we will consider $W^i$ to be $d$-dimensional independent Brownian motions with initial law $\rho_0$, unless differently specified; similarly, we will use $W$ for a $d$-dimensional Brownian motion with initial law $\rho_0$, unless differently specified.

Consider the *empirical process* $z^N_{\boldsymbol{X}}\in\mathcal{P}(S)$,

$$z^N_{\boldsymbol{X}}:=\frac{1}{N}\sum_{i=1}^N\delta_{X^{N,i}},$$

and let $Q^N_{b^N}\in\mathcal{P}(\mathcal{P}(S))$ be the laws of the random variable $z^N_{\boldsymbol{X}}$ induced by $\tilde{Q}^N_{b^N}$. Similarly, let $P^N$ be the law for the empirical process $z^N_{\boldsymbol{W}}$ of the non-interacting system induced by $\tilde{P}^N$. We will use $z^N_{\boldsymbol{x}}$, for $\boldsymbol{x}\in S^N$, to denote the empirical measure associated with $\boldsymbol{x}$, and $z^N$ for the canonical process (that is, the identity process) on $\mathcal{P}(S)$. We will often use the notation $\langle f,\mu\rangle$ to denote $\int f\,d\mu$.

Recall that, as in Section 3, the sequence $P^N$ satisfies an LDP with rate function

$$I:\mathcal{P}(S)\to\overline{\mathbb{R}};\qquad I(\mu):=R(\mu\|\mathbb{W}),$$

where $R(\mu\|\mathbb{W})$ is the relative entropy of $\mu$ with respect to the Wiener measure $\mathbb{W}$.

For the particle system, a Gibbs-like representation holds. Indeed, for a Borel map $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$, we introduce the potentials

$$V_b^N(x,\mu) := -V_b^{N,2}(x,\mu) + \frac{1}{2}V_b^1(x,\mu) \quad \text{with} \quad \begin{cases} V_b^1(x,\mu) := \int_0^T |b_t(x_t,\mu_t)|^2 dt, \\ V_b^{N,2}(x,\mu) := \int_0^T b_t(x_t,\mu_t)\cdot dx_t, \end{cases} \tag{4.2}$$

where $V_b^{N,2} : S \times \mathcal{P}(S) \to \mathbb{R}$ is defined as stochastic integral under the law of $(W^i, z^N_{\boldsymbol{W}})$ on $(x,\mu)$. We define the corresponding log-density as

$$\mathcal{E}_b^N(\mu) := \begin{cases} \int_S V_b^N(x,\mu)\, d\mu(x) & \text{if } V_b^1(\cdot,\mu), V_b^{N,2}(\cdot,\mu) \in L^1(\mu), \\ 0 & \text{otherwise,} \end{cases} \qquad \mu \in \mathcal{P}(S).$$

See Section D.2 for the precise definition and measurability properties of $\mathcal{E}_b^N$. Similarly we define $V_b$ and $\mathcal{E}_b$ by replacing $V_b^{N,2}$ with

$$V_b^2(x,\mu) := \int_0^T b_t(x_t,\mu_t)\cdot dx_t,$$

now as stochastic integral at a deterministic $\mu$. See again Section D.2 for the precise definition of $\mathcal{E}_b$ and its measurability properties. The Gibbs-like representation of the particle system is given by the following lemma, which is essentially a consequence of Girsanov's theorem and the mean field form of the interaction ($b^N$ depending on $z^N_{\boldsymbol{X}}$), the details of the proof are postponed to Subsection D.3.

**Lemma 4.1.** *Fix $N \in \mathbb{N}$. Assume that*

$$\mathbb{E}\left[\exp\left(\frac{N}{2}\int_S\int_0^T |b_t^N(x_t, z^N_{\boldsymbol{W},t})|^2 dt\, dz^N_{\boldsymbol{W},t}(x)\right)\right] < \infty.$$

*Then there exists a weak solution to the system* (4.1)*, which is unique under the constraint*

$$\int_0^T |b_t^N(x_t, z_t^N)|^2 dt \in L^1(S, z^N) \qquad \text{for } Q^N_{b^N}\text{-a.e. } z^N. \tag{4.3}$$

*For this law, we have the following representations:*

$$\frac{d\tilde{Q}^N_{b^N}}{d\tilde{P}^N}(x^1,\dots x^N) = e^{-N\mathcal{E}^N_{b^N}(z^N_{\boldsymbol{x}})}, \qquad \frac{dQ^N_{b^N}}{dP^N}(\mu) = e^{-N\mathcal{E}^N_{b^N}(\mu)}. \tag{4.4}$$

Now we provide a class of drifts $(b^N, b)$ which induce an LDP for the law $Q^N_{b^N}$. Morally, this is the class of Lipschitz drifts (with respect to the 1-Wasserstein distance). The class of $\mathcal{F}$Lip-inducing pairs in the definition below allows for some margin to include the case of drifts without self-interactions, similarly to Lemma 3.3. In the following, $\mathcal{P}_1(\mathbb{R}^d)$ denotes the subset of $\mathcal{P}(\mathbb{R}^d)$ of all probability measures on $\mathbb{R}^d$ with finite first moment; $W_1$ denotes the 1-Wasserstein distance on $\mathcal{P}_1(\mathbb{R}^d)$.

**Definition 4.2.** The class $\mathcal{F}$Lip consists bounded Borel functions $b : [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ such that the map $(x,\mu) \mapsto b(t,x,\mu)$ is globally Lipschitz continuous on $\mathcal{P}_1(\mathbb{R}^d)$ with respect to the 1-Wasserstein distance, uniformly in $t \in [0,T]$, i.e., for some $M_b \geq 0$,

$$|b(t,x,\mu) - b(t,y,\nu)| \leq M_b\big(|x-y| + W_1(\mu,\nu)\big),$$

for all $t \in [0,T]$, $(x,y) \in \mathbb{R}^d$ and $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$.

Moreover, the pair $(\{b^N\}, b)$ (in short $(b^N, b)$) is called $\mathcal{F}$Lip-inducing (subjected to $\{P^N\}$) if it satisfies the following conditions:

(1) $b \in \mathcal{F}\mathrm{Lip}$;
(2) The sequence $b^N$ is uniformly bounded, i.e. $\sup_{N\in\mathbb{N}} \|b^N\|_\infty < \infty$;
(3) There exists a sequence $c_N$ with $c_N \to 0$ as $N \to \infty$ such that

$$\int_0^T \left\langle |b_t - b_t^N|^2(\cdot, z_t^N), z_t^N \right\rangle dt \le c_N, \qquad \text{for } P^N\text{-almost every } z^N. \tag{4.5}$$

**Lemma 4.3.** *Assume that the initial law $\rho_0$ satisfies*

$$\int_{\mathbb{R}^d} e^{\beta|x|}\, d\rho_0(x) < \infty, \qquad \forall \beta > 0. \tag{4.6}$$

*Assume that $(b^N, b)$ is $\mathcal{F}$Lip-inducing. Then $(\mathcal{E}_{b^N}^N, \mathcal{E}_b)$ induces an LDP.*

**Remark 4.4.** The exponential condition (4.6) is required only to apply [CDFM20, Theorem 34]: that result needs (4.6) because it works with the 1-Wasserstein topology instead of the weak topology (the LDP in weak topology follows then by contraction principle). For this reason we suspect that, in the space $\mathcal{P}(S)$ with the weak topology, the condition (4.6) is not necessary.

*Proof of Lemma 4.3.* The LDP induced by $(\mathcal{E}_b^N, \mathcal{E}_b)$ follows from [CDFM20, Theorem 34] (see also [Fis14]). There an LDP is proved for $Q_b^N$ with rate function $R(\mu\|\mathbb{W}^\mu)$, where $\mathbb{W}^\mu$ is defined as in Theorem 4.5 (as the law of the solution to (4.12)). By Lemma D.4, we have

$$R(\mu\|\mathbb{W}^\mu) = R(\mu\|\mathbb{W}) - \mathcal{E}_b(\mu).$$

In particular, the infimum of the $R(\cdot\|\mathbb{W}) - \mathcal{E}_b$ is the infimum of the left-hand side above, that is 0. The Laplace principle is then trivially satisfied as $e^{-N\mathcal{E}_b}(\mu)$ is the density of the Girsanov's transform.

Hence, note that by Theorem 2.8 it is enough to show (2.4) and (2.5a) for $\mathcal{E}_\lambda^N := \mathcal{E}_{b^N}^N$ and $\mathcal{E}_\lambda = \mathcal{E}^N = \mathcal{E} := \mathcal{E}_b$. To derive (2.4a), we have for any $\gamma \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}\left[e^{\gamma N \langle V_{b^N}^{N,2}(\cdot, z_{\boldsymbol{W}}^N), z_{\boldsymbol{W}}^N\rangle}\right] &= \mathbb{E}\left[e^{\gamma \sum_{i=1}^N \int_0^T b_t^N(W_t^i, z_{\boldsymbol{W},t}^N)\cdot dW_t^i}\right] \\ &\le \mathbb{E}\left[e^{2\gamma^2 \sum_{i=1}^N \int_0^T |b_t^N(W_t^i, z_{\boldsymbol{W},t}^N)|^2 dt}\right]^{\frac{1}{2}} \le e^{\gamma^2 NT\|b^N\|_\infty^2}. \end{aligned} \tag{4.7}$$

where Lemma D.2 was applied in the first inequality. Since $V_{b^N}^1$ is trivially bounded by $T\|b^N\|_\infty^2$, we then obtain

$$\mathbb{E}\left[e^{-\gamma N \mathcal{E}^N(z_{\boldsymbol{W}}^N)}\right] \le e^{(\gamma^2+\gamma/2)NT\|b^N\|_\infty^2}.$$

Since $\sup_{N\in\mathbb{N}} \|b^N\|_\infty := M < \infty$, (2.4a) is satisfied.

Moreover, using Lemma B.1 we have for any $\gamma > 1$ and any $\mu \in \mathcal{P}(S)$ with $R(\mu\|\mathbb{W}) < \infty$, that

$$\gamma \mathcal{E}_b(\mu) \le R(\mu\|\mathbb{W}) + \log \mathbb{E}\left[e^{\gamma|V_b|(W,\mu)}\right].$$

In a similar fashion as before, we can estimate the second term on the right-hand side uniformly in $\mu$ to obtain

$$R(\mu\|\mathbb{W}) + \gamma\mathcal{E}_b(\mu) \geq -a,$$

for some constant $a_0 > 0$ independent of $\mu \in \mathcal{P}(S)$, which gives (2.4b).

For (2.5a), we will consider the part of $\mathcal{E}^N_{b^N} - \mathcal{E}^N_b$ determined by $V^1$ and $V^{N,2}$ separately. First, note that $V^{N,2}_{b^N} - V^{N,2}_b = V^{N,2}_{b^N-b}$ and similar to the argument of (4.7) we have for all $\beta \in \mathbb{R}$

$$\mathbb{E}\left[e^{\beta N\langle V^{N,2}_{b^N-b}(\cdot,z^N_{\boldsymbol{W}}),z^N_{\boldsymbol{W}}\rangle}\right] \leq \mathbb{E}\left[e^{2\beta^2\sum_{i=1}^N\int_0^T|b^N_t-b_t|^2(W^i_t,z^N_{\boldsymbol{W},t})dt}\right]^{1/2} \leq e^{c_N N\beta^2}.$$

Secondly, since

$$\beta\left||b^N|^2-|b|^2\right| \leq \beta(|b^N|+|b|)|b^N-b| \leq \tfrac{1}{2}(M+\|b\|_\infty)^2 + \tfrac{\beta^2}{2}|b^N-b|^2,$$

we derive

$$\begin{aligned}\mathbb{E}\left[e^{\beta N\left(\langle V^1_{b^N}(\cdot,z^N_{\boldsymbol{W}}),z^N_{\boldsymbol{W}}\rangle-\langle V^1_b(\cdot,z^N_{\boldsymbol{W}}),z^N_{\boldsymbol{W}}\rangle\right)}\right] &= \mathbb{E}\left[e^{\beta\sum_{i=1}^N\int_0^T\left(|b^N_t|^2-|b_t|^2\right)(W^i_t,z^N_{\boldsymbol{W},t})dt}\right]\\ &\leq \mathbb{E}\left[e^{\frac{TN}{2}(M+\|b\|_\infty)+\frac{\beta^2}{2}\sum_{i=1}^N\int_0^T\left(|b^N_t-b_t|^2\right)(W^i_t,z^N_{\boldsymbol{W},t})dt}\right]\\ &\leq e^{\frac{TN}{2}(M+\|b\|_\infty)+\frac{c_N N\beta^2}{2}}.\end{aligned}$$

Finally, via Cauchy-Schwarz, we conclude that for all $\beta \in \mathbb{R}$

$$\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{\beta\left(\mathcal{E}^N_{b^N}(z^N_{\boldsymbol{W}})-\mathcal{E}^N_b(z^N_{\boldsymbol{W}})\right)}\right] \leq \tfrac{T}{2}(M+\|b\|_\infty).$$

□

4.2. **The main result.** Now we give the main result of this section, which states the LDP for the system (4.1). The assumptions may seem involved at first glance, but their meaning is not difficult: we have an LDP for the system (4.1) as soon as we can approximate in a suitable way the drift $b$ by regular drifts $b_\lambda$, along the Brownian empirical measure $z^N_{\boldsymbol{W}}$ and the couple of Brownian path $W$ and measure $\mu$ with finite relative entropy.

**Theorem 4.5.** *Assume the condition* (4.6) *on the initial law* $\rho_0$. *Suppose there exists a sequence of* $\mathcal{F}$Lip*-inducing drifts* $(b^N_\lambda, b_\lambda)_{\lambda>0}$ *and a sequence* $(g_\lambda)_{\lambda>0}$ *of Borel functions* $g_\lambda : [0,T]\times(\mathbb{R}^d)^k \to [0,+\infty)$, $k \in \mathbb{N}$, *such that the following conditions hold:*

*(i) for every* $\lambda > 0$, *for* $P^N$*-almost every* $z^N \in \mathcal{P}(S)$,

$$\int_0^T\left\langle|b^N_t - b^N_{\lambda,t}|^2(\cdot,z^N_t),z^N_t\right\rangle dt \leq \int_0^T\int_{((\mathbb{R}^d)^k)'} g_\lambda(t,x_1,\dots,x_k)\,d(z^N_t)^{\otimes k}\,dt; \tag{4.8}$$

*(ii) for every* $\lambda > 0$ *and every* $\mu \in \mathcal{P}(S)$ *with* $R(\mu\|\mathbb{W}) < \infty$ *and* $\mathbb{W}$*-almost every* $W$,

$$\int_0^T|b_t - b_{\lambda,t}|^2(W_t,\mu_t)\,dt \leq \int_0^T\int_{(\mathbb{R}^d)^k} g_\lambda(t,W_t,y)\,d\mu_t^{\otimes k-1}(y)\,dt. \tag{4.9}$$

*Suppose also that* $(g_\lambda)_{\lambda>0}$ *satisfies for some* $K>0$,

$$\limsup_{\lambda\to 0}\mathbb{E}\left[e^{\beta\int_0^T g_\lambda(t,W_t^1,\dots,W_t^k)\,dt}\right]\le K,\qquad \forall\beta\in\mathbb{R}, \tag{4.10}$$

*where* $W^1,\dots,W^k$ *are independent Brownian motions with common initial law* $\rho_0$.

*Then the family* $\{Q^N_{b^N}\}$ *of laws of* $z^N_{\boldsymbol{X}}$ *(for* $\boldsymbol{X}=(X^{N,1},\dots,X^{N,N})$ *satisfying* (4.1)*) has an LDP with rate function*

$$\mathcal{F}(\mu)=\begin{cases}R(\mu\|\mathbb{W}^\mu) & \text{if } R(\mu\|\mathbb{W})<\infty\\ +\infty & \text{otherwise,}\end{cases} \tag{4.11}$$

*where* $\mathbb{W}^\mu$ *is the law of the process* $X_t^\mu$ *satisfying the SDE*

$$dX_t^\mu=b_t(X_t^\mu,\mu)\,dt+dW_t. \tag{4.12}$$

**Remark 4.6.** Note that the zeros of the rate function $\mathcal{F}$ are exactly the solution to the McKean–Vlasov SDE

$$\begin{cases}dX_t=b_t(X_t,\mathrm{Law}(X_t))\,dt+dW_t,\\ X_0 \text{ with law } \rho_0,\end{cases}$$

with $R(\mathrm{Law}(X_t)\|\mathbb{W})<\infty$. In particular, since $\mathcal{F}$ has at least one zero, there exists at least one solution to the McKean–Vlasov SDE with finite relative entropy (with respect to $\mathbb{W}$).

We will see in the proof of Theorem 4.5 that, under the assumptions of the above theorem, $Q^N_{b^N}$ is well-defined by Lemma 4.1.

Moreover, since (4.10) is quite general, an application of Khasminskii's lemma provides us with the following sufficient condition:

**Lemma 4.7.** *Let* $(g_\lambda)_{\lambda>0}$ *be a sequence of Borel functions* $g_\lambda:[0,T]\times(\mathbb{R}^d)^k\to[0,+\infty)$, $k\in\mathbb{N}$, *that satisfies*

$$\limsup_{\lambda\to 0}\sup_{x_1,\dots,x_k\in\mathbb{R}^d}\mathbb{E}^{x_1,\dots,x_k}\left[\int_0^T g_\lambda(t,W_t^{1,x_1},\dots,W_t^{k,x_k})\,dt\right]=0, \tag{4.13}$$

*where the expectation is over* $k$ *independent Brownian motions* $W^{1,x_1},\dots,W^{k,x_k}$ *starting at points* $x_1,\dots,x_k\in\mathbb{R}^d$ *respectively. Then* (4.10) *is satisfied.*

*Proof.* By Khasminskii's lemma (cf. Lemma D.5), (4.13) implies that, for every $\beta\ge 0$ and every $0<\alpha<1$,

$$\sup_{(x_1,\dots x_k)\in\mathbb{R}^{kd}}\mathbb{E}\left[e^{\beta\int_0^T g_\lambda(t,W_t^{1,x_1},\dots W_t^{k,x_k})\,dt}\right]\le\frac{1}{1-\alpha},\qquad \text{for all } \lambda>0 \text{ sufficiently small,}$$

and so, averaging $(x_1,\dots x_k)$ over $\rho_0^{\otimes k}$ and using Jensen inequality yields

$$\mathbb{E}\left[e^{\beta\int_0^T g_\lambda(t,W_t^1,\dots W_t^k)\,dt}\right]\le\frac{1}{1-\alpha}\qquad \text{for all } \lambda>0 \text{ sufficiently small.}$$

Therefore we have, for every $\beta\ge 0$,

$$\limsup_{\lambda\to 0}\mathbb{E}\left[e^{\beta\int_0^T g_\lambda(t,W_t^1,\dots,W_t^k)\,dt}\right]=1,$$

which gives (4.10) with $K=1$. □

*Proof of Theorem 4.5.* We would like to apply Theorem 3.10 to $\mathcal{E}^N_{b^N}$, $\mathcal{E}_b$, $\mathcal{E}^N_{b^N_\lambda}$, $\mathcal{E}_{b_\lambda}$, with

$$G^N_\lambda(x^1,\dots,x^k) = G_\lambda(x^1,\dots,x^k) = \int_0^T g_\lambda(t,x^1_t,\dots x^k_t)\,dt, \qquad x^i\in S,\ i=1,\dots,N.$$

We claim that the conditions (2.4), (3.13a), (3.13b) and (3.14a), (3.14b) hold. Then Theorem 3.10 and Lemma 4.3 give an LDP for $Q^N_{b^N}$ with rate function $R(\mu\|\mathbb{W})+\mathcal{E}_b(\mu)$.

We now prove the claims on the above conditions and the form (4.11) for the rate function. In particular, we prove that (4.8) and (4.9) imply (3.13a) and (3.13b) respectively. Finally, note that (4.10) directly implies (3.14a) and (3.14b), and (2.4) follows as in the proof of Lemma 4.3.

*Preliminary uniform bounds:* We call, for $\mu\in\mathcal{P}(S)$,

$$\overline{K}_{\lambda,\beta} := \limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{\beta N\langle\int_0^T|b^N_t-b^N_{\lambda,t}(\cdot,z^N_{\boldsymbol{W},t})|^2dt,z^N\rangle}\right],$$
$$\overline{H}_{\lambda,\beta}(\mu) := \log\mathbb{E}\left[e^{\beta\int_0^T|b-b_\lambda(W,\mu)|^2dt}\right].$$

We will show that there exists $\lambda_0>0$ and $\beta_0\gg 1$ arbitrarily large (more precisely, for every $\beta_0>0$ large, there exists $\lambda_0>0$), such that for all $\beta\le\beta_0$:

(4.14)
$$\overline{K}_{\lambda_0,\beta}\le\overline{K}_{\lambda_0,\beta_0}<\infty,$$
$$\sup_{\mu,R(\mu\|\mathbb{W})<\infty}\overline{H}_{\lambda_0,\beta}(\mu)-(k-1)R(\mu\|\mathbb{W})\le\sup_{\mu,R(\mu\|\mathbb{W})<\infty}\overline{H}_{\lambda_0,\beta_0}(\mu)-(k-1)R(\mu\|\mathbb{W})<\infty.$$

We start with the proof for $\overline{K}_{\lambda_0,\beta_0}$. Applying assumption (4.8), we get

$$\begin{aligned}\overline{K}_{\lambda_0,\beta_0} &:= \limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{\beta_0 N\langle\int_0^T|b^N_t-b^N_{\lambda_0,t}(\cdot,z^N_{\boldsymbol{W},t})|^2dt,z^N_{\boldsymbol{W}}\rangle}\right]\\ &\le\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{\beta_0 N\int_{(S^k)'}G_{\lambda_0}(x^1,\dots,x^k)\,d(z^N_{\boldsymbol{W}})^{\otimes k}(x^1,\dots x^k)}\right].\end{aligned}$$

Since (3.14a) holds, we can apply Remark 3.13: for any $\beta_0$, the above right-hand side is finite for $\lambda_0>0$ sufficiently small. For $\overline{H}_{\lambda_0,\beta_0}(\mu)$, we have

$$\begin{aligned}\overline{H}_{\lambda_0,\beta_0}(\mu)-(k-1)R(\mu\|\mathbb{W}) &:= \log\mathbb{E}\left[e^{\beta_0\int_0^T|b-b_{\lambda_0}(W,\mu)|^2dt}\right]-(k-1)R(\mu\|\mathbb{W})\\ &\le\log\mathbb{E}\left[e^{\beta_0\int_{S^{k-1}}G_{\lambda_0}(W,y)d\mu^{\otimes(k-1)}(y)}\right]-(k-1)R(\mu\|\mathbb{W}).\end{aligned}$$

Since (3.14b) holds, we can apply Remark 3.13: for any $\beta_0$, the above right-hand side is finite and bounded uniformly over $\mu$ for $\lambda_0>0$ sufficiently small.

