Bernoulli **29**(1), 2023, 725–756
https://doi.org/10.3150/22-BEJ1476

A trajectorial approach to relative entropy dissipation of McKean–Vlasov diffusions: Gradient flows and HWBI inequalities

BERTRAM TSCHIDERER[1,a] and LANE CHUN YEUNG[2,b]

[1]*Faculty of Mathematics, University of Vienna, Austria,* [a]*bertram.tschiderer@univie.ac.at*
[2]*Department of Industrial Engineering and Operations Research, Columbia University, New York, USA,* [b]*l.yeung@columbia.edu*

We formulate a trajectorial version of the relative entropy dissipation identity for McKean–Vlasov diffusions, extending recent results which apply to non-interacting diffusions. Our stochastic analysis approach is based on time-reversal of diffusions and Lions' differential calculus over Wasserstein space. It allows us to compute explicitly the rate of relative entropy dissipation along every trajectory of the underlying diffusion via the semimartingale decomposition of the corresponding relative entropy process. As a first application, we obtain a new interpretation of the gradient flow structure for the granular media equation, generalizing a formulation developed recently for the linear Fokker–Planck equation. Secondly, we show how the trajectorial approach leads to a new derivation of the HWBI inequality, which relates relative entropy (H), Wasserstein distance (W), barycenter (B) and Fisher information (I).

Keywords: Relative entropy dissipation; gradient flow; McKean–Vlasov diffusion; granular media equation; HWBI inequality

1. Introduction

We are interested in the relative entropy dissipation of McKean–Vlasov stochastic differential equations of the form

$$\mathrm{d}X_t = -\big(\nabla V(X_t) + \nabla(W * \mathrm{P}_t)(X_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t\,, \qquad 0 \le t \le T, \tag{1.1}$$

where X_0 has some given initial distribution P_0 on $\mathbb{R}^n$. Here, the functions $V, W\colon \mathbb{R}^n \to [0,\infty)$ play the roles of confinement and interaction potentials and are assumed to be suitably regular, $\mathrm{P}_t := \mathrm{Law}(X_t)$ denotes the distribution of the random vector X_t, the symbol $*$ stands for the standard convolution operator, and $(B_t)_{0\le t\le T}$ is a standard n-dimensional Brownian motion. In particular, this SDE is non-local (or non-linear) in the sense that the drift term depends on the distribution of the state variable. Non-local equations of this form arise in the modeling of weakly interacting diffusion equations, after the seminal work of McKean [47].

Since the work of Carrillo–McCann–Villani [13,14], relative entropy dissipation has been known to be an effective method for studying convergence rates to equilibrium and propagation of chaos of McKean–Vlasov equations. Some notable examples include the works [8,11,12,16,45,54]. In a broader context, [34,48] recently applied entropy methods to the mean-field theory of neural networks.

We denote by $\mathscr{P}_{\mathrm{ac}}(\mathbb{R}^n)$ the set of absolutely continuous probability measures on $\mathbb{R}^n$, which we will often identify with their corresponding probability density functions with respect to Lebesgue measure. The *free energy functional*

$$\mathscr{P}_{\mathrm{ac}}(\mathbb{R}^n) \ni p \longmapsto \mathscr{F}(p) := \mathcal{U}(p) + \mathcal{V}(p) + \mathcal{W}(p) \tag{1.2}$$

is defined as the sum of the energy functionals

$$\mathcal{U}(p) \coloneqq \int_{\mathbb{R}^n} p(x)\log p(x)\,\mathrm{d}x, \quad \mathcal{V}(p) \coloneqq \int_{\mathbb{R}^n} V(x)\,p(x)\,\mathrm{d}x, \quad \mathcal{W}(p) \coloneqq \tfrac{1}{2}\int_{\mathbb{R}^n} (W * p)(x)\,p(x)\,\mathrm{d}x \tag{1.3}$$