As a consequence of (4.14), we have (for $N$ large at least)

$$\mathbb{E}\left[e^{\frac12\int_S V^1_{b^N}(x,z^N_{\boldsymbol{W}})\,dz^N_{\boldsymbol{W}}(x)}\right]<\infty, \tag{4.15a}$$

$$\mathbb{E}\left[e^{\frac12\int_S V^1_b(W,\mu)}\right]<\infty \qquad \forall\mu \text{ with } R(\mu\|\mathbb{W})<\infty. \tag{4.15b}$$

In particular, $V^1_{b^N}(\cdot,z^N)$ and $V^{2,N}_{b^N}(\cdot,z^N)$ are in $L^1(z^N)$ for $\mathbb{P}^{\mathbb{N}}$-a.e. $z^N$ and also, for every $\mu$ with $R(\mu\|\mathbb{W})<\infty$, $V^1_b(\cdot,\mu)$ and $V^2_b(\cdot,\mu)$ are in $L^1(\mu)$, see Subsection D.2 for details. Moreover, by Lemma 4.1, at least for $N$ large, the system 4.1 admits a weak solution, with unique (under the additional constraint) law has density given by (4.4). Finally, the inequality

(4.15b) holds replacing 1/2 with any $\beta > 0$ in the exponential, in particular $\mathcal{E}_b$ is Borel by Lemma D.3.

*Verification of* (3.13a) *and* (3.13b)*:* We fix $\lambda_0$ and $\beta_0$ such that (4.14). In view of (3.13a), we show some easy uniform bounds. Using the inequality $|b^N|^2 \le 2|b^N - b^N_{\lambda_0}|^2 + 2|b^N_{\lambda_0}|^2$ and applying Hölder inequality, we get, for every $\ell \ge 0$,

(4.16)
$$\begin{aligned}\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{\ell N \langle \int_0^T \langle |b^N_t(\cdot, z^N_{\boldsymbol{W},t})|^2 dt, z^N_{\boldsymbol{W}}\rangle}\right] &\le \limsup_{N\to\infty} \frac{1}{2N} \log \mathbb{E}\left[e^{4\ell N \langle \int_0^T |b^N_t - b^N_{\lambda_0,t}(\cdot, z^N_{\boldsymbol{W},t})|^2 dt, z^N_{\boldsymbol{W}}\rangle}\right] \\ &\quad + \limsup_{N\to\infty} \frac{1}{2N} \log \mathbb{E}\left[e^{4\ell N \langle \int_0^T |b^N_{\lambda_0,t}(\cdot, z^N_{\boldsymbol{W},t})|^2 dt, z^N_{\boldsymbol{W}}\rangle}\right] \\ &=: \frac{1}{2}\overline{K}_{\lambda_0,4\ell} + \frac{1}{2}\overline{K}'_{\lambda_0,4\ell} < \infty,\end{aligned}$$

where $\overline{K}'_{\lambda_0,4\ell}$ is finite because $\sup_{N\in\mathbb{N}} \|b^N_{\lambda_0}\|_\infty < \infty$. Using now the inequality $|b^N_\lambda|^2 \le 2|b^N - b^N_\lambda|^2 + 2|b^N|^2$ and proceeding similarly, we also get, for every $\ell \ge 0$,

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{\ell N \langle \int_0^T |b^N_{\lambda,t}(\cdot, z^N_{\boldsymbol{W},t})|^2 dt, z^N_{\boldsymbol{W}}\rangle}\right] \le \frac{1}{2}\big(\overline{K}_{\lambda,4\ell} + \overline{K}_{\lambda_0,4\ell} + \overline{K}'_{\lambda_0,4\ell}\big). \tag{4.17}$$

To show (3.13a), we write, for $\beta \in \mathbb{R}$,

(4.18)
$$\mathbb{E}\left[e^{-\beta N(\mathcal{E}^N_{b^N} - \mathcal{E}^N_{b^N_\lambda})(z^N_{\boldsymbol{W}})}\right] \le \mathbb{E}\left[e^{-\beta N \langle (V^1_{b^N} - V^1_{b^N_\lambda})(\cdot, z^N_{\boldsymbol{W}}), z^N_{\boldsymbol{W}}\rangle}\right]^{1/2} \mathbb{E}\left[e^{2\beta N \langle (V^2_{b^N} - V^2_{b^N_\lambda})(\cdot, z^N_{\boldsymbol{W}}), z^N_{\boldsymbol{W}}\rangle}\right]^{1/2}$$

and we control the differences $V^1_{b^N} - V^1_{b^N_\lambda}$ and $V^2_{b^N} - V^2_{b^N_\lambda}$ separately. Using the inequality

$$\beta(|b^N|^2 - |b^N_\lambda|^2) \le |b^N|^2 + |b^N_\lambda|^2 + \frac{\beta^2}{2}|b^N - b^N_\lambda|^2,$$

and applying Hölder's inequality and the bounds (4.16) and (4.17) (with $\ell = 4$), we get

(4.19)
$$\begin{aligned}&\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}[e^{\beta N \langle (V^1_{b^N} - V^1_{b^N_\lambda})(\cdot, z^N_{\boldsymbol{W}}), z^N_{\boldsymbol{W}}\rangle}] \\ &\quad \le \frac{1}{4}\overline{K}_{\lambda,16} + \frac{1}{2}\big(\overline{K}_{\lambda_0,16} + \overline{K}'_{\lambda_0,16}\big) + \limsup_{N\to\infty} \frac{1}{2N} \log \mathbb{E}[e^{\beta^2 N \langle \int_0^T |(b^N_t - b^N_{\lambda,t})(\cdot, z^N_{\boldsymbol{W},t})|^2 dt, z^N_{\boldsymbol{W}}\rangle}] \\ &\quad = \frac{1}{4}\overline{K}_{\lambda,16} + \frac{1}{2}\big(\overline{K}_{\lambda_0,16} + \overline{K}'_{\lambda_0,16}\big) + \frac{1}{2}\overline{K}_{\lambda,\beta^2}.\end{aligned}$$

For $V^2_{b^N} - V^2_{b^N_\lambda}$, we use Lemma D.2 to obtain

$$\mathbb{E}\left[e^{2\beta N \langle V^2_{b^N}(\cdot, z^N_{\boldsymbol{W}}) - V^2_{b^N_\lambda}(\cdot, z^N_{\boldsymbol{W}}), z^N_{\boldsymbol{W}}\rangle}\right] \le \mathbb{E}\left[e^{4\beta^2 \sum_{i=1}^N \int_0^T |b^N_t - b^N_{\lambda,t}|^2 (W^i_t, z^N_{\boldsymbol{W},t})\, dt}\right]^{\frac{1}{2}},$$

and therefore

$$\limsup_{N\to\infty} \frac{1}{N} \log E^N\left[e^{4\beta N \langle V^2_{b^N}(\cdot, z^N_{\boldsymbol{W}}) - V^2_{b^N_\lambda}(\cdot, z^N_{\boldsymbol{W}}), z^N_{\boldsymbol{W}}\rangle}\right] \le \frac{1}{2}\overline{K}_{\lambda,4\beta}. \tag{4.20}$$

Putting together the inequalities (4.18), (4.19) and (4.20), we get, for some constant $c > 0$ (independent of $\beta$) and some $c_\beta > 0$,

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{-\beta N(\mathcal{E}^N_{b^N} - \mathcal{E}^N_{b^N_\lambda})(z^N_{\boldsymbol{W}})}\right] \le \bar{c} + c\overline{K}_{\lambda,c_\beta}, \tag{4.21}$$

with $\bar{c} = (\overline{K}_{\lambda_0,16} + \overline{K}'_{\lambda_0,16})/2 \ge 0$. The assumption (4.8) gives, for some new $c_\beta > 0$,

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{-\beta N(\mathcal{E}^N_{b^N} - \mathcal{E}^N_{b^N_\lambda})(z^N_{\boldsymbol{W}})}\right] \\
&\qquad \le \bar{c} + c \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[\exp\left(c_\beta N \int_0^T \int_{((\mathbb{R}^d)^k)'} g_\lambda(t, x_1, \dots x_k)\, d(z^N_{\boldsymbol{W},t})^{\otimes k}\, dt\right)\right] \\
&\qquad \le \bar{c} + c \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[\exp\left(c_\beta N \int_{(S^k)'} \int_0^T g_\lambda(t, x^1_t, \dots x^k_t)\, dt\, d(z^N_{\boldsymbol{W}})^{\otimes k}\right)\right]
\end{aligned}$$

and (3.13a) follows.

As for (3.13b), we use Lemma B.1 to obtain, for every $\mu$ with $R(\mu\|\mathbb{W}) < \infty$,

$$\beta|\mathcal{E}_b(\mu) - \mathcal{E}_{b_\lambda}(\mu)| \le R(\mu\|\mathbb{W}) + \log \mathbb{E}\left[e^{\beta|V_b(W,\mu) - V_{b_\lambda}(W,\mu)|}\right].$$

Using the same arguments as before, we get, for some $c' > 0$ (independent of $\beta$) and some $c_\beta > 0$,

$$\beta|\mathcal{E}_b(\mu) - \mathcal{E}_{b_\lambda}(\mu)| \le \frac{k+1}{2} R(\mu\|\mu_0) + \bar{c} + c'\overline{H}_{\lambda,c_\beta}, \tag{4.22}$$

where $\bar{c} > 0$ is such that $\bar{c} \ge (\overline{H}_{\lambda_0,16}(\mu) - (k-1)R(\mu\|\mu_0) + \overline{H}'_{\lambda_0,16})/2$ for every $\mu$ with $R(\mu\|\mathbb{W}) < \infty$, and $\overline{H}'_{\lambda_0,\beta} = \log \mathbb{E}[e^{\beta \int_0^T |b_{\lambda_0}(W,\mu)|^2 dt}]$. The assumption (4.9) gives, for some new $c_\beta$,

$$\beta|\mathcal{E}_{b^N}(\mu) - \mathcal{E}_{b^N_\lambda}(\mu)| \le \frac{k+1}{2} R(\mu\|\mu_0) + \bar{c} + c' \log \mathbb{E}\left[e^{c_\beta \int_{S^k} \int_0^T g_\lambda(t, W_t, x^2_t \dots x^k_t)\, dt\, d\mu^{\otimes(k-1)}}\right]$$

and (3.13b) follows.

*Proof of* (4.11)*:* By (4.15b), $b$ satisfies the assumption of Lemma D.4, which then implies the representation formula (4.11) for the rate function. The proof is complete. □

The above proof shows that we can relax some of the assumptions, as we show below.

**Proposition 4.8.** *The results of Theorem 4.5 (namely the LDP for $Q^N_{b^N}$ with rate function $\mathcal{F}$) remain valid if any of the following statements hold:*

*(a) Instead of $\mathcal{F}$*Lip*-inducing drifts $(b^N_\lambda, b_\lambda)_{\lambda>0}$, we assume that for every $\lambda > 0$ the family $Q^N_{b^N_\lambda}$ has an LDP with rate function $\mathcal{F}_{b_\lambda}$ (defined similarly to $\mathcal{F}$ via* (4.11), (4.12)*), and for every $\beta \in \mathbb{R}$,*

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[\exp\left(N\beta \int_0^T \langle |b^N_{\lambda,t}|^2(\cdot, z^N_{\boldsymbol{W},t}), z^N_{\boldsymbol{W},t}\rangle\right) dt\right] < \infty, \\
&\sup_{\mu, R(\mu\|\mathbb{W})<\infty} \log \mathbb{E}\left[\exp\left(\beta \int_0^T |b_{\lambda,t}|^2(W_t, \mu_t)\, dt\right)\right] - R(\mu\|\mathbb{W}) < \infty.
\end{aligned} \tag{4.23}$$

*(b) Instead of the existence of $g_\lambda$ and* (4.8), (4.9), *we assume there exists a constant $K \in \mathbb{R}$ such that for every $\beta \in \mathbb{R}$,*

$$\begin{aligned} &\limsup_{\lambda\to 0}\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[\exp\left(N\beta\int_0^T\langle|b_t^N-b_{\lambda,t}^N|^2(\cdot,z_{\boldsymbol{W},t}^N),z_{\boldsymbol{W},t}^N\rangle\right)dt\right]\le K,\\ &\limsup_{\lambda\to 0}\sup_{\mu,R(\mu\|\mathbb{W})<\infty}\log\mathbb{E}\left[\exp\left(\beta\int_0^T|b_t-b_{\lambda,t}|^2(W_t,\mu_t)\,dt\right)\right]-R(\mu\|\mathbb{W})\le K. \end{aligned} \tag{4.24}$$

*Proof.* The inequalities (4.23) imply via Lemmas 4.1, D.2 and D.4 both the representations of $Q^N_{b^N_\lambda}$ and $\mathcal{F}_{b_\lambda}$ in terms of $\mathcal{E}^N_{b^N_\lambda},\mathcal{E}_{b_\lambda}$, and the estimates (2.4). Moreover, (4.24) combined with (4.21) and (4.22) imply conditions (2.5a) and (2.5b). Hence we can use directly Theorem 2.8 to deduce the LDP (the representation formula (4.11) follows again from Lemma D.4). □

4.3. **Applications to concrete examples.** In this subsection, we consider common form of drifts $b^N$ that appear in applications.

4.3.1. *Example: 2-point interaction.* We start with the example of 2-point, translation-invariant interaction; while this is a particular case of the $k$-point interaction, we discuss this case separately, to highlight the essential ingredients of the result. In this example, we consider a simple class of drifts, that commonly appear in various fields of application, namely

$$b_t^N(x_i,z_{\boldsymbol{x}}^N)=\frac{1}{N}\sum_{j\neq i}\varphi_t(x_i-x_j),\qquad z_{\boldsymbol{x}}^N=\frac{1}{N}\sum_{j=1}^N\delta_{x_j} \tag{4.25}$$

for some Borel map $\varphi:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$.

**Remark 4.9.** Strictly speaking, (4.25) is not a good definition, because the right-hand side depends on $i$ and not just on $x_i$ and $z_{\boldsymbol{x}}^N$. However we can give a rigorous definition with a harmless change of (4.25). Precisely, we can define

$$b_t^N(x,\mu)=\int_{\mathbb{R}^d\setminus\{x\}}\varphi_t(x-y)\,d\mu_t(y). \tag{4.26}$$

Indeed, note that (recall $\boldsymbol{W}=(W^1,\dots,W^N)$ is a $N$-tuple of $d$-dimensional independent Brownian motions with law $\tilde{P}^N=\mathbb{W}^{\otimes N}$)

$$\frac{1}{N}\sum_{j\neq i}\varphi_t(W_t^i-W_t^j)=\int_{\mathbb{R}^d\setminus\{x\}}\varphi_t(x-y)\,dz_{\boldsymbol{W},t}^N(y),\quad\text{for a.e. } t\in[0,T] \text{ and } \tilde{P}^N\text{-a.e. } \boldsymbol{W}. \tag{4.27}$$

because the set $\{t\in[0,T]:W_t^i=W_t^j,j\neq i\}$ has Lebesgue measure zero for $\tilde{P}^N$-a.e. $\boldsymbol{W}=(W^1,\dots,W^N)$. Therefore, if Girsanov's theorem can be applied to the particle system (4.1), as it is the case in the next result, then (4.27) holds also replacing $\boldsymbol{W}$ and $\tilde{P}^N$ respectively with $\boldsymbol{X}^N=(X^{N,1},\dots X^{N,N})$, the solution to the particle system (4.1) with drift (4.26), and its law $\tilde{Q}^N$. Hence such solution $\boldsymbol{X}^N$ solves also the original particle system with drift (4.25).

We introduce also the drift of the associated McKean–Vlasov SDE, namely

$$b_t(x,\mu)=\int_{\mathbb{R}^d}\varphi_t(x-y)\,d\mu_t(y).$$

**Proposition 4.10.** *Assume the condition* (4.6) *on the initial law* $\rho_0$*. Suppose that*

$$\mathbb{E}\left[e^{\beta\int_0^T|\varphi_t|^2(W_t^1-W_t^2)\,dt}\right]<\infty,\qquad \forall\beta\in\mathbb{R}, \tag{4.28}$$

*where* $W^1, W^2$ *are independent Brownian motions with common initial law* $\rho_0$*.*

*Then the family* $\{Q_b^N\}$ *of laws of the empirical processes associated to the interacting system* (4.1) *with drifts of the form* (4.25) *satisfies an LDP with rate function* $\mathcal{F}$ *given in* (4.11)*.*

*In particular,* (4.28) *holds whenever* $\varphi$ *is in* $L_t^q(L_{x-y}^p)$*, for* $p, q$ *satisfying*

$$2\le p,q\le\infty,\quad \frac{d}{p}+\frac{2}{q}<1. \tag{4.29}$$

Proposition 4.10 is a special case of Proposition 4.13 with $k=2$, $p_1=p, p_2=+\infty$, and hence we will postpone the proof.

**Remark 4.11.** The condition $\varphi\in L_t^q(L_{x-y}^p)$ with $p,q$ satisfying (4.29) is well-known in the literature on the so-called regularization by noise phenomenon (where an ill-posed ODE or PDE gains well-posedness by addition of a suitable noise). Indeed, a $d$-dimensional SDE, with additive noise and drift in $L_t^q(L_x^p)$, with $p,q$ satisfying (4.29), has the strong existence and uniqueness property, see for example [KR05, FF11] among many others; on the contrary, if $p,q$ do not satisfy (4.29) not even with equality, there exist counterexamples to well-posedness for SDEs, even in the weak sense, see e.g. [BFGM19, Section 7].

For this reason the exponents of (4.29) are likely optimal for irregular drifts $\varphi$ in our example: we expect that, for $L_t^q(L_{x-y}^p)$ drifts without condition (4.29), even the 2-particle system is not well-posed. However, there are likely drifts that satisfies (4.28) but are not in a $L_t^q(L_x^p)$ class with (4.29).

**Remark 4.12.** A relevant example of function $\varphi$ verifying condition (4.29) is

$$\varphi(t,z)=|z|^\alpha g\left(\frac{z}{|z|}\right)1_{|z|\le R}+h(z)1_{|z|>R},$$

with $g:\mathbb{S}^{d-1}\to\mathbb{S}^{d-1}$ and $h:\mathbb{R}^d\to\mathbb{R}^d$ both Borel bounded, $R>0$, and exponent $\alpha$ satisfying

$$\alpha>-1\ \text{ for }\ d\ge 2,\qquad\text{and}\qquad \alpha>-1/2\ \text{ for }\ d=1.$$

It is unclear whether the restriction $\alpha>-1/2$ in $d=1$, which is due to the assumption $p\ge 2$, is really optimal or is a drawback of the method, nonetheless the assumption $p,q\ge 2$ appears in several works dealing with singular drifts, like [KR05, FF11].

When $d\ge 2$, the above condition includes the case

$$\varphi(x)=-\nabla\Phi(x),\qquad \Phi(t,x)=|x|^\alpha,\qquad \alpha>0,$$

in some neighborhood around $x=0$, and with $\varphi$ bounded outside of this neighborhood. However, $\Phi(x)=\log|x|$ does not fall in this class. Moreover, as shown in [HH16], the exponential moment estimate of (4.28) actually blows up when $\beta$ is large enough.

The latter is also related to the fact that for $d\ge 2$ and a logarithmic potential $\Phi(x)=\log|x|$, the law of SDE (4.1) might no longer be absolutely continuous with respect to the law of non-interacting Brownian motions—it is possible in the case of a sufficiently strong attractive force that the particles hit each other (see [FJ17]). Interestingly enough, as is done in [FJ17], it can still be shown that the particle system converges to the corresponding McKean–Vlasov equation. However, whether in this case a large deviation principle still exists is not known.

4.3.2. *Example: $k$-point interaction.* We can also treat the case of a $k$-interaction drift, namely

$$b_t^N(x_i, z_{\boldsymbol{x}}^N) = \frac{1}{N^{k-1}} \sum_{j_1,\dots,j_{k-1}\neq i \text{ all distinct}} \varphi_t(x_i, x_{j_1}, \dots x_{j_{k-1}}) \tag{4.30}$$

for some Borel function $\varphi : [0,T] \times \mathbb{R}^{kd} \to \mathbb{R}^d$. As in the previous example, the rigorous definition of $b^N$ can be given as in Remark 4.9. Similarly we take the drift of the associated McKean–Vlasov SDE

$$b_t(x,\mu) = \int_{\mathbb{R}^{(k-1)d}} \varphi_t(x, y_1, \dots y_{k-1})\, d\mu_t^{\otimes(k-1)}(y_1, \dots y_{k-1}). \tag{4.31}$$

**Proposition 4.13.** *Assume the condition* (4.6) *on the initial law $\rho_0$. Suppose that*

$$\mathbb{E}\left[e^{\beta \int_0^T |\varphi_t|^2(W_t^1,\dots,W_t^k)\,dt}\right] < \infty, \qquad \forall \beta \in \mathbb{R} \tag{4.32}$$

*where $W^1, \dots, W^k$ are independent Brownian motions with common initial law $\rho_0$.*

*Then the family $\{Q_b^N\}$ of laws of the empirical processes associated to the interacting system* (4.1) *with drifts of the form* (4.30) *satisfies an LDP with rate function $\mathcal{F}$ given in* (4.11).

*In particular,* (4.32) *holds whenever*

$$\varphi \in L_t^q(L_{x_1}^{p_1}(\dots(L_{x_k}^{p_k})\dots)), \tag{4.33}$$

*with $p_1, \dots p_k, q$ satisfying*

$$p_1, \dots p_k, q \in [2,\infty], \quad \frac{d}{p_1} + \dots + \frac{d}{p_k} + \frac{2}{q} < 1. \tag{4.34}$$

**Remark 4.14.** Similarly to Lemma D.7, (4.33) can be replaced by

$$\varphi \in L_t^q(L_{x_{\sigma(1)}}^{p_1}(\dots(L_{x_{\sigma(k)}}^{p_k})\dots))$$

for some permutation $\sigma$ of $\{1, \dots k\}$.

**Remark 4.15.** The space of $\varphi$ satisfying condition (4.32) is a vector space (as easily checked). Hence, condition (4.32) also holds in the more general case of $\varphi = \sum_{j=1}^m \varphi_j$ for some $m$, where, for any $j = 1, \dots m$,

$$\varphi_j \in L_t^{q^{(j)}}(L_{x_1}^{p_1^{(j)}}(\dots(L_{x_k}^{p_k^{(j)}})\dots))$$

with $p_1^{(j)}, \dots p_k^{(j)}, q^{(j)}$ satisfying (4.34). In particular, we can allow $\varphi$ to be a sum of a bounded function and a function satisfying (4.33) and (4.34) as in Theorem 1.1.

*Proof of Proposition 4.13.* We will use three different approximations:

(1) First, we consider $b$ with $\varphi$ Lipschitz bounded and show that $(b^N, b)$ are $\mathcal{F}$Lip-inducing, by bounding away the self-interactions.
(2) Then we consider the case $b$ with $\varphi$ Borel bounded. We apply Theorem 4.5 to get the desired LDP.
(3) Finally, we extend this to $\varphi$ satisfying (4.32) using truncation of $\varphi$ and the approximation given in step (2).

The fact that (4.32) holds under conditions (4.33) and (4.34) follows from Lemma D.8.

**(1)** Assume that $\varphi : [0,T] \times \mathbb{R}^{kd} \to \mathbb{R}^d$ is Borel bounded and that the map $x \mapsto \varphi(t,x)$ is globally Lipschitz continuous for every $t \in [0,T]$, with Lipschitz constant $\mathrm{Lip}(\varphi)$ independent of $t$. We will show that $(b^N, b)$ is $\mathcal{F}$Lip-inducing.

It is well known that $b \in \mathcal{F}\mathrm{Lip}$: indeed, for any $x_1, x_2 \in \mathbb{R}^d$ and $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^d)$, we have

$$\begin{aligned}
&|b_t^N(x_1,\mu_1) - b_t^N(x_2,\mu_2)| \\
&\quad \le \int_{\mathbb{R}^{(k-1)d}} \left|\varphi_t(x_1, y_1, \dots y_{k-1}) - \varphi_t(x_2, y_1', \dots y_{k-1}')\right| \, d\gamma(y_1, y_1') \cdot \ldots \cdot d\gamma(y_{k-1}, y_{k-1}') \\
&\quad \le \mathrm{Lip}(\varphi) \left( |x_1 - x_2| + \int_{\mathbb{R}^{(k-1)d}} \left( \sum_{i=1}^{k-1} |y_i - y_i'| \right) d\gamma(y_1, y_1') \cdot \ldots \cdot d\gamma(y_{k-1}, y_{k-1}') \right) \\
&\quad = \mathrm{Lip}(\varphi) \left( |x_1 - x_2| + (k-1) \int_{\mathbb{R}^{(k-1)d}} |y - y'| \, d\gamma(y, y') \right),
\end{aligned}$$

for any coupling $\gamma$ between $\mu$ and $\nu$; optimizing over all couplings then yields the required Lipschitz estimate. Moreover, clearly $|b^N| \le |b|$ and hence $\sup_N \|b^N\|_\infty < \infty$. Finally, to show (4.5), note that, similarly to Lemma 3.3, we have for every $N$ and $P^N$-almost every $z^N$

$$\begin{aligned}
&\int_0^T \left\langle |b_t - b_t^N|^2(\cdot, z_t^N), z_t^N \right\rangle dt \\
&\le \frac{2}{N} \int_0^T \sum_{i=1}^N \left( \frac{1}{N^{k-1}} \left| \left( \sum_{j_1,\dots,j_{k-1} \neq i \text{ all distinct}} - \sum_{j_1,\dots,j_{k-1} \neq i} \right) \varphi_t(W_t^i, W_t^{j_1}, \dots W_t^{j_{k-1}}) \right| \right)^2 dt \\
&\quad + \frac{2}{N} \int_0^T \sum_{i=1}^N \left( \frac{1}{N^{k-1}} \left| \left( \sum_{j_1,\dots,j_{k-1} \neq i} - \sum_{j_1,\dots,j_{k-1}} \right) \varphi_t(W_t^i, W_t^{j_1}, \dots W_t^{j_{k-1}}) \right| \right)^2 dt \\
&\le 2T \|\varphi\|_\infty^2 \frac{1}{N^{2(k-1)}} \left( \left| \frac{(k-1)(k-2)}{2} N^{k-2} \right|^2 + \left| (k-1) N^{k-2} \right|^2 \right) \le \frac{C}{N^2},
\end{aligned}$$

for some generic constant $C$. Hence $(b^N, b)$ is $\mathcal{F}\mathrm{Lip}$-inducing.

**(2)** Next, suppose that $\varphi : [0,T] \times \mathbb{R}^{kd} \to \mathbb{R}^d$ is Borel bounded. We can find a sequence of Lipschitz approximations $(\varphi_\lambda)_{\lambda > 0}$, with $\varphi_\lambda$ as in **(1)** for each $\lambda > 0$, such that $\varphi_\lambda \to \varphi$ Lebesgue-a.e. as $\lambda \to 0$ and $\|\varphi_\lambda\|_\infty \le \|\varphi\|_\infty$ (for example, we can take $\varphi_\lambda$ as convolutions of $\varphi$ with approximations of identity). We take $b_\lambda^N$, $b_\lambda$ the corresponding drifts for the particle system and the McKean-Vlasov SDE (as respectively in (4.30), (4.31) with $\varphi_\lambda$ in place of $\varphi$). In order to apply Theorem 4.5, with $(b_\lambda^N, b_\lambda)$ as sequence of $\mathcal{F}\mathrm{Lip}$-inducing drifts, we verify now the assumptions (4.8), (4.9) and (4.10). The estimates (4.8) and (4.9) are easily obtained with

$$g_\lambda(t, x_1, \dots, x_k) = |\varphi(t, x_1, \dots, x_k) - \varphi_\lambda(t, x_1, \dots, x_k)|^2 .$$

Concerning (4.10), we start noting that

$$g_\lambda(t, W^1, \dots W_t^k) \to 0 \quad \text{and} \quad g_\lambda(t, W^1, \dots W_t^k) \le 4\|\varphi\|_\infty^2, \quad \text{for } \mathbb{P} \otimes dt\text{-a.e. } (\omega, t).$$

Hence, for each $\beta > 0$, we apply dominated convergence theorem twice, i.e. to the time integral and then to the expectation, thereby obtaining

$$\mathbb{E}\left[ e^{\beta \int_0^T g_\lambda(t, W_t^1, \dots, W_t^k)\, dt} \right] \to 1 \quad \text{as} \quad \lambda \to 0, \tag{4.35}$$

that is (4.10). Hence we can apply Theorem 4.5 and obtain the desired LDP.

**(3)** Finally, assume that $\varphi$ satisfies (4.32). We take

$$\tilde{\varphi}_\lambda = (\varphi \wedge 1/\lambda) \vee (-1/\lambda).$$

We take also an increasing sequence $(\beta_\lambda)_\lambda$ with $\beta_\lambda \to \infty$. For each $\lambda > 0$ fixed, applying (4.35) to $\tilde{\varphi}_\lambda$ in place of $\varphi$, we get the existence of a Lipschitz function $\varphi_\lambda$ as in **(1)**, such that

$$\mathbb{E}\left[e^{\beta_\lambda \int_0^T |\tilde{\varphi}_\lambda - \varphi_\lambda|^2(t, W_t^1, \dots, W_t^k)\, dt}\right] < 1 + \lambda. \tag{4.36}$$

Now we take $b_\lambda^N$, $b_\lambda$ the drifts for the particle system and the McKean-Vlasov SDE, as respectively in (4.30), (4.31) with $\varphi_\lambda$ in place of $\varphi$. As before, in order to apply Theorem 4.5, with $(b_\lambda^N, b_\lambda)$ as sequence of $\mathcal{F}$Lip-inducing drifts, we verify the assumptions (4.8), (4.9) and (4.10). As before, the estimates (4.8) and (4.9) are easily obtained with

$$g_\lambda(t, x_1, \dots, x_k) = |\varphi(t, x_1, \dots, x_k) - \varphi_\lambda(t, x_1, \dots, x_k)|^2 .$$

Concerning (4.10), we split $g_\lambda \le 2|\varphi - \tilde{\varphi}_\lambda|^2 + 2|\tilde{\varphi}_\lambda - \varphi_\lambda|^2$ and study the two terms separately. For the first term, for every $\beta > 0$, we have by dominated convergence theorem

$$\limsup_{\lambda \to 0} \mathbb{E}\left[e^{\beta \int_0^T |\varphi - \tilde{\varphi}_\lambda|^2(t, W_t^1, \dots, W_t^k)\, dt}\right] = 1.$$

For the second term, the bound (4.36) implies, for every $\beta > 0$,

$$\limsup_{\lambda \to 0} \mathbb{E}\left[e^{\beta \int_0^T |\tilde{\varphi}_\lambda - \varphi_\lambda|^2(t, W_t^1, \dots, W_t^k)\, dt}\right] \le \limsup_{\lambda \to 0} \mathbb{E}\left[e^{\beta_\lambda \int_0^T |\tilde{\varphi}_\lambda - \varphi_\lambda|^2(t, W_t^1, \dots, W_t^k)\, dt}\right] = 1.$$

The two bounds above give (4.10). Hence we can apply Theorem 4.5 and obtain the desired LDP. The proof is complete. □

4.3.3. *Example: Measure dependent drift.* As an example of a more general interaction, we consider drifts of the form (cf. [GNP21])

$$b_t^N(x_i, z_{\boldsymbol{x}}^N) = \frac{1}{N} \sum_{j \ne i} \Psi\left(x_i, x_j, \frac{1}{N} \sum_{\ell \ne i} \varphi_t(x_i, x_\ell), \frac{1}{N} \sum_{\ell \ne j} \varphi_t(x_j, x_\ell), (z_{\boldsymbol{x}}^N)_t\right), \tag{4.37}$$

where $\Psi : \mathbb{R}^{4d} \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ and $\varphi : [0, T] \times \mathbb{R}^{2d} \to \mathbb{R}^d$ are Borel maps. As in the previous examples, the rigorous definition of $b^N$ can be given as in Remark 4.9. Similarly we take the drift of the associated McKean–Vlasov SDE

$$b_t(x, \mu) = \int_{\mathbb{R}^d} \Psi\left(x, y, \int_{\mathbb{R}^d} \varphi_t(x, z)\, d\mu_t(z), \int_{\mathbb{R}^d} \varphi_t(y, z)\, d\mu_t(z), \mu_t\right) d\mu_t(y).$$

Throughout we will use both the 1-Wasserstein metric, $W_1$ and the bounded Lipschitz metric, $d_{BL}$ on $\mathcal{P}(\mathbb{R}^d)$, with the latter given by

$$d_{BL}(\mu, \nu) := \sup_{\phi : \|\phi\|_\infty \le 1, \mathrm{Lip}(\phi) \le 1} \left\{\int_{\mathbb{R}^d} \phi(x) d\mu(x) - \int_{\mathbb{R}^d} \phi(x) d\nu(x)\right\}.$$

Note that $d_{BL} \le W_1$.

**Proposition 4.16.** *Assume that the initial law $\rho_0$ satisfies condition* (4.6)*. Furthermore, let $\Psi \in \mathrm{Lip}(\mathbb{R}^{4d} \times (\mathcal{P}(\mathbb{R}^d), d_{BL}))$, and suppose that there exists a constant $L$ such that*

$$\Psi(x, y, a, b, \mu) \le L(1 + |a| + |b|), \qquad x, y, a, b \in \mathbb{R}^d, \mu \in \mathcal{P}(\mathbb{R}^d) \tag{4.38}$$

*and*

$$\mathbb{E}\left[e^{\beta \int_0^T |\varphi_t|^2(W_t^1, W_t^2)\, dt}\right] < \infty, \qquad \text{for all } \beta \in \mathbb{R}, \tag{4.39}$$

*where $W^1, W^2$ are independent Brownian motions with common initial law $\rho_0$.*

*Then the family $Q_b^N$ of laws of the empirical processes associated to the interacting system* (4.1) *induced by drifts of the form* (4.37) *satisfies an LDP with rate function $\mathcal{F}$ given in* (4.11).

*In particular,* (4.39) *holds whenever $\varphi$ is translation invariant with $\varphi \in L_t^q(L_{x-y}^p)$ for $p, q \in [2, \infty]$ satisfying*

$$\frac{d}{p} + \frac{2}{q} < 1. \tag{4.40}$$

**Remark 4.17.** A couple of comments:

(1) Similar to Remark 4.15, Proposition 4.16 holds for any $\varphi = \sum_{j=1}^m \varphi_j$ with each $\varphi_j$ satisfying (4.40) for suitable exponents $p^{(j)}, q^{(j)}$.
(2) As in Remark 3.6, when $\varphi$ is bounded, we can include self-interactions in the summations in (4.25), (4.30) and (4.37) without changing the results above.

**Remark 4.18.** The example of Theorem 1.1 is a particular case of (4.37), which can be seen by setting

$$\Psi(x, x_2, y, x_3, \mu) := \psi(x, \mu, y).$$

The LDP for $Q_b^N$ follows from Proposition 4.16. Moreover, the convergence to the McKean-Vlasov SDE will be shown in the next section, and follows directly from Proposition 4.24.

*Proof of Proposition 4.16.* The proof is an adaptation of the proof for Proposition 4.13. As in step **(1)** of the proof of Lemma 4.13, we first consider suitable bounded Lipschitz functions $\varphi$ with corresponding drifts $b$. Also in this case, it is not difficult to see that in this case $(b^N, b)$ is $\mathcal{F}$Lip-inducing. First, to show that $b \in \mathcal{F}\mathrm{Lip}$, note that $b$ is bounded by (4.38) and the boundedness of $\varphi$ yields

$$|b_t(x_1, \mu_1) - b_t(x_2, \mu_2)| \leq \mathrm{Lip}(\Psi) \iint_{\mathbb{R}^d \times \mathbb{R}^d} \Big( |x_1 - x_2| + |y_1 - y_2| + (\mathrm{I}) + (\mathrm{II}) \Big) \, d\gamma(y_1, y_2) + (\mathrm{III})$$

where $(\mathrm{III}) = \mathrm{Lip}(\Psi)\, d_{BL}(\mu, \nu)$ and

$$(\mathrm{I}) \leq \mathrm{Lip}(\varphi) \iint_{\mathbb{R} \times \mathbb{R}^d} \Big( |x_1 - x_2| + |w_1 - w_2| \Big) \, d\gamma(w_1, w_2),$$
$$(\mathrm{II}) \leq \mathrm{Lip}(\varphi) \iint_{\mathbb{R} \times \mathbb{R}^d} \Big( |y_1 - y_2| + |w_1 - w_2| \Big) \, d\gamma(w_1, w_2).$$

Noting $d_{BL} \leq W_1$, inserting estimates (I) and (II) into the previous inequality and optimizing over all couplings $\gamma$ between $\mu_1$ and $\mu_2$ yields

$$|b_t(x_1, \mu_1) - b_t(x_2, \mu_2)| \leq \mathrm{Lip}(\Psi)(2 + 3\mathrm{Lip}(\varphi)) \big( |x_1 - x_2| + W_1(\mu_1, \mu_2) \big),$$

i.e., the Lipschitz estimates holds. Next, note that we have for $P^N$-almost every $z^N$

$$\int_0^T \big\langle |b_t - b_t^N|^2(\cdot, z_t^N), z_t^N \big\rangle \, dt \leq \left( \mathrm{Lip}(\Psi) \|\varphi\|_\infty \frac{2}{N} + \frac{1}{N} L(1 + 2\|\varphi\|_\infty) \right)^2,$$

which vanishes as $N \to \infty$.

Finally, since $\Psi$ is Lipschitz, we find that for any $\varphi, b$ and approximating sequence $\varphi^\lambda, b^\lambda$ the estimates (4.8) and (4.9) may be obtained with $g_\lambda(t, x, y) = \mathrm{Lip}(\Psi)^2 |\varphi_t - \varphi_t^\lambda|^2$, and the rest of the proof follows similar to Proposition 4.13. □

***Background.*** There are several papers dealing with large deviations for McKean–Vlasov SDEs. One of the first papers is [DG87], which proves an LDP for the paths of empirical measure, assuming continuity on the drift $b$ among other hypotheses. The papers [DMZ03] and [BDF12] prove an LDP (for the empirical measures on the path space, as here), again assuming also continuity of the drift. The work [Lac18] proves propagation of chaos and a large deviation upper bound, assuming $b$ bounded and continuous in the measure argument, but with respect to a stronger topology (the $\tau$ topology). Outside the context of bounded (or linear growth) drift, we are aware only of the result in [Fon04], which shows an LDP for a system of one-dimensional particles with a repulsive two-body interaction of order $1/x$, that is the framework of Subsection 4.3.1 but with $\varphi(x) = 1/x$, which is outside the class we can deal with here.

We also recall the recent papers [Orr20] and [dRST19] on large deviations for interacting diffusions/McKean–Vlasov SDEs when also the noise intensity tends to 0, and [GNP21] for the large deviations of the Brownian one-dimensional hard-rod system.

4.4. **Uniqueness for the McKean–Vlasov SDE.** In this subsection we consider the McKean–Vlasov SDE associated with the particle system (4.1), namely

$$\begin{cases} dX_t = b_t(X_t, \mathrm{Law}(X_t))\, dt + dW_t, \\ X_0 \text{ with law } \rho_0, \end{cases} \tag{4.41}$$

for a given Borel drift $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ and a given initial law $\rho_0$.

We assume conditions on $b$ which include the examples in the previous subsection. We show, under these conditions, uniqueness in law for the McKean–Vlasov SDE (4.41). As a consequence, in all examples in the previous section, the empirical measures associated with the particle system (4.1) converge, as $N \to \infty$, to the law of the solution of the McKean–Vlasov SDE (4.41).

We keep the notation of the previous section, with $S = C([0,T];\mathbb{R}^d)$, $\mathbb{W}$ the $d$-dimensional Wiener measure with initial law $\rho_0$ and $W$ the Brownian motion with initial law $\rho_0$ (similarly $W^i$ are independent Brownian motion starting from $\rho_0$). We fix $m > 2$ and, for $\eta > 0$, we set

$$M_\eta := \left\{ \mu \in \mathcal{P}(S) \;\middle|\; \left\| \frac{d\mu}{d\mathbb{W}} \right\|_{L^m(S,\mathbb{W})} \le \eta \right\}.$$

We introduce some notation and give some remarks on the restriction from paths on $[0,T]$ to path on a subinterval $[0,T_2]$. We call $\pi_{[0,T_2]} : S = C([0,T];\mathbb{R}^d) \to S_{[0,T_2]} := C([0,T_2];\mathbb{R}^d)$ the restriction map; for $\gamma \in S$, we call $\gamma_{[0,T_2]} = \pi_{[0,T_2]}(\gamma)$ its restriction to $[0,T_2]$. For $\rho \in \mathcal{P}(S)$, we call $\rho_{[0,T_2]}$ the restriction of $\rho$ to the interval $[0,T_2]$, that is $\rho_{[0,T_2]} = (\pi_{[0,T_2]})_\# \rho$, which is a probability measure on $S_{[0,T_2]}$; similarly we use $d\rho_{[0,T_2]}/d\mathbb{W}$ for the density of $\rho_{[0,T_2]}$ with respect to $\mathbb{W}_{[0,T_2]}$. Note that $d\rho_{[0,T_2]}/d\mathbb{W}$ is a version of the conditional expectation of $d\rho/d\mathbb{W}$ given $\pi_{[0,T_2]}$; in particular, if $d\rho/d\mathbb{W} \in L^\gamma(S,\mathbb{W})$ for some $\gamma \ge 1$, then $d\rho_{[0,T_2]}/d\mathbb{W}$ is in $L^\gamma(S_{[0,T_2]}, \mathbb{W}_{[0,T_2]})$ and we have (by Jensen's inequality for conditional expectation)

$$\|d\rho/d\mathbb{W}\|_{L^\gamma} \le \|d\rho_{[0,T_2]}/d\mathbb{W}\|_{L^\gamma}.$$

We keep a similar notation for $\rho^{\otimes k}_{[0,T_2]}$; note that $d\rho^{\otimes k}_{[0,T_2]}/d\mathbb{W}^{\otimes k}$ (the density of $\rho^{\otimes k}_{[0,T_2]}$ with respect to $\mathbb{W}^{\otimes k}_{[0,T_2]}$) coincides with $(d\rho_{[0,T_2]}/d\mathbb{W})^{\otimes k}$ (the density tensorised $k$ times). Finally, we call $V_{b,[0,T_2]}$ for the map $V_b$ (as defined in Subsection 4.1) with final time $T_2$ instead of $T$.

**Theorem 4.19.** *Fix the initial law $\rho_0 \in \mathcal{P}(\mathbb{R}^d)$. Fix $m > 2$. Assume that, for every $\eta > 0$, for every $\beta > 0$,*

$$\sup_{\mu \in \mathcal{P}(S),\, R(\mu\|\mathbb{W}) \le \eta} \mathbb{E}\left[\exp\left(\beta \int_0^T |b_t(W_t, \mu_t)|^2 dt\right)\right] < \infty. \tag{4.42}$$

*Assume also that there exist a non-negative Borel function $g : [0,T] \times (\mathbb{R}^d)^k \to \mathbb{R}$ and, for every $\eta > 0$, a constant $C_\eta \ge 0$, such that, for every $0 \le T_1 < T_2 \le T$,*

$$\begin{aligned} &\int_{T_1}^{T_2} |b_t(x_t, \mu_t) - b_t(x_t, \nu_t)|^2 dt \\ &\qquad \le C_\eta \int_{S^{k-1}_{[0,T_2]}} \left(\int_{T_1}^{T_2} g_t(x_t, y_t)\, dt\right) \left|u^{\otimes(k-1)}_{[0,T_2]}(y) - v^{\otimes(k-1)}_{[0,T_2]}(y)\right|^2 d\mathbb{W}^{\otimes(k-1)}_{[0,T_2]}(y) \\ &\qquad\qquad \text{for } \mathbb{W}\text{-a.e. } x \text{ and every } \mu, \nu \in M_\eta, \end{aligned} \tag{4.43}$$

*where $u_{[0,T_2]} = d\mu_{[0,T_2]}/d\mathbb{W}$ and $v_{[0,T_2]} = d\nu_{[0,T_2]}/d\mathbb{W}$, and that, for some $\tilde{m} > m$ with $\tilde{m} \ge (m/2)' = m/(m-2)$,*

$$\mathbb{E}\left[\left(\int_0^T g_t(W_t^1, \dots W_t^k)\, dt\right)^{\tilde{m}}\right] < \infty. \tag{4.44}$$

*Then uniqueness in law holds for the McKean–Vlasov equation (4.41) among laws with finite relative entropy with respect to $\mathbb{W}$, that is: if $X$ and $Y$ are two solutions with laws $\bar{\mu}$ and $\bar{\nu}$ respectively, and if $R(\bar{\mu}\|\mathbb{W}) < \infty$ and $R(\bar{\nu}\|\mathbb{W}) < \infty$, then $\bar{\mu} = \bar{\nu}$.*

*Proof.* For any $\mu \in \mathcal{P}(S)$ with $R(\mu\|\mathbb{W}) < \infty$, assumption (4.42) gives via Girsanov's theorem D.1 the existence of a weak solution $X^\mu$ to the SDE

$$dX_t^\mu = b_t(X_t^\mu, \mu_t)\, dt + dW_t, \tag{4.45}$$

with law $F(\mu) := \mathbb{W}^\mu = \mathrm{Law}(X^\mu)$ given by

$$\frac{dF(\mu)}{d\mathbb{W}}(W) = \exp\big(V_b(W, \mu)\big),$$

with $V_b$ as defined in Section 4.1. Note that $F(\mu)$ is the unique law solving (4.45) and having finite entropy with respect to $\mathbb{W}$. Indeed, if $\nu$ is the law of another solution $Y^\mu$ to (4.45) with $R(\nu\|\mathbb{W}) < \infty$, then, by Lemma B.1,

$$\int_S \int_0^T |b_t(x_t, \mu_t)|^2 dt\, d\nu(x) \le R(\nu\|\mathbb{W}) + \log \mathbb{E}\left[e^{\int_0^T |b_t(W_t, \mu_t)|^2 dt}\right] < \infty.$$

Therefore the uniqueness condition (D.1.3) is met under $\nu$ and so $\nu = F(\mu)$.