corresponding to internal ($\mathcal{U}$), potential ($\mathcal{V}$) and interaction ($\mathcal{W}$) energy, respectively. Defining the *relative entropy dissipation functional*

$$\mathscr{P}_{\mathrm{ac}}(\mathbb{R}^n) \ni p \longmapsto \mathscr{D}(p) \coloneqq \int_{\mathbb{R}^n} |\nabla \log p(x) + \nabla V(x) + \nabla(W * p)(x)|^2\, p(x)\,\mathrm{d}x, \tag{1.4}$$

the well-known relative entropy dissipation identity takes the form

$$\mathscr{F}(p_t) - \mathscr{F}(p_{t_0}) = -\int_{t_0}^{t} \mathscr{D}(p_u)\,\mathrm{d}u. \tag{1.5}$$

This identity is of a deterministic nature: it only depends on the curve of probability density functions $(p_t)_{0\le t\le T}$, but not on the trajectories of the underlying process $(X_t)_{0\le t\le T}$ itself. It is then natural to ask whether there is a process-level analogue of the relative entropy dissipation identity (1.5), depending directly on the trajectories of the McKean–Vlasov process $(X_t)_{0\le t\le T}$. The main contribution of this paper is to give an affirmative answer to this question, by formulating a trajectorial version of the relative entropy dissipation identity via a stochastic analysis approach.

Before going into details, let us briefly describe the main ideas. We draw inspiration from prior literature [29,40] based on a simpler (linear) setting without interaction, i.e., $W \equiv 0$. In this case, the McKean–Vlasov SDE (1.1) reduces to a Langevin–Smoluchowski diffusion equation of the form

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t\,, \qquad 0 \le t \le T. \tag{1.6}$$

In particular, the drift term does not depend on the distribution of X_t. Moreover, there is an explicit stationary distribution (also known as the Gibbs distribution [30,35,52]) with density proportional to $\mathbb{R}^n \ni x \mapsto q(x) \coloneqq \mathrm{e}^{-V(x)}$. Defining the likelihood ratio function (or Radon–Nikodym derivative) $\ell_t(x) \coloneqq p_t(x)/q(x)$, the free energy at time t can be expressed as $\mathscr{F}(p_t) = \mathbb{E}[\log \ell_t(X_t)]$, and the resulting stochastic process

$$\log \ell_t(X_t) = \log p_t(X_t) + V(X_t), \qquad 0 \le t \le T \tag{1.7}$$

is called *free energy* or *relative entropy process*. As shown in [29,40], the time-reversal

$$\big(\log \ell_{T-s}(X_{T-s})\big)_{0\le s\le T}$$

of this process is a submartingale, and Itô calculus can be used to obtain its Doob–Meyer decomposition

$$\log \ell_{T-s}(X_{T-s}) - \log \ell_T(X_T) = M_{T-s} + F_{T-s}. \tag{1.8}$$

Here, $(M_{T-s})_{0\le s\le T}$ is a martingale and $(F_{T-s})_{0\le s\le T}$ is an increasing process of finite first variation, both with explicit expressions. This decomposition describes exactly the rate of relative entropy dissipation along every trajectory of the Langevin–Smoluchowski diffusion. Therefore, it can be viewed as a trajectorial analogue of the (deterministic) relative entropy dissipation identity (1.5).

Let us now return to our McKean–Vlasov setting. In order to take into account the interaction potential W, it is natural to consider a generalized relative entropy process of the form

$$\log p_t(X_t) + V(X_t) + \tfrac{1}{2}(W * \mathrm{P}_t)(X_t), \qquad 0 \le t \le T. \tag{1.9}$$