Moreover, the density of $F(\mu)$ with respect to $\mathbb{W}$ is in $L^\gamma(S, \mathbb{W})$ for every finite $\gamma > 1$: Indeed, following standard computations (similarly to the proof of Theorem 4.5) and using assumption (4.42), we have for some $c_\gamma$,

$$\mathbb{E}\left[\left|\frac{dF(\mu)}{d\mathbb{W}}(W)\right|^\gamma\right] \le \mathbb{E}\left[\exp\left(c_\gamma \int_0^T |b_t(W_t, \mu_t)|^2 dt\right)\right] < \infty. \tag{4.46}$$

Finally, note that a process $X$, with $R(\mathrm{Law}(X)\|\mathbb{W}) < \infty$, is a solution to the McKean–Vlasov SDE if and only if its law is a fixed point for $F$. In particular, if $\bar{\mu}$ is the law of a solution

to the McKean–Vlasov SDE, with $R(\bar\mu\|\mathbb{W})<\infty$, then $\bar\mu\in L^\gamma(S,\mathbb{W})$ for every finite $\gamma$ and satisfies (4.46) (with $\bar\mu$ in place of $\mu$ and $F(\mu)$).

Let $\bar\mu$ and $\bar\nu$ be the laws of two solutions to the McKean–Vlasov SDE and $\mathcal{T}$ be the subset of $[0,T]$ consisting of all times $T_1$ such that $\bar\mu$ and $\bar\nu$ coincide when restricted to $[0,T_1]$ (that is, $\bar\mu_{[0,T_1]}=\bar\nu_{[0,T_1]}$). Clearly $\mathcal{T}$ is an interval containing 0. We will show that $\mathcal{T}$ is both open and closed in $[0,T]$, thus implying that $\mathcal{T}$ must coincide with $[0,T]$, and hence $\bar\mu=\bar\nu$.

The fact that $\mathcal{T}$ is closed follows easily from the fact that $\bar\mu$ and $\bar\nu$ are measures on continuous paths: indeed, if $(T_n)_n$ is an increasing sequence in $[0,T]$ converging to $T'$ and $\bar\mu$ and $\bar\nu$ coincide up to $T_n$ for every $n$, then $\bar\mu$ and $\bar\nu$ coincide up to $T'$ as well.

We will now show that $\mathcal{T}$ is open, that is, if $\bar\mu$ and $\bar\nu$ coincide up to $T_1$, then there exists $T_2>T_1$ such that $\bar\mu$ and $\bar\nu$ coincide up to $T_2$. Fix $\eta>0$ such that $\bar\mu$ and $\bar\nu$ belong to $M_\eta$ (such $\eta$ exists due to (4.46)). Fix also $T_2>T_1$ to be specified later. Unless differently specified, all constants in the proof are independent of $T_2$.

Using the elementary inequality $|e^a-e^b|\le\frac12(e^a+e^b)|a-b|$, we get

$$\begin{aligned}\left|\frac{d\bar\mu_{[0,T_2]}}{d\mathbb{W}}-\frac{d\bar\nu_{[0,T_2]}}{d\mathbb{W}}\right|&=\left|\frac{dF(\bar\mu)_{[0,T_2]}}{d\mathbb{W}}-\frac{dF(\bar\nu)_{[0,T_2]}}{d\mathbb{W}}\right|\\&\le\frac12\left(\frac{dF(\bar\mu)_{[0,T_2]}}{d\mathbb{W}}+\frac{dF(\bar\nu)_{[0,T_2]}}{d\mathbb{W}}\right)|V_{b,[0,T_2]}(\cdot,\bar\mu)-V_{b,[0,T_2]}(\cdot,\bar\nu)|.\end{aligned}$$

So by Hölder's inequality with $1/\gamma+1/\tilde m=1/m$ (with shorthand $L^\gamma=L^\gamma(S_{[0,T_2]},\mathbb{W}_{[0,T_2]})$, $\gamma>1$),

$$\left\|\frac{dF(\bar\mu)_{[0,T_2]}}{d\mathbb{W}}-\frac{dF(\bar\nu)_{[0,T_2]}}{d\mathbb{W}}\right\|_{L^m}\le\frac12\left(\left\|\frac{dF(\bar\mu)}{d\mathbb{W}}\right\|_{L^\gamma}+\left\|\frac{dF(\bar\nu)}{d\mathbb{W}}\right\|_{L^\gamma}\right)\|V_b(\cdot,\bar\mu)-V_b(\cdot,\bar\nu)\|_{L^{\tilde m}}.$$

By (4.46), (4.42) and the fact that $R(\rho\|\mathbb{W})\le\|d\rho/d\mathbb{W}\|^m_{L^m(S,\mathbb{W})}$ for any $\rho\in\mathcal{P}(S)$, we have, for some $C_\eta>0$,

$$\begin{aligned}\left\|\frac{dF(\bar\mu)}{d\mathbb{W}}\right\|_{L^\gamma}+\left\|\frac{dF(\bar\nu)}{d\mathbb{W}}\right\|_{L^\gamma}&\le\left\|\frac{dF(\bar\mu)}{d\mathbb{W}}\right\|_{L^\gamma(S,\mathbb{W})}+\left\|\frac{dF(\bar\nu)}{d\mathbb{W}}\right\|_{L^\gamma(S,\mathbb{W})}\\&\le2\sup_{\rho\in M_\eta}\mathbb{E}\left[\exp\left(c_\gamma\int_0^T|b_t(W_t,\rho_t)|^2dt\right)\right]\\&\le2\sup_{\rho\in\mathcal{P}(S),R(\rho\|\mathbb{W})\le\eta^\gamma}\mathbb{E}\left[\exp\left(c_\gamma\int_0^T|b_t(W_t,\rho_t)|^2dt\right)\right]\le C_\eta.\end{aligned}$$

By definition of $V_{b,[0,T_2]}$ (in Subsection 4.1 with final time $T_2$), we obtain

$$\begin{aligned}\|V_{b,[0,T_2]}(\cdot,\bar\mu)-V_{b,[0,T_2]}(\cdot,\bar\nu)\|_{L^{\tilde m}}&\le\left\|\int_0^{T_2}(b_t(W_t,\bar\mu_t)-b_t(W_t,\bar\nu_t))\cdot dW_t\right\|_{L^{\tilde m}}\\&\quad+\frac12\left\|\int_0^{T_2}\Big||b_t(W_t,\bar\mu_t)|^2-|b_t(W_t,\bar\nu_t)|^2\Big|dt\right\|_{L^{\tilde m}}=\text{(I)}+\text{(II)}.\end{aligned}$$

For the first term, we obtain by the Burkholder–Davis–Gundy inequality

$$\text{(I)}\le\tilde c_1\left\|\int_0^{T_2}|b_t(W_t,\bar\mu_t)-b_t(W_t,\bar\nu_t)|^2dt\right\|^{\frac12}_{L^{\tilde m/2}}\le c_1\left\|\int_0^{T_2}|b_t(W_t,\bar\mu_t)-b_t(W_t,\bar\nu_t)|^2dt\right\|^{\frac12}_{L^{\tilde m}}$$

for some constants $\tilde{c}_1, c_1 > 0$. As for the second term, we estimate as follows

$$\begin{aligned}
\text{(II)} &\leq \tilde{c}_2 \left\| \left( \int_0^{T_2} \Big( |b_t(W_t, \bar{\mu}_t)|^2 + |b_t(W_t, \bar{\nu}_t)|^2 \Big) dt \right)^{\frac{1}{2}} \cdot \left( \int_0^{T_2} |b_t(W_t, \bar{\mu}_t) - b_t(W_t, \bar{\nu}_t)|^2 dt \right)^{\frac{1}{2}} \right\|_{L^{\tilde{m}}} \\
&\leq c_2 \left( \left\| \int_0^{T_2} |b_t(W_t, \bar{\mu}_t)|^2 dt \right\|_{L^{\tilde{m}}}^{\frac{1}{2}} + \left\| \int_0^{T_2} |b_t(W_t, \bar{\nu}_t)|^2 dt \right\|_{L^{\tilde{m}}}^{\frac{1}{2}} \right) \left\| \int_0^{T_2} |b_t(W_t, \bar{\mu}_t) - b_t(W_t, \bar{\nu}_t)|^2 dt \right\|_{L^{\tilde{m}}}^{\frac{1}{2}}
\end{aligned}$$

for some constants $\tilde{c}_2, c_2 > 0$. Using the inequality $a^{\tilde{m}} \leq e^{\tilde{m}a}$, we obtain the estimate

$$\begin{aligned}
\left\| \int_0^{T_2} |b_t(W_t, \bar{\mu}_t)|^2 dt \right\|_{L^{\tilde{m}}}^{\frac{1}{2}} &\leq \mathbb{E} \left[ \exp \left( \tilde{m} \int_0^{T_2} |b_t(W_t, \bar{\mu}_t)|^2 dt \right) \right]^{\frac{1}{2\tilde{m}}} \\
&\leq \sup_{\rho \in M_\eta} \mathbb{E} \left[ \exp \left( \tilde{m} \int_0^{T} |b_t(W_t, \rho_t)|^2 dt \right) \right]^{\frac{1}{2\tilde{m}}} \leq C_\eta,
\end{aligned}$$

where we have used again (4.42) and the fact that $R(\rho \| \mathbb{W}) \leq \| d\rho / d\mathbb{W} \|_{L^m(S,\mathbb{W})}^m$. The same argument holds for the term with $\int_0^{T_2} |b_t(W_t, \bar{\nu}_t)|^2 dt$.

Putting the terms (I) and (II) together, and using that $\bar{\mu}_{[0,T_1]} = \bar{\nu}_{[0,T_1]}$, we then obtain

$$\begin{aligned}
\| V_{b,[0,T_2]}(\cdot, \bar{\mu}) - V_{b,[0,T_2]}(\cdot, \bar{\nu}) \|_{L^{\tilde{m}}} &\leq c_3 \left\| \int_0^{T_2} |b_t(W_t, \bar{\mu}_t) - b_t(W_t, \bar{\nu}_t)|^2 dt \right\|_{L^{\tilde{m}}}^{\frac{1}{2}} \\
&= c_3 \left\| \int_{T_1}^{T_2} |b_t(W_t, \bar{\mu}_t) - b_t(W_t, \bar{\nu}_t)|^2 dt \right\|_{L^{\tilde{m}}}^{\frac{1}{2}}
\end{aligned}$$

for some constant $c_3 > 0$. Using assumption (4.43) (with the notation $\boldsymbol{W} = (W^1, \ldots, W^N)$ and $\boldsymbol{W}^{,1} = (W^2, \ldots, W^N)$, where $W^i$ are independent $d$-dimensional Brownian motions with

initial law $\rho_0$), we further estimate the right-hand side to obtain

$$
\begin{aligned}
&\left\| \int_{T_1}^{T_2} |b_t(W_t, \bar{\mu}_t) - b_t(W_t, \bar{\nu}_t)|^2 dt \right\|_{L^{\tilde{m}}}^{1/2} \\
&\leq c_4 \mathbb{E}^{W^1} \left[ \left( \mathbb{E}^{\boldsymbol{W}, \mathbf{1}} \left[ \int_{T_1}^{T_2} g_t(\boldsymbol{W}_t) dt \left| \frac{d\bar{\mu}_{[0,T_2]}}{d\mathbb{W}}^{\otimes(k-1)} (\boldsymbol{W}^{\mathbf{1}}_{[0,T_2]}) - \frac{d\bar{\nu}_{[0,T_2]}}{d\mathbb{W}}^{\otimes(k-1)} (\boldsymbol{W}^{\mathbf{1}}_{[0,T_2]}) \right|^2 \right] \right)^{\tilde{m}} \right]^{1/(2\tilde{m})} \\
&\leq c_5 \mathbb{E}^{W^1} \left[ \left( \mathbb{E}^{\boldsymbol{W}, \mathbf{1}} \left[ \int_{T_1}^{T_2} g_t(\boldsymbol{W}_t)\, dt \cdot \right. \right. \right. \\
&\qquad \left. \left. \left. \cdot \sum_{j=2}^{k} \left| \frac{d\bar{\mu}_{[0,T_2]}}{d\mathbb{W}}(W^j) - \frac{d\bar{\nu}_{[0,T_2]}}{d\mathbb{W}}(W^j_{[0,T_2]}) \right|^2 \prod_{2 \leq \ell \leq k, \ell \neq j} \left( \frac{d\bar{\mu}_{[0,T_2]}}{d\mathbb{W}} \vee \frac{d\bar{\nu}_{[0,T_2]}}{d\mathbb{W}}(W^\ell_{[0,T_2]}) \right)^2 \right) \right]^{\tilde{m}} \right]^{1/(2\tilde{m})} \\
&\leq c_6 \sum_{j=2}^{k} \mathbb{E}^{W^1} \left[ \left( \mathbb{E}^{\boldsymbol{W}, \mathbf{1}} \left[ \left( \int_{T_1}^{T_2} g_t(\boldsymbol{W}_t)\, dt \right)^{(m/2)'} \right] \right)^{\tilde{m}/(m/2)'} \cdot \right. \\
&\quad \left. \cdot \left( \mathbb{E}^{\boldsymbol{W}, \mathbf{1}} \left[ \left| \frac{d\bar{\mu}_{[0,T_2]}}{d\mathbb{W}}(W^j_{[0,T_2]}) - \frac{d\bar{\nu}_{[0,T_2]}}{d\mathbb{W}}(W^j_{[0,T_2]}) \right|^m \prod_{2 \leq \ell \leq k, \ell \neq j} \left( \frac{d\bar{\mu}_{[0,T_2]}}{d\mathbb{W}} \vee \frac{d\bar{\nu}_{[0,T_2]}}{d\mathbb{W}}(W^\ell_{[0,T_2]}) \right)^m \right] \right)^{2\tilde{m}/m} \right]^{1/(2\tilde{m})} \\
&\leq c_7 \left( \mathbb{E} \left[ \int_{T_1}^{T_2} g_t(\boldsymbol{W}_t)\, dt \right]^{\tilde{m}} \right)^{1/(2\tilde{m})} \left\| \frac{d\bar{\mu}_{[0,T_2]}}{d\mathbb{W}} \vee \frac{d\bar{\nu}_{[0,T_2]}}{d\mathbb{W}} \right\|_{L^m}^{k-2} \left\| \frac{d\bar{\mu}_{[0,T_2]}}{d\mathbb{W}} - \frac{d\bar{\nu}_{[0,T_2]}}{d\mathbb{W}} \right\|_{L^m} \\
&\leq c_8 \left( \mathbb{E} \left[ \int_{T_1}^{T_2} g_t(\boldsymbol{W}_t)\, dt \right]^{\tilde{m}} \right)^{1/(2\tilde{m})} \eta^{k-2} \left\| \frac{d\bar{\mu}_{[0,T_2]}}{d\mathbb{W}} - \frac{d\bar{\nu}_{[0,T_2]}}{d\mathbb{W}} \right\|_{L^m}
\end{aligned}
$$

for appropriate constants $c_i > 0$, $i = 1, \ldots, 8$. We conclude that, for some $C_\eta > 0$ (independent of $T_1$ and $T_2$),

$$
\begin{aligned}
\left\| \frac{d\bar{\mu}_{[0,T_2]}}{d\mathbb{W}} - \frac{d\bar{\nu}_{[0,T_2]}}{d\mathbb{W}} \right\|_{L^m} &= \left\| \frac{dF(\bar{\mu})_{[0,T_2]}}{d\mathbb{W}} - \frac{dF(\bar{\nu})_{[0,T_2]}}{d\mathbb{W}} \right\|_{L^m} \\
&\leq C_\eta \left( \mathbb{E} \left( \int_{T_1}^{T_2} g_t(\boldsymbol{W}_t)\, dt \right)^{\tilde{m}} \right)^{1/(2\tilde{m})} \left\| \frac{d\bar{\mu}_{[0,T_2]}}{d\mathbb{W}} - \frac{d\bar{\nu}_{[0,T_2]}}{d\mathbb{W}} \right\|_{L^m}.
\end{aligned}
$$

By assumption (4.44), we can find $T_2 > T_1$ close enough to $T_1$ such that

$$
C_\eta \left( \mathbb{E} \left[ \int_{T_1}^{T_2} g_t(\boldsymbol{W}_t)\, dt \right]^{\tilde{m}} \right)^{1/(2\tilde{m})} < 1.
$$

Hence, for such $T_2$, we get that $d\bar{\mu}_{[0,T_2]}/d\mathbb{W} = d\bar{\nu}_{[0,T_2]}/d\mathbb{W}$, that is $\bar{\mu}$ and $\bar{\nu}$ coincide up to $T_2$. The proof is complete. □

**Remark 4.20.** As the proof shows, assumption (4.43) may be replaced by the weaker one

(4.47)
$$\int_{T_1}^{T_2} |b_t(x_t,\mu_t) - b_t(x_t,\nu_t)|^2 dt$$
$$\leq C_\eta \int_{S^{k-1}} \left( \int_{T_1}^{T_2} g(t,x_t,y_t)\, dt \right) \sum_{j=2}^{k} \left| u_{[0,T_2]}(y^j) - v_{[0,T_2]}(y^j) \right|^2 \prod_{\substack{2\leq \ell\leq k \\ \ell\neq j}} (u_{[0,T_2]} \vee v_{[0,T_2]})(y^\ell)^2 d\mathbb{W}^{\otimes(k-1)}(y)$$

for $\mathbb{W}$-a.e. $x$ and every $\mu,\nu \in M_\eta$,

where $u_{[0,T_2]} = d\mu_{[0,T_2]}/d\mathbb{W}$ and $v_{[0,T_2]} = d\nu_{[0,T_2]}/d\mathbb{W}$.

**Remark 4.21.** The proof of Theorem 4.19 shows, morally, a contraction bound for $F$ in $M_\eta$, provided that $T_2$ is sufficiently close to $T_1$. From this bound it should be possible to get local existence for the McKean–Vlasov SDE. Unfortunately, the iteration in time scheme does not work easily, because the constant $\eta$ bounding the $L^m$ norm of the solution and the time step $T_2 - T_1$ can change from one time step to the next. Nonetheless, in all the examples where our LPD result Theorem 4.5 applies, existence for the McKean-Vlasov SDE holds due to existence of a minimizer of the rate function $\mathcal{F}$.

**Corollary 4.22.** *Under the assumptions of Theorems 4.5 and 4.19 (possibly modified as in Remark 4.20), as $N \to \infty$, the family of empirical measures $z^N_{\boldsymbol{X}}$ associated with the particle system* (4.1) *converges almost surely to the (unique) law $\bar{\mu}$ of the McKean–Vlasov SDE* (4.41) *(with finite relative entropy with respect to $\mathbb{W}$).*

*Proof.* The zeros of the rate function $\mathcal{F}$ in Theorem 4.5 are exactly the laws of the solutions to the McKean–Vlasov SDE (4.41) with finite relative entropy with respect to $\mathbb{W}$. In particular, since $\mathcal{F}$ is a good rate function, there exists at least one law solution to the McKean–Vlasov SDE. By Theorem 4.19, such a law is unique (among the probability measures with finite entropy with respect to $\mathbb{W}$), and it is then the unique zero of the rate function $\mathcal{F}$ of the LDP for $(z^N_{\boldsymbol{X}})_N$ in Theorem 4.5. Hence Lemma 2.4 applies. □

**Remark 4.23.** As noted by Sznitman in [Szn98, Szn91], convergence in law of the empirical measures $z^N$ to the constant variable $\bar{\mu}$ implies that the sequence $\tilde{Q}^N_{b^N} \in \mathcal{P}(S^N)$ is $\bar{\mu}$-chaotic, in the sense that for every $k \in \mathbb{N}$,

$$\mathrm{Law}\left(X^{N,1},\dots,X^{N,k}\right) \to \bar{\mu}^{\otimes k},$$

weakly as $N \to \infty$ on $S = C([0,T];\mathbb{R}^d)$. In particular, we have a form of propagation of chaos, namely that for all $k \in \mathbb{N}$,

$$\mathrm{Law}\left(X^{N,1}_t,\dots,X^{N,k}_t\right) \to \bar{\mu}_t^{\otimes k}, \qquad \forall t \in [0,T].$$

Now we come back to the examples in Subsections 4.3.1, 4.3.2, 4.3.3, recalling the drift for the corresponding McKean–Vlasov SDEs, namely:

- for the example in Subsection 4.3.1,

$$b_t(x,\mu) = \int_{\mathbb{R}^d} \varphi_t(x,y)\, d\mu_t(y);$$

- for the example in Subsection 4.3.2,

$$b_t(x,\mu) = \int_{\mathbb{R}^{(k-1)d}} \varphi_t(x, y_2, \dots y_k)\, d\mu_t(y_2, \dots y_k);$$

- for the example in Subsection 4.3.3,

$$b_t(x,\mu) = \int_{\mathbb{R}^d} \Psi\left(x, y, \int_{\mathbb{R}^d} \varphi_t(x,z)\, d\mu_t(z), \int_{\mathbb{R}^d} \varphi_t(y,z)\, d\mu_t(z), \mu\right) d\mu_t(y).$$

**Proposition 4.24.** *For the examples in Subsections 4.3.1, 4.3.2, 4.3.3, under the assumptions of Propositions 4.10, 4.13 and 4.16 respectively, the corresponding McKean–Vlasov SDEs admit a unique solution $X$ with $R(\mathrm{Law}(X)\|\mathbb{W}) < \infty$, and the family of empirical measures $z^N_{\boldsymbol{X}}$ converges almost surely to the law of $X$.*

*Proof.* We have seen in the previous section that the examples above satisfy the assumptions of Theorem 4.5, so it is enough to verify the assumptions of Theorem 4.19, possibly modified as in Remark 4.20. Assumption (4.42) is a consequence of (4.14) in the proof of Theorem 4.5. Assumptions (4.43) and (4.44) are satisfied, with any given $m$ and $\tilde{m}$ as in Theorem 4.19:

- for the example in Subsection 4.3.1, taking $g(t,x,y) = |\varphi(t,x,y)|^2$;
- for the example in Subsection 4.3.2, taking $g(t,x_1,\dots x_k) = |\varphi(t,x_1,\dots x_k)|^2$.

We will not show the proof of this fact, which is similar to, and easier than, the next proof in the example in Subsection 4.3.3.

For the example in Subsection 4.3.3, we will show that assumptions (4.47) and (4.44) are satisfied, taking

$$g(t,x,y,z) = \left(\mathrm{Lip}(\Psi)^2 + L^2\right)\left(1 + |\varphi_t(x,z)|^2 + |\varphi_t(y,z)|^2\right),$$

where $L$ is the linear growth constant for $\Psi$ as in (4.38). We start showing (4.47). We denote, for $0 \le t \le T_2$,

$$\Psi_t(x,y,\mu^1_{[0,T_2]},\mu^2_{[0,T_2]},\mu^3_{[0,T_2]}) = \Psi\left(x_t, y_t, \int_{S_{[0,T_2]}} \varphi_t(x_t,z_t) d\mu^1_{[0,T_2]}(z), \int_{S_{[0,T_2]}} \varphi_t(y_t,z_t) d\mu^2_{[0,T_2]}(z), \mu^3_t\right),$$

and also $u_{[0,T_2]} = d\mu_{[0,T_2]}/d\mathbb{W}$ and $v_{[0,T_2]} = d\nu_{[0,T_2]}/d\mathbb{W}$. For simplicity of notation, in the following proof of (4.47), we often omit the subscript $[0,T_2]$ from $\mu$, $\nu$, $u$, $v$, $\mathbb{W}$, $S$ (up to the end of the proof of (4.47), every time we write $\mu$, $\nu$, ... we mean $\mu_{[0,T_2]}$, $\nu_{[0,T_2]}$). We have the following estimates, with a generic constant $C > 0$ (independent of $T_1$ and $T_2$): for any $T_1, T_2$

with $0 \le T_1 < T_2 \le T$,

$$\begin{aligned}
\int_{T_1}^{T_2} |b_t(x_t,\mu_t) - b_t(x_t,\nu_t)|^2 dt &= \int_{T_1}^{T_2} \left| \int_S \Psi_t(x,y,\mu,\mu,\mu)\, d\mu(y) - \int_S \Psi_t(x,y,\nu,\nu,\nu) d\nu(y) \right|^2 dt \\
&\le C \int_{T_1}^{T_2} \left| \int_S \Psi_t(x,y,\mu,\mu,\mu) d(\mu-\nu)(y) \right|^2 dt \\
&\quad + C \int_{T_1}^{T_2} \left| \int_S [\Psi_t(x,y,\mu,\mu,\mu) - \Psi_t(x,y,\nu,\mu,\mu)] d\nu(y) \right|^2 dt \\
&\quad + C \int_{T_1}^{T_2} \left| \int_S [\Psi_t(x,y,\nu,\mu,\mu) - \Psi_t(x,y,\nu,\nu,\mu)] d\nu(y) \right|^2 dt \\
&\quad + C \int_{T_1}^{T_2} \left| \int_S [\Psi_t(x,y,\nu,\nu,\mu) - \Psi_t(x,y,\nu,\nu,\nu)] d\nu(y) \right|^2 dt \\
&=: \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)}.
\end{aligned}$$

For the term (I), we have

$$\begin{aligned}
\text{(I)} &= C \int_{T_1}^{T_2} \left| \int_S \Psi_t(x,y,\mu,\mu,\mu) d(\mu-\nu)(y) \right|^2 dt \\
&\le CL^2 \int_{T_1}^{T_2} \left| \int_S \left( 1 + \left| \int_S \varphi_t(x_t,z_t) d\mu(z) \right| + \left| \int_S \varphi_t(y_t,z_t) d\mu(z) \right| \right) |u(y) - v(y)| d\mathbb{W}(y) \right|^2 dt \\
&\le CL^2 \int_{T_1}^{T_2} \left| \int_S \int_S \left(1 + |\varphi_t(x_t,z_t)| + |\varphi_t(y_t,z_t)|\right) |u(y) - v(y)| d\mathbb{W}(y) d\mu(z) \right|^2 dt \\
&= CL^2 \int_{T_1}^{T_2} \left| \int_{S^2} \left(1 + |\varphi_t(x_t,z_t)| + |\varphi_t(y_t,z_t)|\right) |u(y) - v(y)| u(z) d\mathbb{W}^{\otimes 2}(y,z) \right|^2 dt \\
&\le 3C \int_{S^2} \left( \int_{T_1}^{T_2} L^2 (1 + |\varphi(t,x_t,z_t)|^2 + |\varphi(t,y_t,z_t)|^2) dt \right) |u(y) - v(y)|^2 u(z)^2 d\mathbb{W}^{\otimes 2}(y,z).
\end{aligned}$$

For the term (II), we have

$$\begin{aligned}
\text{(II)} &\le C \int_{T_1}^{T_2} \int_S \mathrm{Lip}(\Psi)^2 \left| \int_S \varphi(t,x_t,z_t) d(\mu-\nu)(z) \right|^2 d\nu(y) dt \\
&= C \int_{T_1}^{T_2} \mathrm{Lip}(\Psi)^2 \left| \int_{S^2} \varphi(t,x_t,z_t)(u(z) - v(z)) u(y) d\mathbb{W}^{\otimes 2}(y,z) \right|^2 dt \\
&\le C \int_{S^2} \left( \int_{T_1}^{T_2} \mathrm{Lip}(\Psi)^2 |\varphi(t,x_t,z_t)|^2 dt \right) |u(z) - v(z)|^2 u(y)^2 d\mathbb{W}^{\otimes 2}(y,z),
\end{aligned}$$

where $u(y)$ has been added artificially to satisfy (4.47) with $k = 3$. As for (III),

$$\begin{aligned}
\text{(III)} &\leq C \int_{T_1}^{T_2} \int_S \mathrm{Lip}(\Psi)^2 \left| \int_S \varphi(t, y_t, z_t) d(\mu - \nu)(z) \right|^2 d\nu(y) dt \\
&= C \int_{T_1}^{T_2} \int_S \mathrm{Lip}(\Psi)^2 \left| \int_S \varphi(t, y_t, z_t)(u(z) - v(z)) d\mathbb{W}(z) \right|^2 v(y) d\mathbb{W}(y) dt \\
&\leq C \int_{S^2} \left( \int_{T_1}^{T_2} \mathrm{Lip}(\Psi)^2 |\varphi(t, y_t, z_t)|^2 dt \right) |u(z) - v(z)|^2 v(y)^2 d\mathbb{W}^{\otimes 2}(y, z).
\end{aligned}$$

In view of the term (IV), we recall that, for any $t \in [0, T_2]$,

$$d_{BL}(\mu_t, \nu_t) \leq d_{BL}(\mu_{[0,T_2]}, \nu_{[0,T_2]}),$$

which follows from the fact that the map $e_t : S_{[0,T_2]} \to \mathbb{R}^d$, $e_t(x) := x_t$, is 1-Lipschitz. Hence,

$$\begin{aligned}
\text{(IV)} &= C \int_{T_1}^{T_2} \left| \int_S [\Psi_t(x, y, \nu, \nu, \mu) - \Psi_t(x, y, \nu, \nu, \nu)] d\nu(y) \right|^2 dt \\
&\leq C \int_{T_1}^{T_2} \mathrm{Lip}(\Psi)^2 \, d_{BL}(\mu_t, \nu_t)^2 dt \\
&\leq C(T_2 - T_1) \, \mathrm{Lip}(\Psi)^2 \, d_{BL}(\mu, \nu)^2.
\end{aligned}$$

We also recall that the bounded Lipschitz metric $d_{BL}$ is bounded by the total variation distance $d_{TV}$. Therefore we derive, adding again an artificial term $u(z)$,

$$\begin{aligned}
\text{(IV)} &\leq C(T_2 - T_1) \, \mathrm{Lip}(\Psi)^2 d_{TV}(\mu, \nu)^2 = C(T_2 - T_1) \, \mathrm{Lip}(\Psi)^2 \|u - v\|_{L^1}^2 \\
&= C \left( \int_{T_1}^{T_2} \mathrm{Lip}(\Psi)^2 \, dt \right) \left( \int_{S^2} |u(y) - v(y)| u(z) d\mathbb{W}^{\otimes 2}(y, z) \right)^2 \\
&\leq C \int_{S^2} \left( \int_{T_1}^{T_2} \mathrm{Lip}(\Psi)^2 \, dt \right) |u(y) - v(y)|^2 u(z)^2 d\mathbb{W}^{\otimes 2}(y, z).
\end{aligned}$$

Putting together (I), (II), (III), (IV), we get (4.47).

Finally, to complete the proof, we verify assumption (4.44) for $g$ (with any fixed $m$ and $\tilde{m}$ as in Theorem 4.19). Note that it is enough to prove

$$\mathbb{E} \left[ \left( \int_0^T |\varphi_t(W_t^1, W_t^2)|^2 \, dt \right)^{\tilde{m}} \right] < \infty.$$

But this easily follows from the assumption (4.39) on $\varphi$,

$$\mathbb{E} \left[ e^{\beta \int_0^T |\varphi_t|^2 (W_t^1, W_t^2) \, dt} \right] < \infty, \qquad \forall \beta \in \mathbb{R},$$

which implies the finiteness of all moments of the variable $\int_0^T |\varphi_t(W_t^1, W_t^2)|^2 dt$. □

***Background.*** The convergence of the particle system to the McKean–Vlasov SDE, in the sense of Corollary 4.22, is classical in the case of Lipschitz bounded drift, see e.g. [Szn91]. The case of non-Lipschitz drift has also been treated in various works and we mention only some of them. In the context of the example in Subsection 4.3.1, the paper [JW18] proves the convergence, with quantitative bounds, for $\varphi$ in $W^{-1,\infty}$ (which includes our example) such that $\mathrm{div}(\varphi)$ is in $W^{-1,\infty}$. The work [GQ15] covers the case $\varphi = -\nabla\Phi$ with $\Phi(x) = |x|^\alpha$ with $\alpha$ in $(0, 1)$, which is a relevant example of Subsection 4.3.1 for $d \geq 2$ (see Remark 4.12).

Both in [JW18] and [GQ15] the initial conditions are assumed to be diffuse in a suitable sense. The paper [Tom20] (which has appeared after our paper came on arXiv), exploiting the technique in [JTT18], shows well-posedness and convergence of the particle system for $\varphi$ as in our example in Subsection 4.3.1 (namely $\varphi$ in $L^q_t(L^p_x)$ with $p, q$ satisfying (4.29)), also possibly non-translation invariant, with a condition of continuity of $\varphi$ outside a set of singular points. As examples of convergence in critical cases, that are not covered by our results, we recall [FJ17], for $\varphi = -\nabla\Phi$ with $\Phi(x) = \log|x|$, and [FHM14], for the 2D Navier-Stokes equations and the associated vortex system, that is $\varphi(x) = x^\perp/|x|^2$. Outside the context of Subsection 4.3.1, we already mentioned [Lac18], proving convergence in the $\tau$ topology when $b$ is bounded and satisfies a suitable continuity assumption in the measure argument. The paper [Jab19] proves convergence for a general measure-dependent drift, assuming a quite weak condition but depending on the solution to the McKean–Vlasov itself; the result is then applied to the case of bounded drifts. However we are not aware of a convergence result that covers our Corollary 4.22 and Proposition 4.24, although the recent work [HRZ22] (which has also appeared after our paper came on arXiv) proves convergence for a class of singular drifts similar to that in Proposition 4.24.

Concerning (weak or strong) uniqueness for McKean–Vlasov SDEs with irregular drifts, we mention [MS19, HvS21, CdR20, RZ21, MV20, FPZ19]. It is also worth mentioning [Del19] on a regularization by noise phenomenon, via an infinite-dimensional noise, for a related mean field game problem.

## Appendix A. Limits of convex functions

Let $V$ be a vector space. We consider a sequence of convex functions $\phi_n : V \to \overline{\mathbb{R}}$ (with domains $D_n$), and two convex functions $\phi_L$ and $\phi_U$ with $\phi_L \leq \phi_U$ on $V$ (with respective domains $D_L$, $D_U$), which will act as asymptotic lower and upper bounds for $\phi_n$ in a sense specified below.

Moreover, we consider pairs of sequences and points $(\{y_n\}, y) \in V^{\mathbb{N}} \times V$, which for simplicity will be referred to as the pairs $(y_n, y)$, and denote $(y_n, y) \in D$ for the statement $y \in D_U$ and $y_n \in D_n$ for every $n$. We then have the following approximation result.

**Theorem A.1.** *Let $\gamma > 1$, and let the sequence of pairs $(y_{\lambda,n}, y_\lambda)$ be such that for every $\lambda > 0$,*

$$\limsup_{n\to\infty} \phi_n(\gamma y_{\lambda,n}) < \infty, \tag{A.1a}$$

$$\phi_U(\gamma y_\lambda) < \infty, \tag{A.1b}$$

$$\phi_L(y_\lambda) \leq \liminf_{n\to\infty} \phi_n\big(y_{\lambda,n}\big) \leq \limsup_{n\to\infty} \phi_n\big(y_{\lambda,n}\big) \leq \phi_U(y_\lambda). \tag{A.2}$$

*Moreover, let the couple $(x_n, x)$ be such that there exists a $K \in \mathbb{R}$ such that for all $\beta \in \mathbb{R}$,*

$$\limsup_{\lambda\to 0} \limsup_{n\to\infty} \phi_n\big(\beta(y_{\lambda,n} - x_n)\big) \leq K, \tag{A.3a}$$

$$\limsup_{\lambda\to 0} \phi_U\big(\beta(y_\lambda - x)\big) \leq K. \tag{A.3b}$$

*Then for any $0 < \gamma' < \gamma$,*

$$\limsup_{n\to\infty} \phi_n(\gamma' x_n) < \infty, \tag{A.4a}$$

$$\phi_U(\gamma' x) < \infty, \tag{A.4b}$$

*and that*

$$\phi_L(x) \leq \liminf_{n\to\infty} \phi_n\big(x_n\big) \leq \limsup_{n\to\infty} \phi_n\big(x_n\big) \leq \phi_U(x). \tag{A.5}$$

**Remark A.2.** Some comments on these assumptions:

(i) Note that the convex functions $\phi_L, \phi_N, \phi$ are all allowed to be *improper*, i.e. allowed to be equal to $-\infty$ on their domain, or to have empty domain.

(ii) The constant $K$ in (A.3) is independent of $\beta$, which is crucial in proving the convergence. As an example, note that when $\phi_U$ is even with $\phi_U(0) = 0$ the assumptions imply that (A.4b) holds for $K = 0$ as well.

First, we will establish some technical properties for limits of convex functions: a generalization of the classical statement on continuity of convex function on the interior of their domains to certain pointwise limits of convex functions, and a result that shows how under (A.3a) the limits in $n, \lambda$ of $\phi_n(x), \phi_n(y_{\lambda,n})$ in effect 'commute'.

**Lemma A.3.** *Let $g_n : \mathbb{R} \to \overline{\mathbb{R}}$ be a sequence of convex functions such that for some $a, b \in \mathbb{R}$*

$$\limsup_{n\to\infty} g_n(a) < \infty, \qquad \limsup_{n\to\infty} g_n(b) < \infty.$$

*Then the functions*

$$\bar{g}(z) := \limsup_{n\to\infty} g_n(z), \qquad g(z) := \liminf_{n\to\infty} g_n(z)$$

*are both either equal to $-\infty$ on $(a,b)$, or finite and continuous on $(a,b)$.*

*Proof.* First, note that there exists a large enough $N$ such that for all $n \geq N$ both $g_n(a)$ and $g_n(b)$ are bounded from above, and in particular $[a,b] \subset D(g_n)$. Moreover, it is easy to verify that $\bar{g} = \limsup_{n\to\infty} g_n(z)$ is convex as well, $[a,b] \subset D(\bar{g})$, and with $\bar{g}$ bounded from above on $[a,b]$. We will show that either $g \equiv -\infty$ on $(a,b)$ or that $\limsup_{n\to\infty} \|g_n\|_\infty$ is finite — in which case we will derive that $g$ is Lipschitz continuous at any proper sub-interval $[c,d] \subsetneq [a,b]$ with $a < c < d < b$. The case for $\bar{g}$ follows from a similar argument and the fact that $\bar{g}$ is convex itself.

Now, suppose that $\limsup_{n\to\infty} \|g_n\|_\infty = \infty$. Since $\bar{g}(z)$ is bounded from above on $[a,b]$, it follows that there exists a subsequence $n' \in \mathbb{N}$ and a sequence $z_{n'} \in [a,b]$ such that

$$\lim_{n'\to\infty} g_{n'}(z_{n'}) = -\infty.$$

By compactness in $[a,b]$ we can choose a converging sequence $z_{n'} \to z^*$. We will treat the case $z^* \in (a,b)$ and $z^* = a, b$ separately. First, assume the former and fix $s \in (a,b)$. Then for any small enough $\epsilon > 0$ and large enough $n'$ such that $g_{n'}(z_{n'}) < \infty$ and $|z_{n'} - z^*| < \epsilon$ with $a + \epsilon < z^* < b - \epsilon$, we have by convexity of $g_{n'}$

$$\begin{aligned} g_{n'}(s) &\leq \frac{z_{n'} - s}{z_{n'} - a} g_{n'}(a) + \frac{s-a}{z_{n'} - a} g_{n'}(z_{n'}) \\ &\leq g_{n'}(a)^+ + \frac{s-a}{z^* - a - \epsilon} g_{n'}(z_{n'}), \end{aligned}$$

whenever $s \in (a, z_{n'}]$, and where $g_{n'}(a)^+ = \min(0, g_{n'}(a))$. Repeating the argument for $s \in [z_{n'}, b)$ we derive

$$\begin{aligned} \limsup_{n'\to\infty} g_{n'}(s) &\leq \limsup_{n'\to\infty} \max\left(g_{n'}^+(a), g_{n'}^+(b)\right) + \max\left(\frac{s-a}{z^* - a - \epsilon}, \frac{b-s}{b - z^* - \epsilon}\right) \limsup_{n'\to\infty} g_{n'}(z_{n'}) \\ &= -\infty, \end{aligned}$$

In particular we conclude that $g(s) := \liminf_{n\to\infty} g_n(s) = -\infty$. The case for $z^* = b$ (or $z^* = a$) is similar, using the fact that $s \in (a, z_n')$ for large enough $n'$.

Next, suppose otherwise, i.e. $\limsup_{n\to\infty} \|g_n\|_\infty < \infty$. Recall from classical convex analysis on $\mathbb{R}^d$ (e.g. similar to [EG15, Theorem 6.7]) that bounded convex functions on convex sets $O$ are uniformly Lipschitz on certain subsets $A \subsetneq O$ with $d(A, O^c) > 0$. In particular, when $f : [a,b] \to \mathbb{R}$ is convex and bounded, $a + \delta \leq c < d \leq b - \delta$, for some $\delta > 0$, the Lipschitz constant of $f$ on the interval $[c,d]$ is bounded by

$$K := 2\delta^{-1} \|f\|_\infty.$$

Since pointwise limits or pointwise minima of $K$-Lipschitz functions are also $K$-Lipschitz, and

$$g(z) = \lim_{n\to\infty} \left( \lim_{m\to\infty} \min_{n \leq l \leq m} g_l(z) \right),$$

it follows that $g$ is $K$-Lipschitz as well. Since $c, d$ are arbitrary, we conclude that $g$ is continuous on $(a,b)$. □

**Lemma A.4.** *Let $(\phi_n)_{n\in\mathbb{N}}$ be a sequence of convex functions on $V$ and $(y_{\lambda,n})$ a family of sequences and $(x_n)$ a sequence in $V$. Suppose there exists a $\gamma > 1$ such that*

$$\limsup_{n\to\infty} \phi_n\big(\gamma y_{\lambda,n}\big) < \infty, \qquad \textit{for all } \lambda > 0, \tag{A.6}$$

*and suppose there exists some constant $K \in \mathbb{R}$ such that*

$$\limsup_{\lambda\to 0}\limsup_{n\to\infty}\phi_n\big(\beta(y_{\lambda,n}-x_n)\big)\le K \qquad \text{for all } \beta\in\mathbb{R}. \tag{A.7}$$

*Then for any $0\le\gamma'<\gamma$,*

$$\limsup_{n\to\infty}\phi_n(\gamma' x_n)<\infty, \tag{A.8}$$

*and*

$$\lim_{\lambda\to 0}\limsup_{n\to\infty}\phi_n\big(y_{\lambda,n}\big)=\limsup_{n\to\infty}\phi_n\big(x_n\big), \tag{A.9a}$$

$$\lim_{\lambda\to 0}\liminf_{n\to\infty}\phi_n\big(y_{\lambda,n}\big)=\liminf_{n\to\infty}\phi_n\big(x_n\big). \tag{A.9b}$$

*Proof.* First, note that if $\phi$ is convex, we have for any $x,y\in V$ and $\alpha\in(0,1)$ for which both $\alpha^{-1}(y-x),(1-\alpha)^{-1}x\in D(\phi)$ that $x+y\in D(\phi)$

$$\phi(x+y)\le\alpha\phi\big(\alpha^{-1}x\big)+(1-\alpha)\phi\big((1-\alpha)^{-1}y\big),$$

and therefore,

$$\begin{aligned}\phi(y)&=\phi(y-x+x)\\&\le\alpha\phi\left(\alpha^{-1}(y-x)\right)+(1-\alpha)\phi\left((1-\alpha)^{-1}x\right).\end{aligned} \tag{A.10}$$

Now, we begin by establishing (A.8). By (A.6) and (A.7) there exists a large enough $N$ such that for all $n\ge N^*$, $\lambda$, $\beta\in R$, and all $\beta\in\mathbb{R}$ both $\gamma y_{\lambda,n}\in D_n$ and $\beta(x-y_\lambda)\in D_n$. Therefore, for any $\gamma'\in[0,\gamma)$, by (A.10) we have

$$\phi_n(\gamma' x_n)\le(1-\alpha)\phi_n\left((1-\alpha)^{-1}\gamma'(x_n-y_{\lambda,n})\right)+\alpha\phi_n(\gamma y_{\lambda,n}).$$

with $\alpha=\gamma'/\gamma$. Passing to the limes superior in $n$, we obtain

$$\limsup_{n\to\infty}\phi_n(\gamma' x_n)\le(1-\alpha)\limsup_{n\to\infty}\phi_n\left((1-\alpha)^{-1}\gamma'(x_n-y_{\lambda,n})\right)+\alpha\limsup_{n\to\infty}\phi_n(\gamma y_{\lambda,n}).$$

Note that the left-hand side is independent of $\lambda>0$, while the second term on the right-hand side is finite for every $\lambda>0$ by (A.6). Moreover, by (A.7) it follows that the first term on the right-hand side is finite for sufficiently small $\lambda>0$. Hence, the left-hand side is finite as well and in particular, we have shown (A.8).

Next, from (A.10) we find for any $\alpha\in(0,1)$

$$\begin{aligned}\phi_n(\alpha x_n)&\le(1-\alpha)\phi_n\left(\frac{\alpha}{1-\alpha}(x_n-y_{\lambda,n})\right)+\alpha\phi_n(y_{\lambda,n}),\\ \phi_n(y_{\lambda,n})&\le(1-\alpha)\phi_n\left((1-\alpha)^{-1}(y_{\lambda,n}-x_n)\right)+\alpha\phi_n\left(\alpha^{-1}x_n\right).\end{aligned}$$

Taking limits in $\lambda$ and $n$, we obtain

$$\limsup_{n\to\infty}\phi_n(\alpha x_n)\le(1-\alpha)K+\alpha\liminf_{\lambda\to 0}\limsup_{n\to\infty}\phi_n(y_{\lambda,n}), \tag{A.11a}$$

$$\limsup_{\lambda\to 0}\limsup_{n\to\infty}\phi_n(y_{\lambda,n})\le(1-\alpha)K+\alpha\limsup_{n\to\infty}\phi_n\left(\alpha^{-1}x_n\right), \tag{A.11b}$$

where we made use of (A.7). Note that $\bar g(z):=\limsup_{n\to\infty}\phi_n(zx_n)$ is bounded from above around $z=1$, and by Lemma A.3 is either continuous in $(0,\gamma')$ or $\bar g(z)\equiv-\infty$ in a neighborhood around $z=1$. In both cases, passing to the limit $\alpha\to 1$ in (A.11a) and (A.11a), we recover (A.9a).