The task is now to compute the semimartingale decomposition of this process. We will provide a detailed analysis of this extension, which is subtler than might appear at first sight. The main difficulty is that, even when it exists, the stationary distribution of the McKean–Vlasov diffusion does not have a closed-form expression and is not even unique in general; see the works [6,20,21,33,54–57]. This prevents us from defining the likelihood ratio function in a straightforward manner as in the setting of Langevin–Smoluchowski diffusions, where one can rely on the invariant Gibbs distribution. An appropriate definition of the generalized likelihood ratio function turns out to be that (1.9) should be viewed as a function of the form $\log \ell_t(X_t, \mathrm{P}_t)$, depending explicitly on the distribution P_t of X_t itself, in addition to the state X_t. This form of generalized likelihood ratio function allows us to take the *L-derivative* with respect to the probability distribution P_t. The notion of L-differentiation for functions of probability measures was introduced by Lions [43]. We refer to the monograph [10, Chapter 5] for a detailed discussion of differential calculus and stochastic analysis over spaces of probability measures. In particular, we will use a generalized form of Itô's formula for functions of curves of measures, to derive the dynamics of the time-reversal of the relative entropy process (1.9), in terms of the semimartingale decomposition

$$\log \ell_{T-s}(X_{T-s}, \mathrm{P}_{T-s}) - \log \ell_T(X_T, \mathrm{P}_T) = M_{T-s} + F_{T-s}, \qquad 0 \le s \le T, \tag{1.10}$$

where $(M_{T-s})_{0\le s\le T}$ is a martingale and $(F_{T-s})_{0\le s\le T}$ is a process of finite first variation, both of which will be explicitly computed. Similar to the case of Langevin–Smoluchowski dynamics, this decomposition can be viewed as the trajectorial rate of relative entropy dissipation. The classical (deterministic) identity (1.5) can then be recovered by taking expectations.

1.1. Gradient flow structure of the granular media equation

As a first application of our trajectorial approach we obtain a new interpretation of the gradient flow structure of the *granular media equation*

$$\partial_t p_t(x) = \operatorname{div}\Big(\nabla p_t(x) + p_t(x)\nabla V(x) + p_t(x)\nabla(W * p_t)(x)\Big), \qquad (t,x) \in (0,T)\times\mathbb{R}^n, \tag{1.11}$$

which describes the evolution of the curve of probability density functions $(p_t)_{0\le t\le T}$ corresponding to the McKean–Vlasov diffusion $(X_t)_{0\le t\le T}$ of (1.1). When $n = 1$, this PDE appears in the modeling of the time evolution of granular media [7,17,59]; in that context, the granular medium is modeled as system of particles performing inelastic collisions, and $p_t(x)$ is regarded as the velocity of a representative particle in the system at time t and position x, while V and W represent the friction and the inelastic collision forces, respectively. Note that in the interaction-free case $W \equiv 0$, the equation (1.11) reduces to a linear *Fokker–Planck equation*. As is well known from [13,14], this curve of probability densities can be characterized as a *gradient flow* in $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$, the space of absolutely continuous probability measures with finite second moments. Roughly speaking, this is an optimality property stating that the curve $(p_t)_{0\le t\le T}$ evolves in the direction of steepest possible descent for the free energy functional (1.2) with respect to the quadratic Wasserstein distance

$$W_2(\mu,\nu) = \Big(\inf_{Y\sim\mu, Z\sim\nu} \mathbb{E}|Y-Z|^2\Big)^{1/2}, \qquad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^n). \tag{1.12}$$

The Wasserstein gradient flow structure of the linear Fokker–Planck equation was first discovered by Jordan, Kinderlehrer and Otto in the seminal work [36]. In the paper [50], Otto and Villani developed a formal Riemannian structure on the space of probability measures with finite second moments, leading to heuristic proofs of gradient flow properties as in [49], where the porous medium equation was

studied. This pioneering approach is often referred to as "Otto calculus". Later, a rigorous framework based on minimizing movement schemes and curves of maximal slope was introduced in [3]. Recently, a trajectorial approach to the gradient flow properties of Langevin–Smoluchowski diffusions [40] and Markov chains [38] was established. We will follow this approach and adapt it to our McKean–Vlasov setting. For gradient flows of McKean–Vlasov equations on discrete spaces we refer to [24].