Similarly, to establish (A.9b), we first note that after repeating the argument of (A.11), we find

$$\liminf_{n\to\infty} \phi_n(\alpha x_n) \le (1-\alpha)K + \alpha \liminf_{\lambda\to 0} \liminf_{n\to\infty} \phi_n(y_{\lambda,n}),$$
$$\limsup_{\lambda\to 0} \liminf_{n\to\infty} \phi_n(y_{\lambda,n}) \le (1-\alpha)K + \alpha \liminf_{n\to\infty} \phi_n\left(\alpha^{-1}x_n\right).$$

Set $g_n(z) := \phi_n(zx_n)$, $g(z) := \liminf_{n\to\infty} g_n(z)$, and recall that $\bar{g}(z)$ is bounded from above on $[0,\gamma']$. Hence, again applying Lemma A.4 and passing to the limit $z \to 1$ in $g(z)$, and $\alpha \to 1$ above, we conclude the proof. □

Now, as a special case, for when $\phi_n = \phi$, $x_n = x$ and $y_{\lambda,n} = y_\lambda$, we have the following statement.

**Corollary A.5.** *Let $\phi$ be a convex function on $V$. Suppose there is a $\gamma > 1$ such that $\phi(\gamma y_\lambda) < \infty$ for all $\lambda > 0$, and that there exists a constant $K \in \mathbb{R}$ such that*

$$\limsup_{\lambda\to 0} \phi\big(\beta(x - y_\lambda)\big) \le K \qquad \text{for all } \beta \in \mathbb{R}.$$

*Then for any $0 < \gamma' < \gamma$, $\phi(\gamma' x) < \infty$, and $\phi(x) = \lim_{\lambda\to 0} \phi(y_\lambda)$.*

Together, these results show how to connect $\phi_L(x)$ to $\phi_L(y_\lambda)$, $\phi_n(x_n)$ to $\phi_n(y_{\lambda,n})$, and $\phi_U(x)$ to $\phi_U(y_\lambda)$. Since by assumption $\phi_n(y_{\lambda,n})$ is connected to $\phi_L(y_\lambda)$ and $\phi_U(y_\lambda)$, we derive corresponding statements for $(x_n, x)$. Indeed, we now show how the above results imply Theorem A.1.

*Proof of Theorem A.1.* Applying Corollary A.5 to both $\phi_L$ and $\phi_U$ separately, we obtain from (A.1b) and (A.3b) the convergences

$$\phi_L(x) = \lim_{\lambda\to 0} \phi(y_\lambda), \qquad \phi_U(x) = \lim_{\lambda\to 0} \phi(y_\lambda).$$

Similarly, from (A.1a) and (A.3a) it follows from Lemma A.4 that

$$\lim_{\lambda\to 0} \limsup_{n\to\infty} \phi_n\big(y_{\lambda,n}\big) = \limsup_{n\to\infty} \phi_n\big(x_n\big),$$
$$\lim_{\lambda\to 0} \liminf_{n\to\infty} \phi_n\big(y_{\lambda,n}\big) = \liminf_{n\to\infty} \phi_n\big(x_n\big).$$

Finally, we use the above and the relationship between $y_{\lambda,n}$ and $y_\lambda$ assumed in (A.2), to obtain

$$\begin{aligned}\phi_L(x) = \lim_{\lambda\to 0} \phi_L(y_\lambda) &\le \lim_{\lambda\to 0} \liminf_{n\to\infty} \phi_n\big(y_{\lambda,n}\big) \\ &= \liminf_{n\to\infty} \phi_n\big(x_n\big) \qquad \le \limsup_{n\to\infty} \phi_n\big(x_n\big) \\ &\qquad\qquad\qquad\qquad = \lim_{\lambda\to 0} \limsup_{n\to\infty} \phi_n\big(y_{\lambda,n}\big) \le \lim_{\lambda\to 0} \phi_U\big(y_\lambda\big) = \phi_U\big(x\big).\end{aligned}$$

The boundedness conditions follow similarly. □

## Appendix B. Variational representation of exponential integrals

Below we will give an extension of the classic variational formulation of exponential integrals. This is exploited in various arguments of the paper.

**Lemma B.1.** *Let $(X, \mathcal{A})$ be a measurable space and $V : X \to [0, \infty]$ be a non-negative $\mathcal{A}$-measurable function. Then*

$$\log \int_X e^V d\mu = \sup_{\nu \in D} \left\{ \int_X V \, d\nu - R(\nu \| \mu) \right\}, \tag{B.1}$$

*where*

$$D = \left\{ \nu \in \mathcal{P}(X) \,\middle|\, R(\nu \| \mu) < \infty \right\}.$$

In the case of Polish spaces $X$, the result also follows directly from [DE97, Proposition 4.5.1], which deals with potentials $V$ that are either bounded from below or from above. Nevertheless, for completeness, we give here a direct and elementary proof.

*Proof.* We will first prove that for every $\nu \in D$,

$$\int_X V \, d\nu - R(\nu \| \mu) \leq \log \int_X e^V d\mu, \tag{B.2}$$

and then we will show by approximation of bounded functions that

$$\log \int_X e^V d\mu \leq \sup_{\nu \in D} \left\{ \int_X V \, d\nu - R(\nu \| \mu) \right\}, \tag{B.3}$$

which together will imply (B.1).

For the first inequality, we take any $\nu \in D$ and assume, without loss of generality, that $e^V$ is integrable with respect to $\mu$. In this case, Young's inequality (in the form $ab \leq e^a - b + b \log b$) and the finiteness of $R(\nu \| \mu) < +\infty$ imply the finiteness of $\int_X V \, d\nu$. Let $\mu_V \in \mathcal{P}(S)$ be defined by

$$\mu_V = \frac{1}{Z_V} e^V \mu \qquad \text{with} \qquad Z_V = \int_X e^V d\mu.$$

Now let us rewrite the expression on the left-hand side of (B.2) as follows,

$$\begin{aligned} \int_X V \, d\nu - R(\nu \| \mu) &= \int_X V \, d\nu - \int_X \log \frac{d\nu}{d\mu} \, d\nu \\ &= \log \int_X e^V d\mu + \int_X \log e^V d\nu - \log \int_X e^V d\mu - \int_X \log \frac{d\nu}{d\mu} \, d\nu \\ &= \log \int_X e^V d\mu - \int_X \log \left( \frac{d\nu}{d\mu} \left( \frac{d\mu_V}{d\mu} \right)^{-1} \right) d\nu \\ &= \log \int_X e^V d\mu - R(\nu \| \mu_V) \end{aligned} \tag{B.4}$$

By non-negativity of the entropy $R(\cdot \| \mu_V)$, (B.2) holds for any $\nu \in D$.

To show (B.3), consider the sequence $V_n = \min\{V, n\}$ for $n \in \mathbb{N}$. Note that $V_n$ is a non-decreasing sequence of non-negative bounded functions converging pointwise to $V$. Since $V_n$ is bounded, we have that $R(\mu_{V_n} \| \mu) < +\infty$ and that $\int_X e^{V_n} d\mu < \infty$, and we define again $\mu_{V_n}$ as above. Hence, repeating the argument of (B.4) for $V_n$ it follows that

$$\int_X V_n \, d\nu - R(\nu \| \mu) = \log \int_X e^{V_n} \, d\mu - R(\nu \| \mu_{V_n}).$$

In particular, taking $\nu = \mu_{V_n}$, we get

$$\log \int_X e^{V_n}\, d\mu = \int_X V_n\, d\mu_{V_n} - R(\mu_{V_n}\|\mu) \le \int_X V\, d\mu_{V_n} - R(\mu_{V_n}\|\mu)$$

Maximizing over $n$, we conclude (B.3). □

**Remark B.2.** In the proof for bounded $V$ the equality in B.1 follows by checking that $\nu = \mu_V$ is the unique maximizer (B.1). However, this is not possible for a general unbounded $V$, even when $e^V$ is integrable with respect to $\mu$. A simple example over $X = [0, 1/2]$ follows from taking $V(x) = -\log x + \alpha \log\log x^{-1} + 1$ with any $1 < \alpha < 2$: for this example, $\mu_V$ is well defined and clearly $R(\mu_V\|\mu_V) = 0$, but both $R(\mu_V\|\mu)$ and $\int V d\mu_V$ are infinite. In particular, $\mu_V$ does not belong to $D$ and is not a maximizer of (B.1), even though $\mu_{V_n}$ is a maximizing sequence.

In contrast, if $V$ satisfies a slightly stronger exponential integrability condition, then $R(\nu\|\mu_V) < \infty$ does imply $R(\nu\|\mu) < \infty$ and $\int V d\nu < \infty$, as we show below.

**Corollary B.3.** *Let $V : X \to \overline{\mathbb{R}}$ be a measurable function and $\mu_V = Z_V^{-1} e^V \mu$ with normalization constant $Z_V$. Further, suppose there exists $\gamma > 1$ such that*

$$\int_X e^{\gamma|V|} d\mu < \infty.$$

*Then*

$$R(\nu\|\mu_V) < \infty \iff R(\nu\|\mu) < \infty.$$

*Moreover, the following equality holds:*

$$R(\nu\|\mu_V) = R(\nu\|\mu) - \int_X V d\nu + \log \int_X e^V d\mu. \tag{B.5}$$

*Proof.* Repeating carefully the proof of (B.4), one gets formula (B.5) if $V$ is in $L^1(\nu)$ and one of the two conditions $R(\nu\|\mu) < \infty$ and $R(\nu\|\mu_V) < \infty$ holds. Hence it remains to show that each of the conditions $R(\nu\|\mu) < \infty$ and $R(\nu\|\mu_V) < \infty$ implies that $V$ is in $L^1(\nu)$.

In the case $R(\nu\|\mu) < \infty$, by (B.2) we have

$$\int_X |V|\, d\nu \le R(\nu\|\mu) + \log \int_X e^{|V|} d\mu,$$

which is finite by assumption.

In the case $R(\nu\|\mu_V) < \infty$, we apply again (B.2) but with base measure $\mu_V$ instead of $\mu$ and $(\gamma - 1)V$ instead of $V$, getting

$$\begin{aligned}\int_X (\gamma - 1)|V|\, d\nu &\le R(\nu\|\mu_V) + \log \int_X e^{(\gamma-1)|V|} d\mu_V \\ &\le R(\nu\|\mu_V) + \log \int_X e^{\gamma|V|} d\mu - \log \int_X e^V d\mu,\end{aligned}$$

which is finite by assumption. The proof is complete. □

**Lemma B.4.** *Let $F : X^k \to [0, \infty)$, $k \in \mathbb{N}$ be a nonnegative measurable function satisfying*

$$\int_{X^k} \exp(F(x_1, \dots, x_k))\, d\mu^{\otimes k} < \infty,$$

*and $\nu \in \mathcal{P}(S)$ is such that $R(\nu\|\mu) < \infty$. Then*

$$\log \int_X \exp\left(\int_{X^{k-1}} F(x,y)\, d\nu^{\otimes k-1}(y)\right) d\mu(x) \leq (k-1)R(\nu\|\mu) + \log \int_{X^k} e^F d\mu^{\otimes k}.$$

*Proof.* A simple application of Lemma B.1 yields

$$\begin{aligned}
&\log \int_X \exp\left(\int_{X^{k-1}} F(x,y)\, d\nu^{\otimes k-1}(y)\right) d\mu(x) \\
&\quad = \sup_\rho \left\{ \left\langle \rho, \int_{X^{k-1}} F(x,y)\, d\nu^{\otimes k-1}(y) \right\rangle - R(\rho\|\mu) \right\} \\
&\quad = \sup_\rho \left\{ \langle \rho \otimes \nu^{\otimes k-1}, F\rangle - R(\rho \otimes \nu^{\otimes k-1} \| \mu \otimes \mu^{\otimes k-1}) \right\} + (k-1)R(\nu\|\mu) \\
&\quad \leq \sup_\sigma \left\{ \langle \sigma, F\rangle - R(\sigma\|\mu^{\otimes k}) \right\} + (k-1)R(\nu\|\mu) = \log \int_{X^k} e^F d\mu^{\otimes k} + (k-1)R(\nu\|\mu),
\end{aligned}$$

which is the desired estimate. □

## Appendix C. On measurability and exponential approximations

C.1. **Measurability of integral maps.** In this subsection we give a measurability result for integral maps. This in particular implies measurability of $\mathcal{E}_V$ and $\mathcal{E}_V^N$ in Section 3. In the following, $X$ and $Y$ are Polish spaces, we recall that $\mathcal{P}(Y)$ endowed with the weak topology is also Polish; $X$, $Y$ and $\mathcal{P}(Y)$ are endowed with their Borel $\sigma$-algebras.

**Theorem C.1.** *Given any Borel function*

$$X \times Y \times \mathcal{P}(Y) \ni (x,y,\mu) \mapsto f(x,y,\mu) \in [-\infty, +\infty],$$

*then the set $\{\mu \in \mathcal{P}(Y) \mid f \in L^1(\mu)\}$ is Borel and the mapping*

$$F_f : X \times \mathcal{P}(Y) \ni (x,\mu) \mapsto 1_{f \in L^1(\mu)} \int_Y f(x,y,\mu)\, d\mu(y)$$

*is also Borel.*

The case when $f(x,y,\mu) = W(y)$ for some measurable $W$ is a classical question, and treated in great generalization in for example [Bog07, Chapter 8]. However, for simplicity, in our setting we stick to the case of metric spaces, and adapt an argument of [Jor06] (Theorem 15.13). First, we provide a generalization of Lemma 7.3.12 of [DZ10].

**Lemma C.2.** *Suppose that $f$ is in $C_b(X \times Y \times \mathcal{P}(Y))$. Then $F_f$ is in $C_b(X \times \mathcal{P}(Y))$.*

*Proof.* In the following, we denote $Z = X \times Y \times \mathcal{P}(Y)$. The idea of the proof is similar to the fact that for Polish spaces $X$ and $Y$, the set of functions $\{f(x)g(y) \mid f \in C_b(X), g \in C_b(Y)\}$ is convergence determining for $\mathcal{P}(X \times Y)$ (see for example Theorem 3.4.5b of [EK05]), which is used in Lemma 7.3.12 of [DZ10]. Namely, first note for any $g(x,y,\mu) := g_1(x)g_2(y)g_3(\mu)$, with $g_1 \in C_b(X)$, $g_2 \in C_b(Y)$ and $g_3 \in C_b(\mathcal{P}(Y))$, the boundedness and continuity of $F_g$ is trivial.

Now, consider a sequence $(x_n, \mu_n)_n$ converging to $(x^*, \mu^*)$. In particular the subsets $L := \{x_n\} \cup \{x^*\} \subset X$ and $M := \{\mu_n\}_{n\geq 1} \cup \{\mu^*\} \subset \mathcal{P}(Y)$ are compact in $X$ and $\mathcal{P}(Y)$ respectively. In particular, the set $M \subset \mathcal{P}(Y)$ is tight, i.e., for every $\epsilon > 0$ there exists a compact set $K_\epsilon \subset Y$ such that $\mu^*(K^c) < \epsilon$ and $\mu_n(K_\epsilon^c) < \epsilon$ for all $n \geq 1$.

Therefore, fix any $\epsilon > 0$. By applying Stone-Weierstrass on the compact set $B_\epsilon = L \times K_\epsilon \times M$, we find a sequence $(g^{\epsilon,l})_{l\geq 1} \in C_b(B_\epsilon)$ with $g^{\epsilon,l}(x,y,\mu) := g_1^{\epsilon,l}(x)g_2^{\epsilon,l}(y)g_3^{\epsilon,l}(\mu)$, where $g_1^{\epsilon,l} \in C_b(L)$, $g_2^{\epsilon,l} \in C_b(K_\epsilon)$ and $g_3^{\epsilon,l} \in C_b(M)$, such that

$$\lim_{l\to\infty} \|f - g^{\epsilon,l}\|_{C_b(B_\epsilon)} = 0.$$

Now also fix $l \in \mathbb{N}$. By Tietze's extension theorem we find a $\tilde{g}_2^{\epsilon,l} \in C_b(Y)$ such that

$$\tilde{g}_2^{\epsilon,l}(y) = g_2^{\epsilon,l}(y) \text{ on } K_\epsilon, \qquad \|\tilde{g}_2^{\epsilon,l}\|_{C_b(Y)} \leq \|g_2^{\epsilon,l}\|_{C_b(K_\epsilon)}.$$

Hence, define on $\tilde{B} = L \times Y \times M$ the continuous function $\tilde{g}^{\epsilon,l}(x,y,\mu) := g_1^{\epsilon,l}(x)\tilde{g}_2^{\epsilon,l}(y)g_3^{\epsilon,l}(\mu)$, and note that by construction $\|\tilde{g}^{\epsilon,l}\|_{C_b(\tilde{B})} \leq \|g^{\epsilon,l}\|_{C_b(B_\epsilon)}$.

We now compute for any $l, n \geq 1$,

$$\begin{aligned}
|F_f(x^*,\mu^*) - F_f(x_n,\mu_n)| &= \left| \int_Y f(x^*,y,\mu^*)\,d\mu^*(y) - \int_Y f(x_n,y,\mu_n)\,d\mu_n(y) \right| \\
&\leq \left| \int_{K_\epsilon} f(x^*,y,\mu^*)\,d\mu^*(y) - \int_{K_\epsilon} f(x_n,y,\mu_n)\,d\mu_n(y) \right| + 2\epsilon\|f\|_{C_b(Z)} \\
&\leq \left| \int_{K_\epsilon} g^{\epsilon,l}(x^*,y,\mu^*)\,d\mu^*(y) - \int_{K_\epsilon} g^{\epsilon,l}(x_n,y,\mu_n)\,d\mu_n(y) \right| \\
&\qquad + 2\|f - g^{\epsilon,l}\|_{C_b(B_\epsilon)} + 2\epsilon\|f\|_{C_b(Z)} \\
&\leq \left| \int_Y \tilde{g}^{\epsilon,l}(x^*,y,\mu^*)\,d\mu^*(y) - \int_Y \tilde{g}^{\epsilon,l}(x_n,y,\mu_n)\,d\mu_n(y) \right| \\
&\qquad + 2\epsilon\|g^{\epsilon,l}\|_{C_b(B_\epsilon)} + 2\|f - g^{\epsilon,l}\|_{C_b(B_\epsilon)} + 2\epsilon\|f\|_{C_b(Z)}.
\end{aligned}$$

Thus, by the continuity of $F_{\tilde{g}^{\epsilon,l}}$ on $\tilde{B}$,

$$\limsup_{n\to\infty} |F_f(x^*,\mu^*) - F_f(x_n,\mu_n)| \leq 2\epsilon\|g^{\epsilon,l}\|_{C_b(B_\epsilon)} + 2\|f - g^{\epsilon,l}\|_{C_b(B_\epsilon)} + 2\epsilon\|f\|_{C_b(Z)}.$$

Taking subsequent limits, first in $l \to \infty$ and then in $\epsilon \to 0$ we conclude the proof. □

Next, we paraphrase Theorem 4.33 of [Jor06].

**Theorem C.3** (Monotone class theorem)**.** *Let $X$ be a metrizable space, and let $\mathcal{F}$ be a vector subspace of $B_b(X)$ including $C_b(X)$. Then $\mathcal{F} = B_b(X)$ if and only if $\mathcal{F}$ is closed under monotone sequential pointwise limits in $B_b(X)$.*

*Proof of theorem C.1.* Let $\mathcal{F}$ be the set of bounded Borel-measurable functions given by

$$\mathcal{F} = \Big\{ f \in B_b(X \times Y \times \mathcal{P}(Y)) \;\Big|\; F_f : X \times \mathcal{P}(Y) \to \mathbb{R} \text{ is Borel-measurable} \Big\}.$$

It is clear that $\mathcal{F}$ is a vector subspace of $B_b(X \times Y \times \mathcal{P}(Y))$, and by Lemma C.2 it contains $C_b(X \times Y \times \mathcal{P}(Y))$. To show stability under monotone pointwise limit, consider any sequence $f_n$ in $\mathcal{F}$ with $f_n \uparrow f$ and $f$ in $B_b(X \times Y \times \mathcal{P}(Y))$. For any measure $\mu \in \mathcal{P}(Y)$, it follows from monotone convergence that

$$\lim_{n\to\infty} F_{f_n}(x,\mu) = \lim_{n\to\infty} \int_Y f_n(x,y,\mu)\,d\mu(y) = \int_Y f_n(x,y,\mu)\,d\mu(y) = F_f(x,\mu).$$

Hence by the monotone class theorem, we get that $\mathcal{F} = B_b(X \times Y \times \mathcal{P}(Y))$, that is $F_f$ is Borel for any Borel bounded function $f$.

For a Borel non-negative function $f$, it follows from approximation with bounded function and monotone convergence that

$$(x,\mu) \mapsto \int_Y f(x,y,\mu)\, d\mu(y)$$

is Borel, in particular the set $\{\mu \mid f \in L^1(\mu)\}$ is Borel. The case of a general Borel function $f$ follows splitting $f$ into $f^+$ and $f^-$. The proof is complete. □

**Corollary C.4.** *Let $S$ be a Polish space and let $V : S^k \to \overline{\mathbb{R}}$ be Borel-measurable. Then both $\mathcal{E}_V, \mathcal{E}_V^N : \mathcal{P}(S) \to \overline{\mathbb{R}}$ are also Borel-measurable.*

C.2. **Exponential approximation theorem in Polish spaces.**

**Theorem C.5.** *Suppose $X$ is Polish, $V : X \to \overline{\mathbb{R}}$ is Borel measurable such that*

$$\int_X e^{\beta|V|}\, d\mu < +\infty \qquad \text{for any } \beta \geq 0. \tag{C.1}$$

*Then there exists continuous bounded functions $V_\lambda : S \to \mathbb{R}$ such that*

$$\lim_{\lambda\to 0} \log \int_X e^{\beta|V-V_\lambda|}\, d\mu = 0 \qquad \text{for any } \beta \geq 0.$$

The argument is similar to the denseness of $C_b(S)$ in $L^p(S)$ spaces, and rely on Tietze's extension theorem and Lusin's Theorem, see for example [Jor06, Theorems 2.47 and 12.8] (note that any Polish space $S$ is normal, and so these theorems apply).

*Proof of Theorem C.5.* First, suppose that $V$ is bounded. Then by Lusin's theorem, for any $\epsilon > 0$, there exists a compact set $K_\epsilon$ such that $\mu(S \setminus K_\epsilon) < \epsilon$.

By Tietze's extension theorem, we obtain a continuous function $V_\epsilon$ such that $V_\epsilon = V$ on the compact set $K_\epsilon$, and satisfying $\sup_{x\in S} |V_\epsilon(x)| \leq \sup_{x\in S} |V(x)|$. Therefore, for any $\beta \geq 0$,

$$\begin{aligned}\int_X e^{\beta|V-V_\epsilon|}\, d\mu &= \int_{K_\epsilon} e^{\beta|V-V_\epsilon|}\, d\mu + \int_{S\setminus K_\epsilon} e^{\beta|V-V_\epsilon|}\, d\mu \\ &= \int_{K_\epsilon} e^0\, d\mu + \int_{S\setminus K_\epsilon} e^{\beta|V-V_\epsilon|}\, d\mu \leq 1 + \epsilon \cdot e^{2\beta \sup_{x\in S}|V|(x)}.\end{aligned}$$

Note that the choices $\epsilon, K_\epsilon, V_\epsilon$ are independent of $\beta$, and thus, by continuity of the logarithm,

$$\lim_{\epsilon\to 0} \log \int_X e^{\beta|V-V_\epsilon|}\, d\mu = 0.$$

Next, for the case of unbounded $V$, suppose (C.1) holds. Let $V_n$ be the bounded truncation of $V$ to the interval $[-n, n]$, i.e.

$$V_n := \min\{n, \max\{-n, V\}\}, \qquad n \in \mathbb{N}.$$

It is clear that $V_n$ is bounded and $\lim_{n\to\infty} |V - V_n|(x) \to 0$ pointwise in $x \in X$. Moreover, $|V - V_n| \leq |V|$ for all $n \in \mathbb{N}$. Hence, by the assumption on $V$ and the dominated convergence,

$$\lim_{n\to\infty} \int_X e^{\beta|V-V_n|}\, d\mu = 1 \qquad \text{for any } \beta \geq 0,$$

which implies that

$$\lim_{n\to\infty} \log \int_X e^{\beta|V-V_n|}\, d\mu = 0 \qquad \text{for any } \beta \geq 0.$$

From the previous argument on bounded functions, we obtain, for each $n \in \mathbb{N}$, a sequence of bounded continuous functions $(V_{n,\epsilon})_{\epsilon>0}$ such that

$$\lim_{\epsilon\to 0} \log \int_X e^{\beta|V_n - V_{n,\epsilon}|}\, d\mu = 0 \qquad \text{for any } \beta \geq 0.$$

Finally, since by convexity,

$$\log \int_X e^{\beta|V - V_{n,\epsilon}|}\, d\mu \leq \frac{1}{2} \log \int_X e^{2\beta|V - V_n|}\, d\mu + \frac{1}{2} \log \int_X e^{2\beta|V_n - V_{n,\epsilon}|}\, d\mu,$$

we can find an appropriate sequence $V_\lambda := V_{n(\lambda),\epsilon(\lambda)}$ such that

$$\lim_{\lambda\to 0} \log \int_X e^{\beta|V - V_\lambda|}\, d\mu = 0 \qquad \text{for any } \beta \geq 0,$$

thereby concluding the proof. □

## Appendix D. Stochastic estimates and technical parts of Section 4

### D.1. Basic facts on the Girsanov theorem and the Novikov condition.

**Theorem D.1** (Girsanov)**.** *Let $W$ be an $m$-dimensional Brownian motion on a filtered probability space (satisfying the standard assumption), with general initial law* $\mathrm{Law}(W_0)$*, let $b : [0,T] \times \mathbb{R}^m \to \mathbb{R}^m$ be a Borel function. Consider the SDE*

$$\begin{aligned} dX_t &= b(t, X_t)\, dt + dW_t, \qquad t \in [0,T], \\ \mathrm{Law}(X_0) &= \mathrm{Law}(W_0). \end{aligned} \tag{D.1.1}$$

*If*

$$\mathbb{E}\left[ e^{\frac{1}{2}\int_0^T |b(t,W_t)|^2\, dt} \right] < \infty, \tag{D.1.2}$$

*then there exists a weak solution to* (D.1.1)*. Its law $\tilde{P}^X$ is equivalent to the Wiener measure $\mathbb{W}$ (starting with the same initial law of $X_0$) and satisfies,*

$$\frac{dP^X}{d\mathbb{W}}(W) := \exp\left( -\frac{1}{2}\int_0^T |b(t,W_t)|^2\, dt + \int_0^T b(t,W_t) \cdot dW_t \right).$$

*Moreover, if $Y$ is another weak solution to* (D.1.1) *(defined possibly on another probability space satisfying the standard assumption), with law $\tilde{P}^Y$, such that*

$$\int_0^T |b(t,Y_t)|^2 dt < \infty \quad \tilde{P}^Y\text{-a.s.}, \tag{D.1.3}$$

*then $\tilde{P}^Y$ coincides with $\tilde{P}^X$.*

This result is classical, here we recall uniqueness, in the line of [Fed09], Section 3.

*Proof.* Existence and the representation formula are a classical consequence of Girsanov's theorem, which can be applied thanks to Novikov's condition (D.1.2). When the initial distribution $\rho_0$ is a Dirac delta, uniqueness follows from [LS01, Theorem 7.7]. Uniqueness in the general case follows from conditioning $X_0$ to be a single point. □

**Lemma D.2.** *Let $W$ be an $m$-dimensional Brownian motion and let $b : \mathbb{R}^m \to \mathbb{R}$ be a Borel function such that*

$$\mathbb{E}\left[e^{2\int_0^T |b(t,W_t)|^2 dt}\right] < \infty.$$

*Then*

$$\mathbb{E}\left[e^{\int_0^T b(t,W_t)\cdot dW_t}\right] \leq \mathbb{E}\left[e^{2\int_0^T |b(t,W_t)|^2 dt}\right]^{\frac{1}{2}}$$

*Proof.* The proof is classical, we sketch the idea: By Novikov's criterion, the exponential local martingale

$$\exp\left(-\frac{1}{2}\int_0^T |2b(t,W_t)|^2 dt + \int_0^T (2b)(t,W_t)\cdot dW_t\right)$$

is a true martingale, in particular it has expectation 1. Then it is enough to apply Hölder inequality to get the required estimate. □

D.2. **Definition of the log-densities in Subsection 4.1.** Here we give the precise definition of $\mathcal{E}_b^N$ and $\mathcal{E}_b$ in Subsection 4.1.

We recall that $S = C([0,T];\mathbb{R}^d)$ and $\mathbb{W}$ is the Wiener measure on $S$. For $b : [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ Borel function, we define

$$V_b^1(x,\mu) := \int_0^T |b_t(x_t,\mu_t)|^2 dt, \quad x \in S,\ \mu \in \mathcal{P}(S).$$

Note that $V_b^1$ is a Borel map (by Fubini theorem applied to the map $(t,x,\mu) \mapsto |b_t(x_t,\mu_t)|^2$). For $\mu$ in $\mathcal{P}(S)$, if $V_b^1(x,\mu)$ is finite for $\mathbb{W}$-a.e. $x$, then we can define the stochastic integral

$$\int_0^T b_t(x_t,\mu_t)\cdot dx_t, \tag{D.2.1}$$

on the space $(S,(\mathcal{G}_t)_t,\mathbb{W})$, where $\mathcal{G}_t$ is the $\sigma$-algebra generated on $S$ by the $\mathbb{W}$-negligible sets and by the projection $\pi_{[0,t]} : S \to S_t = C([0,t];\mathbb{R}^d)$ on $[0,t]$; in particular $\mathcal{G}_T$ is the completion of $\mathcal{B}(S)$ with respect to $\mathbb{W}$. Hence we can define a map $V_b^2(\cdot,\mu) : S \to \mathbb{R}$ which is a representative of the stochastic integral (D.2.1) and is measurable with respect to $\mathcal{B}(S)$.

Now we define

$$\mathcal{E}_b(\mu) := \begin{cases} \int_S \left(\frac{1}{2}V_b^1(x,\mu) - V_b^2(x,\mu)\right) d\mu & \text{if } V_b^1(\cdot,\mu), V_b^2(\cdot,\mu) \in L^1(S,\mu), \\ 0 & \text{otherwise} \end{cases} \qquad \mu \in \mathcal{P}(S)$$

(for $\mu$ which is absolutely continuous with respect to $\mathbb{W}$, $\mathcal{E}(\mu)$ does not depend on the specific choice of $V_b^2(\cdot,\mu)$). Note that, if $R(\mu\|\mathbb{W}) < \infty$ and $\mathbb{E}[e^{\frac{1}{2}V_b^1(W,\mu)}] < \infty$, then $V_b^1(\cdot,\mu)$ and $V_b^2(\cdot,\mu)$ are in $L^1(\mu)$: indeed, by Lemma B.1,

$$\frac{1}{2}\int_S V_b^1(x,\mu)\, d\mu \leq R(\mu\|\mathbb{W}) + \log\int_S e^{\frac{1}{2}V_b^1(x,\mu)} d\mathbb{W} < \infty,$$

and, by Lemmas B.1 and D.2 (using $e^{|a|} \leq e^a + e^{-a}$),

$$\begin{aligned}\frac{1}{2}\int_S |V_b^2(x,\mu)| d\mu &\leq R(\mu\|\mathbb{W}) + \log\left(\int_S e^{\frac{1}{2}V_b^2(x,\mu)} d\mathbb{W} + \int_S e^{-\frac{1}{2}V_b^2(x,\mu)} d\mathbb{W}\right) \\ &\leq R(\mu\|\mathbb{W}) + \log\left(2\int_S e^{\frac{1}{2}V_b^1(x,\mu)} d\mathbb{W}\right) < \infty.\end{aligned}$$

In Lemma D.3 we show that $\mathcal{E}_b$ is Borel (at least on the set $\{\mu \mid R(\mu\|\mathbb{W}) < \infty\}$).

Coming to $\mathcal{E}_b^N$, we recall that $W^i$, $1 \le i \le N$, are independent $d$-dimensional Brownian motions on some filtered probability space $(\Omega, (\mathcal{F}_t)_t, \mathbb{P})$ (under the standard assumption) and $z^N_{\boldsymbol{W}}$ is the empirical measure associated with $W^i$. We assume on $b$ that

$$\int_S V_b^1(x, z^N_{\boldsymbol{W}})\, dz^N_{\boldsymbol{W}}(x) < \infty \qquad \mathbb{P}\text{-a.s..}$$

Under this assumption, we can define the stochastic integral

$$\int_0^T b_t(x_t, \mu_t) \cdot dx_t \tag{D.2.2}$$

on the space $(\bar\Omega, (\mathcal{H}_t)_t, \bar P)$. Here $\bar\Omega = S \times \mathcal{P}(S)$ and $\bar P$ is the law of $W^1 \otimes z^N_{\boldsymbol{W}}$, or equivalently of $W^i \otimes z^N_{\boldsymbol{W}}$ for any $1 \le i \le N$, under $\mathbb{W}$. Also $\mathcal{H}_t$ is the $\sigma$-algebra on $\bar\Omega$ generated by the $\bar P$-negligible sets and by $\pi_{[0,t]} \otimes (\pi_{[0,t]})_\#$, where $\pi_{[0,t]} : S \to S_t = C([0,t]; \mathbb{R}^d)$ is the projection on time $[0,t]$ and $(\pi_{[0,t]})_\# : \mathcal{P}(S) \to \mathcal{P}(S_t)$ is the corresponding image measure map; in particular, $\mathcal{H}_T$ is the completion under $\bar P$ of $\mathcal{B}(S) \otimes \mathcal{B}(\mathcal{P}(S))$. Hence we can define a map $V_b^{2,N} : S \times \mathcal{P}(S) \to \mathbb{R}$ which is a representative of the stochastic integral (D.2.2) and is measurable with respect to $\mathcal{B}(S) \otimes \mathcal{B}(\mathcal{P}(S))$. Note that, for every $1 \le i \le N$,

$$V_b^{2,N}(W^i, z^N_{\boldsymbol{W}}) = \int_0^T b_t(W^i_t, z^N_{\boldsymbol{W},t})\, dW^i_t \qquad \mathbb{P}\text{-a.s.} \tag{D.2.3}$$

and that, $\mathbb{P}$-a.s., $V_b^{2,N}(\cdot, z^N_{\boldsymbol{W}})$ is in $L^1(z^N_{\boldsymbol{W}})$.

Now we define

$$\mathcal{E}_b^N(\mu) := \begin{cases} \int_S \left(\frac{1}{2} V_b^1(x,\mu) - V_b^{2,N}(x,\mu)\right) d\mu & \text{if } V_b^1(\cdot,\mu), V^{2,N}(\cdot,\mu) \in L^1(S,\mu), \\ 0 & \text{otherwise} \end{cases} \qquad \mu \in \mathcal{P}(S).$$

By Theorem C.1, the function $\mathcal{E}_b^N$ is Borel.

**Lemma D.3.** *Assume that $\mathbb{E}[e^{\beta V_b^1(W,\mu)}] < \infty$ for every $\beta > 0$. Then the map*

$$\mathcal{P}(S) \ni \mu \mapsto \mathcal{E}_b(\mu) 1_{R(\mu\|\mathbb{W})<\infty} \quad \textit{is Borel.}$$

*Proof.* The map

$$\mu \mapsto \int_S \frac{1}{2} V_b^1(x,\mu)\, d\mu$$

is Borel by Theorem C.1, so it is enough to show Borel measurability of

$$F : \mu \mapsto \left(\int_S \frac{1}{2} V_b^2(x,\mu)\, d\mu\right) 1_{R(\mu\|\mathbb{W})<\infty} = \mathbb{E}\left[\int_0^T b_t(x_t,\mu_t) \cdot dW_t \, \frac{d\mu}{d\mathbb{W}}(W)\right] 1_{R(\mu\|\mathbb{W})<\infty}.$$

We start with the case of $b$ in $C_b([0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d))$ ($\mathcal{P}(\mathbb{R}^d)$ being endowed with the weak topology). We take a sequence $\Pi^n$ of partitions $0 = t_0 < t_1 < \ldots < t_n = T$ on $[0,T]$ with size tending to 0 in $n$. For each $n$, we call $b^n : [0,T] \times S \times \mathcal{P}(S) \to \mathbb{R}$ the Borel function

$$b^n(t,x,\mu) = \sum_i b(t_i, x_{t_i}, \mu_{t_i}) 1_{t \in [t_i, t_{i+1})}$$

and $I^n : S \times \mathcal{P}(S) \to \mathbb{R}$ the Borel function defined by

$$I^n(\gamma,\mu) = \sum_i b(t_i, \gamma_{t_i}, \mu_{t_i}) \cdot (\gamma_{t_{i+1}} - \gamma_{t_i}).$$

Again by Theorem C.1, the map

$$F_n : \mu \mapsto \mathbb{E}[I^n(W,\mu)\frac{d\mu}{d\mathbb{W}}(W)]1_{R(\mu\|\mathbb{W})<\infty}$$

is Borel. Now, for each $\mu$ with $R(\mu\|\mathbb{W}) < \infty$, we have by Lemmas B.1 and D.2, for every $\beta > 0$,

$$\begin{aligned}\beta|F(\mu) - F_n(\mu)| &\leq \log \mathbb{E}[e^{\beta|V_b^2(W,\mu)-I^n(W,\mu)|}] + R(\mu\|\mathbb{W})\\ &= \log \mathbb{E}[e^{\beta|\int_0^T (b_t(W_t,\mu_t)-b^n(t,W,\mu))\cdot dW_t|}] + R(\mu\|\mathbb{W})\\ &\leq \log\left(2\mathbb{E}[e^{2\beta^2\int_0^T |b_t(W_t,\mu_t)-b^n(t,W,\mu)|^2\cdot dt}]\right) + R(\mu\|\mathbb{W}).\end{aligned}$$

Since $b$ is continuous, $b_t(W_t,\mu_t) - b^n(t,W,\mu)$ tends to 0 for every $t$ in $[0,T)$ and every $W$ and $\mu$. Hence, for every fixed $\beta > 0$, by dominated convergence theorem and boundedness of $b$, $\mathbb{E}[e^{2\beta^2\int_0^T |b_t(W_t,\mu_t)-f^n(t,W,\mu)|^2\cdot dt}]$ tends to 1 and so

$$\limsup_n |F(\mu) - F_n(\mu)| \leq \frac{1}{\beta}(\log 2 + R(\mu\|\mathbb{W})).$$

By arbitrariness of $\beta$, $F$ is the pointwise limit of the Borel functions $F_n$, hence $F$ is Borel (for $b$ continuous and bounded).

The case of $b$ Borel bounded follows from the case of $b$ continuous bounded via a monotone class argument (cf. Theorem C.3): the stability assumption needed for the monotone class theorem can be verified as in the proof of convergence of $F_n$ to $F$. Finally, the case of general $b$ (satisfying $\mathbb{E}[e^{\beta V_b^1(W,\mu)}] < \infty$ for every $\beta$) follows approximating $b$ with bounded $b^n$ and proceeding as in the proof of convergence of $F_n$ to $F$. The proof is complete. □

D.3. **Proof of Lemma 4.1 and relative entropy representation.** In this subsection, we assume the setting at the beginning of Section 4.

*Proof of Lemma 4.1.* The SDE (4.1) is an SDE on $\mathbb{R}^{dN}$ for the vector $\boldsymbol{X}^N$, where the $i$-th component of the drift is $(t,\boldsymbol{x}) \mapsto b_t^N(x_i, z_{\boldsymbol{x}}^N)$. Note that, for this SDE, Novikov's condition is satisfies, indeed

$$\mathbb{E}\left[e^{\frac{1}{2}\sum_{i=1}^N \int_0^T |b_t^N(W_t^i, z_{\boldsymbol{W},t}^N)|^2 dt}\right] = \mathbb{E}\left[\exp\left(\frac{N}{2}\int_S\int_0^T |b_t^N(x_t, z_{\boldsymbol{W},t}^N)|^2 dt\, dz_{\boldsymbol{W}}^N(x)\right)\right] < \infty.$$

Girsanov's theorem gives then the existence of a weak solution $\boldsymbol{X}$, with law $\tilde{Q}_{b^N}^N$. The uniqueness in law condition (D.1.3) reads here

$$\sum_{i=1}^N \int_0^T |b_t^N(x_t^i, z_{\boldsymbol{x},t}^N)|^2 dt < \infty \quad \tilde{Q}_{b^N}^N\text{-a.s.},$$

which is equivalent to (4.3). The representation formula of the law $\tilde{Q}^N_{b^N}$ in Girsanov's theorem reads here (recall the definition of $V^1_b$ and $V^{2,N}_b$ in (4.2))

$$\begin{aligned}\frac{d\tilde{Q}^N_{b^N}}{d\tilde{P}^N}(\boldsymbol{W}) &= \exp\left(-\frac{1}{2}\sum_{i=1}^N\int_0^T |b^N_t(W^i_t, z^N_{\boldsymbol{W},t})|^2\,dt + \sum_{i=1}^N\int_0^T b^N_t(W^i_t, z^N_{\boldsymbol{W},t})\cdot dW^i_t\right)\\ &= \exp\left(-\frac{1}{2}N\int_S V^1_{b^N}(x, z^N_{\boldsymbol{W}})dz^N_{\boldsymbol{W}}(x) + N\int_S V^{2,N}_{b^N}(x, z^N_{\boldsymbol{W}})dz^N_{\boldsymbol{W}}(x)\right)\\ &= \exp\left(-N\mathcal{E}^N_{b^N}(z^N_{\boldsymbol{W}})\right),\end{aligned}$$

where we used (D.2.3) for the stochastic integral. The first formula in (4.4) is proved. The second formula (for the law $Q^N_{b^N}$ of the empirical measure) follows from the first one by a standard argument from measure theory. □

**Lemma D.4** (Relative entropy representation of $\mathcal{E}$)**.** *Assume that*

$$\mathbb{E}\left[e^{\frac{1}{2}\int_0^T |b_t(W_t,\mu_t)|^2 dt}\right] < \infty \quad \forall \mu \text{ with } R(\mu\|\mathbb{W}) < \infty. \tag{D.3.1}$$

*Then we have the following representation formula:*

$$R(\mu\|\mathbb{W}) + \mathcal{E}_b(\mu) = \begin{cases} R(\mu\|\mathbb{W}^\mu) & \text{if } R(\mu\|\mathbb{W}) < \infty \\ +\infty, & \text{otherwise,}\end{cases} \tag{D.3.2}$$

*where $\mathbb{W}^\mu$ is the law of the process $X^\mu$ satisfying the SDE*

$$dX^\mu_t = b_t(X^\mu_t, \mu_t)\,dt + dW_t$$

*with initial law $\rho_0$ (the law $\mathbb{W}^\mu$ exists and is uniquely determined by Girsanov's theorem D.1).*

*Moreover, when $b$ is in the class $\mathcal{F}$Lip, the restriction $R(\mu\|\mathbb{W}) < \infty$ in* (D.3.2) *may be dropped.*

*Proof.* When $R(\mu\|\mathbb{W}) = \infty$, the representation formula holds trivially (recall that $\mathcal{E}_b(\mu)$ is finite for every $\mu$). Now fix $\mu$ with $R(\mu\|\mathbb{W}) < \infty$. By the condition (D.3.1), we can apply Girsanov's theorem, which gives the formula

$$\begin{aligned}\frac{d\mathbb{W}^\mu}{d\mathbb{W}}(W) &= \exp\left(-\frac{1}{2}\int_0^T |b_t(W_t,\mu_t)|^2\,dt + \int_0^T b_t(W_t,\mu_t)\cdot dW_t\right)\\ &= \exp\left(-\frac{1}{2}V^1_b(W,\mu) + V^2_b(W,\mu)\right).\end{aligned} \tag{D.3.3}$$

In particular $\mathbb{W}$ and $\mathbb{W}^\mu$ are equivalent and so $\mu$ is absolutely continuous also with respect to $\mathbb{W}^\mu$. Hence we can compute the relative entropy

$$\begin{aligned}R(\mu\|\mathbb{W}^\mu) &= \int \log\frac{d\mu}{d\mathbb{W}^{\mu,b}}d\mu = \int \log\frac{d\mu}{d\mathbb{W}}d\mu - \int \log\frac{d\mathbb{W}^\mu}{d\mathbb{W}}d\mu\\ &= R(\mu\|\mathbb{W}) - \int\left(-\frac{1}{2}V^1_b(W,\mu) + V^2_b(W,\mu)\right)d\mu(W)\\ &= R(\mu\|\mathbb{W}) + \mathcal{E}_b(\mu),\end{aligned}$$

which is the desired representation formula.

For $b$ in $\mathcal{F}$Lip we have to prove that for every $\mu$,

$$R(\mu\|\mathbb{W}^{\mu,b}) < \infty \iff R(\mu\|\mathbb{W}) < \infty. \tag{D.3.4}$$

Note that $\mathbb{W}^{\mu,b}$ is well-defined for every $\mu$ when $b$ is in $\mathcal{F}\mathrm{Lip}$. Now, fix any $\mu$, and note that in particular $b$ is bounded and hence Girsanov's formula (D.3.3) still holds. Moreover, for every $\beta > 0$ and $i = 1, 2$, by boundedness of $b$ and Lemma D.2 we have,

$$\mathbb{E}\left[e^{\beta V_b^i(W,\mu)}\right] < \infty.$$

Applying Corollary B.3 to the measures $\mathbb{W}$ and $\mathbb{W}^{\mu,b}$ we easily deduce (D.3.4). □

D.4. $L_t^q(L_x^p)$**-estimates.** Khasminskii's Lemma is classical, see for example [Has59], [Szn98, Chapter 1, Lemma 2.1], [FF11, Lemma 13].