Returning to the setting of this paper, our main result leads to a new formulation of the gradient flow property of the granular media equation. To show this steepest descent property, the main idea is to consider a *perturbed* McKean–Vlasov diffusion of the form

$$\mathrm{d}X_t = -\big(\nabla V(X_t) + \nabla \boldsymbol{\beta}(X_t) + \nabla(W * \mathrm{P}_t^{\boldsymbol{\beta}})(X_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^{\boldsymbol{\beta}}, \qquad t_0 \le t \le T \tag{1.13}$$

which is constructed by adding a perturbation $\boldsymbol{\beta}\colon \mathbb{R}^n \to \mathbb{R}$ to the confinement potential[1] of the original McKean–Vlasov SDE (1.1). In other words, from time t_0 onward, the perturbed diffusion drifts in a direction different from that of the original diffusion, hence the perturbed curve of time-marginal distributions $(\mathrm{P}_t^{\boldsymbol{\beta}})_{t_0 \le t \le T}$ also evolves differently from the unperturbed curve $(\mathrm{P}_t)_{t_0 \le t \le T}$. In parallel with the unperturbed case, we may compute the dynamics of the perturbed relative entropy process associated with (1.13). As a consequence, we derive the rate of relative entropy dissipation for the perturbed McKean–Vlasov diffusion. On the other hand, the rate of change of the Wasserstein distance along the perturbed curve $(\mathrm{P}_t^{\boldsymbol{\beta}})_{t_0 \le t \le T}$ can be computed based on the general theory of metric derivative of absolutely continuous curves, see [3]. Finally, comparing these two rates in both the perturbed and unperturbed settings, allows us to establish the gradient flow property.

1.2. The HWBI inequality

The second application of our trajectorial approach deals with the HWBI inequality [2, Theorem 4.2], which is an extension of the HWI inequality [50]. It relates not only relative entropy (H), Wasserstein distance (W), and relative Fisher information (I), but also barycenter (B). These quantities are defined as follows: for two probability measures $\nu, \mu \in \mathscr{P}(\mathbb{R}^n)$, the *relative entropy* of ν with respect to μ is defined by

$$H(\nu \,|\, \mu) := \begin{cases} \displaystyle\int_{\mathbb{R}^n} \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\mu, & \text{if } \nu \ll \mu \\ +\infty, & \text{otherwise,} \end{cases} \tag{1.14}$$

the *relative Fisher information* of ν with respect to μ is given by

$$I(\nu \,|\, \mu) := \begin{cases} \displaystyle\int_{\mathbb{R}^n} \left|\nabla \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)\right|^2 \mathrm{d}\mu, & \text{if } \nu \ll \mu \\ +\infty, & \text{otherwise,} \end{cases} \tag{1.15}$$

and the *barycenter* of a probability measure $\nu \in \mathscr{P}_2(\mathbb{R}^n)$ is defined as $b(\nu) := \int_{\mathbb{R}^n} x\,\mathrm{d}\nu(x) \in \mathbb{R}^n$, where the integral is understood as a Bochner integral. Informally, the HWBI inequality then states that any two probability measures $\nu_0, \nu_1 \in \mathscr{P}_2(\mathbb{R}^n)$ satisfy

$$H(\nu_0 \,|\, \mu_0) - H(\nu_1 \,|\, \mu_1) \le \sqrt{I(\nu_0 \,|\, \mu_0^{\uparrow})}\, W_2(\nu_0, \nu_1) - \tfrac{\kappa_V + \kappa_W}{2}\, W_2^2(\nu_0, \nu_1) + \tfrac{\kappa_W}{2}\, |b(\nu_0) - b(\nu_1)|^2, \tag{1.16}$$