**Lemma D.5** (Khasminskii's Lemma)**.** *Let $W$ be a $d$-dimensional Brownian motion starting from 0, let $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ be a non-negative Borel function and assume that*

$$\alpha_f := \sup_{x\in\mathbb{R}^d} \mathbb{E} \int_0^T f(t, x + W_t)\, dt < 1.$$

*Then it holds*

$$\sup_{x\in\mathbb{R}^d} \mathbb{E}\left[\exp\left[\int_0^T f(t, x + W_t)\, dt\right]\right] \le \frac{1}{1-\alpha_f}.$$

**Lemma D.6** ($L_t^q(L_x^p)$ estimates)**.** *Let $W$ be a $d$-dimensional Brownian motion starting from 0. Take $1 \le p, q \le \infty$ satisfying*

$$\frac{d}{p} + \frac{2}{q} < 2. \tag{D.4.1}$$

*Then there exists a constant $C$ (depending on $p, q, d$ and $T$) such that, for every $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ non-negative Borel function,*

$$\sup_{x\in\mathbb{R}^d} \int_0^T \mathbb{E}[f(t, x + W_t)]\, dt \le C\|f\|_{L_t^q(L_x^p)}. \tag{D.4.2}$$

This bound is classical (see e.g. [FF11, Lemma 11]), with an elementary proof that we recall here.

*Proof.* We use Hölder's inequality applied at $t$ and $x$ fixed for the convolution with the Gaussian density $p_t$:

$$\mathbb{E}[f(t, x + W_t)] = f_t \star p_t(x) \le \|f_t\|_{L_x^p} \|p_t\|_{L_x^{p'}}$$

We recall that $\|p_t\|_{L_x^{p'}} \le ct^{-d/2p}$ for some constant $c$ depending on $p$ and $d$ (as one can see via the change of variable $x' = t^{-1/2}x$). Therefore, for every $x$, we get by Hölder's inequality,

$$\int_0^T f(t, x + W_t)\, dt \le c \int_0^T \|f_t\|_{L_x^p} t^{-d/2p}\, dt \le c\|f\|_{L_t^q(L_x^p)} \left(\int_0^T t^{-dq'/2p}\, dt\right)^{1/q'} \quad \text{for } q < \infty$$

(note that (D.4.1) implies $q > 1$, so $q' < \infty$). As for $q = \infty$, we estimate similarly,

$$\int_0^T f(t, x + W_t)\, dt \le c\|f_t\|_{L_t^\infty(L_x^p)} \int_0^T t^{-d/2p}\, dt.$$

Now the assumption on $p$ and $q$ is equivalent to $dq'/2p < 1$ for $q < \infty$ and to $d < 2p$ for $q = \infty$. Hence the time integral of $t^{-dq'/2p}$ is finite. Hence the bound (D.4.2) holds with $C = c\left(\int_0^T t^{-dq'/2p} dt\right)^{1/q'}$. The proof is complete. □

The previous bound can be easily generalized to the case of $k$ independent Brownian motions, as in the following:

**Lemma D.7.** *Let $W^1, \dots W^k$ be $k$ independent $d$-dimensional Brownian motions starting from 0. Take $1 \le p_1, \dots p_k, q \le \infty$ satisfying*

$$\frac{d}{p_1} + \dots \frac{d}{p_k} + \frac{2}{q} < 2.$$

*Then there exists a constant $C$ (depending on $p_1, \dots p_k, q, d$ and $T$) such that, for every $f : [0,T] \times \mathbb{R}^{kd} \to \mathbb{R}$ non-negative Borel function,*

$$\sup_{x_1,\dots x_k \in \mathbb{R}^d} \int_0^T \mathbb{E}[f(t, x_1 + W_t^1, \dots x_k + W_t^k)]dt \le C\|f\|_{L_t^q(L_{x_1}^{p_1}(\dots(L_{x_k}^{p_k})\dots))}. \tag{D.4.3}$$

*More generally, one can replace the above right-hand side by $\|f\|_{L_t^q(L_{x_{\sigma(1)}}^{p_1}(\dots(L_{x_{\sigma(k)}}^{p_k})\dots))}$ for any permutation $\sigma$ of $\{1, \dots k\}$.*

*Proof.* The proof is similar to the previous one. We write

$$\mathbb{E}[f(t, x_1 + W_t^1, \dots x_k + W_t^k)] = f_t \star p_t^{\otimes k}(x_1 \dots x_k)$$

and use Hölder inequality in the $x_k$ variable, to get

$$\mathbb{E}[f(t, x_1 + W_t^1, \dots x_k + W_t^k)] \le c\|f_t(x_1, \dots x_{k-1}, \cdot)\|_{L_{x_k}^{p_k}} t^{-d/2p_k} p_t^{\otimes k}(x_1, \dots x_{k-1}).$$

Then we proceed similarly with the other variables and get

$$\mathbb{E}[f(t, x_1 + W_t^1, \dots x_k + W_t^k)] \le c\|f_t\|_{L_t^q(L_{x_1}^{p_1}(\dots(L_{x_k}^{p_k})\dots))} t^{-d/2p_1} \cdot \dots \cdot t^{-d/2p_k}.$$

We then conclude on (D.4.3) as in the previous proof, taking

$$C = C_T = c\left(\int_0^T t^{-dq'(1/2p_1 + \dots 1/2p_k)} dt\right)^{1/q'} = c_1 T^{1-1/q-d/(2p_1)-\dots-d/(2p_k)}, \tag{D.4.4}$$

for some constant $c_1 > 0$ independent of $T$. The bound for a general permutation $\sigma$ follows from (D.4.3) applied to $f(x_{\sigma^{-1}(1)}, \dots x_{\sigma^{-1}(k)})$. □

Finally, we put together the previous bounds to obtain an exponential estimate for $L^q(L^p)$ functions (see [FF11, Corollary 14] for a similar statement).

**Lemma D.8.** *Let $W^1, \dots W^k$ be $k$ independent $d$-dimensional Brownian motions starting from 0. Take $1 \le p_1, \dots p_k, q \le \infty$ satisfying*

$$\frac{d}{p_1} + \dots \frac{d}{p_k} + \frac{2}{q} < 2. \tag{D.4.5}$$

*Then there exists a constant $c > 0$ (depending on $p_1, \dots p_k, q, T$) such that, for every $f : [0,T] \times \mathbb{R}^{kd} \to \mathbb{R}$ non-negative Borel function with $f \in L_t^q(L_{x_1}^{p_1} \dots (L_{x_k}^{p_k}) \dots)$,*

$$\sup_{x_1,\dots x_k \in \mathbb{R}^d} \mathbb{E}\left[\exp\left[\int_0^T f(t, x_1 + W_t^1, \dots x_k + W_t^k)\, dt\right]\right] \le \exp\left[c\left(1 + \|f\|_{L_t^q(L_{x_1}^{p_1} \dots (L_{x_k}^{p_k})\dots)}^{1/(1-\alpha)}\right)\right],$$

*with $\alpha = 1 - 1/q - d/(2p_1) - \dots - d/(2p_k)$.*

*Proof.* We take

$$h = \Big(2c_1\|f\|_{L^q_t(L^{p_1}_{x_1}\dots(L^{p_k}_{x_k})\dots)}\Big)^{-1/(1-\alpha)} \wedge T,$$

and let $t_j = hj \wedge T$, and $m$ the first positive integer with $t_m = T$, in particular

$$m = \left\lceil \frac{T}{h} \right\rceil \le T\Big(2c_1\|f\|_{L^q_t(L^{p_1}_{x_1}\dots(L^{p_k}_{x_k})\dots)}\Big)^{1/(1-\alpha)} + 1$$

With this choice of $h$, we have

$$C_h \sup_{j=0,\dots m-1} \left(\int_{t_j}^{t_{j+1}} \|f_t\|^q_{L^{p_1}_{x_1}(\dots(L^{p_k}_{x_k})\dots)} dt\right)^{1/q} \le C_h\|f\|_{L^q_t(L^{p_1}_{x_1}\dots(L^{p_k}_{x_k})\dots)} \le \frac{1}{2},$$

$C_h$ being the constant in (D.4.4). As a consequence of Lemma D.7, we have

$$\begin{aligned}\sup_{j=0,\dots,m-1} \sup_{x_1,\dots x_k\in\mathbb{R}^d} \int_{t_j}^{t_{j+1}} \mathbb{E}[f(t, x_1 + W^1_t - W^1_{t_j}, \dots x_k + W^k_t - W^k_{t_j})]\, dt \\ \le C_h \sup_{j=0,\dots,m-1} \left(\int_{t_j}^{t_{j+1}} \|f_t\|^q_{L^{p_1}_{x_1}(\dots(L^{p_k}_{x_k})\dots)} dt\right)^{1/q} \le \frac{1}{2}.\end{aligned}$$

Hence, we can apply Lemma 4.7 and get

(D.4.6)

$$\sup_{j=0,\dots,m-1} \sup_{x_1,\dots x_k\in\mathbb{R}^d} \mathbb{E}\left[\exp\left[\int_{t_j}^{t_{j+1}} f(t, x_1 + W^1_t - W^1_{t_j}, \dots x_k + W^k_t - W^k_{t_j})\, dt\right]\right] \le 2.$$

Now we come back to the bound on the whole time interval $[0, T]$. We split the time integral over $[0, T]$ into the integrals over $[t_j, t_{j+1}]$ and use conditional expectation with respect to $\mathcal{F}_{t_{m-1}}$: we have, for every $x_1, \dots, x_k \in \mathbb{R}^d$,

$$\begin{aligned}&\mathbb{E}\left[\exp\left[\int_0^T f(t, x_1 + W^1_t, \dots x_k + W^k_t)\, dt\right]\right] \\ &\quad = \mathbb{E}\left[\prod_{j=0}^{m-1} \exp\left[\int_{t_j}^{t_{j+1}} f(t, x_1 + W^1_t, \dots x_k + W^k_t)\, dt\right]\right] \\ &\quad = \mathbb{E}\left[\prod_{j=0}^{m-2} \exp\left[\int_{t_j}^{t_{j+1}} f(t, x_1 + W^1_t, \dots x_k + W^k_t)\, dt\right] \cdot\right. \\ &\qquad\qquad \left.\cdot\, \mathbb{E}\left[\exp\left[\int_{t_{m-1}}^{t_m} f(t, x_1 + W^1_t, \dots x_k + W^k_t)\, dt\right] \,\middle|\, \mathcal{F}_{t_{m-1}}\right]\right]\end{aligned}$$

(all exponentials are $\ge 1$ and so the above products make sense and we can use the rule $\mathbb{E}[XY] = \mathbb{E}[X\mathbb{E}[Y \mid \mathcal{F}_s]]$ for $X$ $\mathcal{F}_s$-measurable). Now we apply the Markov property and the

bound (D.4.6) and get

$$\begin{aligned}
&\mathbb{E}\left[\exp\left[\int_0^T f(t, x_1 + W_t^1, \dots x_k + W_t^k)\, dt\right]\right] \\
&= \mathbb{E}\left[\prod_{j=0}^{m-2} \exp\left[\int_{t_j}^{t_{j+1}} f(t, x_1 + W_t^1, \dots x_k + W_t^k)\, dt\right] \cdot\right. \\
&\quad \left.\cdot \mathbb{E}\left[\exp\left[\int_{t_{m-1}}^{t_m} f(t, y_1 + W_t^1 - W_{t_{m-1}}^1, \dots y_k + W_t^k - W_{t_{m-1}}^k)\, dt\right]\right]\Bigg|_{y_1 = x_1 + W_{t_{m-1}}^1, \dots, y_k = x_k + W_{t_{m-1}}^k}\right] \\
&\leq 2\mathbb{E}\left[\prod_{j=0}^{m-2} \exp\left[\int_{t_j}^{t_{j+1}} f(t, x_1 + W_t^1, \dots x_k + W_t^k)\, dt\right]\right].
\end{aligned}$$

Iterating this argument on $j$, we get finally, for every $x_1, \dots x_k \in \mathbb{R}^d$,

$$\mathbb{E}\left[\exp\left[\int_0^T f(t, x_1 + W_t^1, \dots x_k + W_t^k) dt\right]\right] \leq 2^m \leq 2^{T(2c_1 \|f\|_{L_t^q(L_{x_1}^{p_1} \dots (L_{x_k}^{p_k}) \dots)})^{1/(1-\alpha)} + 1},$$

which concludes the proof. □

## References


[BDF12] A. Budhiraja, P. Dupuis, and M. Fischer. Large deviation properties of weakly interacting processes via weak convergence methods. *Ann. Probab.*, 40(1):74–102, 01 2012.

[Ber18] R. J. Berman. On large deviations for Gibbs measures, mean energy and gamma-convergence. *Constr. Approx.*, 48(1):3–30, 2018.

[BFGM19] L. Beck, F. Flandoli, M. Gubinelli, and M. Maurelli. Stochastic ODEs and stochastic linear PDEs with critical drift: regularity, duality and uniqueness. *Electron. J. Probab.*, 24:72 pp., 2019.

[BG99] T. Bodineau and A. Guionnet. About the stationary states of vortex systems. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 35(2):205–237, apr 1999.

[BJW19] D. Bresch, P.-E. Jabin, and Z. Wang. On mean-field limits and quantitative estimates with a large class of singular kernels: Application to the Patlak–Keller–Segel model. *Comptes Rendus Mathematique*, pages 1–13, 2019.

[BO19] R. J. Berman and M. Önnheim. Propagation of chaos for a class of first order models with singular mean field interactions. *SIAM J. Math. Anal.*, 51(1):159–196, 2019.

[Bog07] V. I. Bogachev. *Measure Theory*, volume II. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[CDFM20] M. Coghi, J.-D. Deuschel, P. K. Friz, and M. Maurelli. Pathwise McKean-Vlasov theory with additive noise. *Ann. Appl. Probab.*, 30(5):2355–2392, 2020.

[CdR20] P. E. Chaudru de Raynal. Strong well posedness of McKean–Vlasov stochastic differential equations with Hölder drift. *Stochastic Process. Appl.*, 130(1):79–107, 2020.

[CGZ14] D. Chafaï, N. Gozlan, and P. A. Zitt. First-order global asymptotics for confined particles with singular pair repulsion. *Annals of Applied Probability*, 24(6):2371–2413, 2014.

[DE97] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 1997.

[Del19] F. Delarue. Restoring uniqueness to mean-field games by randomizing the equilibria. *Stoch. Partial Differ. Equ. Anal. Comput.*, 7(4):598–678, 2019.

[DFMS17] J. D. Deuschel, P. K. Friz, M. Maurelli, and M. Slowik. The enhanced Sanov theorem and propagation of chaos. *Stochastic Processes and their Applications*, pages 1–42, 2017.

[DG87] D. A. Dawson and J. Gärtner. Large deviations from the McKean–Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, apr 1987.

[DLR20] P. Dupuis, V. Laschos, and K. Ramanan. Large deviations for configurations generated by gibbs distributions with energy functionals consisting of singular interaction and weakly confining potentials. *Electron. J. Probab.*, 25:41 pp., 2020.

[DMZ03] P. Del Moral and T. Zajic. A note on the Laplace–Varadhan integral lemma. *Bernoulli*, 9(1):49–65, 02 2003.

[dRST19] G. dos Reis, W. Salkeld, and J. Tugaut. Freidlin–Wentzell LDP in path space for McKean–Vlasov equations and the functional iterated logarithm law. *Ann. Appl. Probab.*, 29(3):1487–1540, 2019.

[DZ10] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[EG15] L. C. Evans and R. Gariepy. *Measure Theory and Fine Properties, revised edition*. New York: Chapman and Hall/CRC, 2015.

[Eic04] P. Eichelsbacher. Refined large deviations for von Mises statistics. *Teor. Veroyatn. Primen.*, 49(1):197–204, 2004.

[EK05] S. Ethier and T. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA, mar 2005.

[ES98] P. Eichelsbacher and U. Schmock. Exponential approximations in completely regular topological spaces and extensions of Sanov's theorem. *Stochastic Processes and their Applications*, 77(2):233–251, sep 1998.

[ES02] P. Eichelsbacher and U. Schmock. Large deviations of U-empirical measures in strong topologies and applications. *Annales de l'institut Henri Poincare (B) Probability and Statistics*, 38(5):779–797, 2002.

[EZ03] P. Eichelsbacher and T. Zajic. Moderate deviations for mean-field Gibbs measures. *Bernoulli*, 9(1):67–95, 2003.

[Fed09] E. Fedrizzi. Uniqueness and Flow Theorems for solutions of SDEs with low regularity of the drift. Master's thesis, University of Pisa, 2009.

[FF11] E. Fedrizzi and F. Flandoli. Pathwise uniqueness and continuous dependence of SDEs with non-regular drift. *Stochastics*, 83(3):241–257, 2011.

[FHM14] N. Fournier, M. Hauray, and S. Mischler. Propagation of chaos for the 2D viscous vortex model. *Journal of the European Mathematical Society*, 16(7):1423–1466, 2014.

[Fis14] M. Fischer. On the form of the large deviation rate function for the empirical measures of weakly interacting systems. *Bernoulli*, 20(4):1765–1801, 2014.

[FJ17] N. Fournier and B. Jourdain. Stochastic particle approximation of the Keller–Segel equation and two-dimensional generalization of Bessel processes. *Ann. Appl. Probab.*, 27(5):2807–2861, 2017.

[Fon04] J. Fontbona. Uniqueness for a weak nonlinear evolution equation and large deviations for diffusing particles with electrostatic repulsion. *Stochastic Processes and their Applications*, 112(1):119–144, 2004.

[FPZ19] F. Flandoli, E. Priola, and G. Zanco. A mean-field model with discontinuous coefficients for neurons with spatial interaction. *Discrete Contin. Dyn. Syst.*, 39(6):3037–3067, 2019.

[GNP21] N. Gavish, P. Nyquist, and M. Peletier. Large deviations and gradient flows for the Brownian one-dimensional hard-rod system. *Potential Anal.*, 2021.

[GQ15] D. Godinho and C. Quiñinao. Propagation of chaos for a subcritical Keller–Segel model. *Annales de l'institut Henri Poincare (B) Probability and Statistics*, 51(3):965–992, 2015.

[GZ19] D. García-Zelada. A large deviation principle for empirical measures on Polish spaces: Application to singular Gibbs measures on manifolds. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(3):1377–1401, 2019.

[Has59] R. Z. Has′minskiĭ. On positive solutions of the equation $\mathcal{U} + Vu = 0$. *Theor. Probability Appl.*, 4:309–318, 1959.

[HH16] Y. Hariya and K. Hasegawa. On divergence of expectations of the Feynman–Kac type with singular potentials. *J. Math. Soc. Japan*, 68(3):1271–1296, 2016.

[HLSS18] D. P. Hardin, T. Leblé, E. B. Saff, and S. Serfaty. Large deviation principles for hypersingular Riesz gases. *Constr. Approx.*, 48(1):61–100, 2018.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[Hoe17] J. Hoeksema. Quasi-continuity and large-deviation principles for singularly interacting particle systems. Master's thesis, Technical University of Eindhoven, 2017.

[HRZ22] Z. Hao, M. Röckner, and X. Zhang. Strong convergence of propagation of chaos for McKean-Vlasov SDEs with singular interactions. *arXiv:2204.07952*, 2022. preprint.

[HvS21] W. R. P. Hammersley, D. Šiška, and L. u. Szpruch. McKean-Vlasov SDEs under measure dependent Lyapunov conditions. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(2):1032–1057, 2021.

[Jab19] J.-F. Jabir. Rate of propagation of chaos for diffusive stochastic particle systems via Girsanov transformation. *arXiv:1907.09096*, 2019. preprint.

[Jor06] P. E. T. Jorgensen. *Infinite Dimensional Analysis*, volume 2-2. Springer-Verlag, Berlin/Heidelberg, 2006.

[JTT18] J.-F. Jabir, D. Talay, and M. Tomašević. Mean-field limit of a particle approximation of the one-dimensional parabolic-parabolic Keller-Segel model without smoothing. *Electron. Commun. Probab.*, 23:Paper No. 84, 14, 2018.

[JW18] P. E. Jabin and Z. Wang. Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. *Inventiones Mathematicae*, 214(1):523–591, 2018.

[KR05] N. V. Krylov and M. Röckner. Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Related Fields*, 131(2):154–196, 2005.

[Lac18] D. Lacker. On a strong form of propagation of chaos for McKean–Vlasov equations. *Electron. Commun. Probab.*, 23:Paper No. 45, 11, 2018.

[Léo87] C. Léonard. Large deviations and law of large numbers for a mean field type interacting particle system. *Stochastic processes and their applications*, 25:215–235, 1987.

[LS01] R. S. Liptser and A. N. Shiryaev. *Statistics of random processes. I*, volume 5 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, expanded edition, 2001. General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.

[LW20] W. Liu and L. Wu. Large deviations for empirical measures of mean-field Gibbs measures. *Stochastic Process. Appl.*, 130(2):503–520, 2020.

[MS19] S. Mehri and W. Stannat. Weak solutions to Vlasov–McKean equations under Lyapunov-type conditions. *Stoch. Dyn.*, 19(6):1950042, 23, 2019.

[MV20] Y. S. Mishura and A. Y. Veretennikov. Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations. *Theor. Probability and Math. Statist.*, 103:59–101, 2020.

[Orr20] C. Orrieri. Large deviations for interacting particle systems: joint mean-field and small-noise limit. *Electron. J. Probab.*, 25:Paper No. 111, 44, 2020.

[PdH96] P. D. Pra and F. den Hollander. McKean–Vlasov limit for interacting random processes in random media. *Journal of Statistical Physics*, 84(3-4):735–772, aug 1996.

[PS19] M. A. Peletier and M. C. Schlottke. Large-deviation principles of switching Markov processes via Hamilton–Jacobi equations. *arXiv:1901.08478*, 2019. Preprint.

[Rey18] J. Reygner. Equilibrium large deviations for mean-field systems with translation invariance. *Ann. Appl. Probab.*, 28(5):2922–2965, 2018.

[RZ21] M. Röckner and X. Zhang. Well-posedness of distribution dependent SDEs with singular drifts. *Bernoulli*, 27(2):1131–1158, 2021.

[Szn91] A.-S. Sznitman. Topics in propagation of chaos. In P.-L. Hennequin, editor, *Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989*, pages 165–251, Berlin, Heidelberg, 1991. Springer Berlin Heidelberg.

[Szn98] A.-S. Sznitman. *Brownian motion, obstacles and random media*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998.

[Tan84] H. Tanaka. Limit theorems for certain diffusion processes with interaction. In K. Itô, editor, *Stochastic Analysis*, volume 32 of *North-Holland Mathematical Library*, pages 469–488. Elsevier, 1984.

[Tom20] M. Tomasevic. Propagation of chaos for stochastic particle systems with singular mean-field interaction of $L^q$–$L^p$ type. *hal-03086253*, 2020. preprint.

[Var84] S. R. S. Varadhan. *Large Deviations and Applications*. Society for Industrial and Applied Mathematics, 1984.