[1] As we will see, the steepest descent property is already visible by perturbing the confinement potential from V to $V + \boldsymbol{\beta}$, thus we avoid complicating the setup further by adding another perturbation to the interaction potential W.

where μ_0, μ_1, $\mu_0^{\uparrow}$ are some appropriate σ-finite reference measures depending on the potentials V, W (see Subsection 3.3 for the details), and $\kappa_V, \kappa_W \in \mathbb{R}$ are the moduli of uniform convexity for V, W. This inequality describes the evolution of the relative entropy along the *displacement interpolation* $(\nu_t)_{0\le t\le 1}$ between ν_0 and ν_1. Compared with the HWI inequality, there are two additional terms on the right-hand side of (1.16) contributed by the interaction energy functional $\mathcal{W}$ of (1.3). Intuitively, the κ_W-uniform convexity of W leads to the first additional term $-\frac{\kappa_W}{2}W_2^2(\nu_0,\nu_1)$, which alone would correspond to the κ_W-uniform displacement convexity of $\mathcal{W}$ along $(\nu_t)_{0\le t\le 1}$. But since $\mathcal{W}(p)$ is invariant under any translation of p, the functional $\mathcal{W}$ might fail to be uniformly displacement convex when the barycenter shifts. This suggests that the barycentric shift along $(\nu_t)_{0\le t\le 1}$ should be factored out of the consideration of the displacement convexity of $\mathcal{W}$, which is intuitively why the second additional term $\frac{\kappa_W}{2}|b(\nu_0)-b(\nu_1)|^2$ in (1.16) appears.

Coming back to our second application, we illustrate how our approach yields a trajectorial proof of the inequality (1.16), in the slightly strengthened form of [18, Theorem 4.1] and [25, Theorem D.50]. Much of this consists of arguments similar in spirit to our main result (1.10), but with one key difference: instead of the time-marginals of the McKean–Vlasov diffusion, we apply the trajectorial approach to the displacement interpolation $(\nu_t)_{0\le t\le 1}$. In this regard, our derivation can be seen as a generalization of the trajectorial proof of the HWI inequality in [40, Section 4.2]; see also [38, Section 9.4], where the same idea was used to derive a discrete version of the HWI inequality in a Riemannian-geometric framework. Let us also point out that for the proof of the HWBI inequality we shall impose convexity assumptions (see Assumptions 3.18) on the potentials V, W. We do not require these assumptions in the rest of the paper.

In the literature, similar trajectorial approaches have also been applied in the context of martingale inequalities [1,5], functional inequalities [4,15,31,41], and their stability estimates [22,23]. In particular, we refer to [4, Corollary 1.4] for a related HWI inequality derived from the entropic interpolation of the mean-field Schrödinger problem.

1.3. Organization of the paper

We set up the probabilistic framework and discuss some regularity assumptions in Section 2. In Section 3 we state our main trajectorial results, Theorem 3.1 and Theorem 3.9, and develop two explicit examples for illustration. As immediate consequences, we derive the classical relative entropy dissipation identities in Corollary 3.4 and Corollary 3.10. Building on these results, we formulate the gradient flow property of the granular media equation in Theorem 3.15. The HWBI inequality is then stated in Theorem 3.19. The proofs of the trajectorial results and of the HWBI inequality are developed in Section 4. Some proofs of auxiliary results postponed in previous sections are contained in Section 5.

2. The probabilistic framework

2.1. The setting

We fix a terminal time $T \in (0,\infty)$ and let $\Omega := C([0,T];\mathbb{R}^n)$ be the path space of $\mathbb{R}^n$-valued continuous functions defined on $[0,T]$. We denote by $(X_t)_{0\le t\le T}$ the canonical process defined by $X_t(\omega) := \omega(t)$ for $\omega \in \Omega$, and fix a probability distribution $\mathrm{P}_0 \in \mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$.

As will be shown in Lemma 2.2, under the Assumptions 2.1 below, the SDE (1.1) with initial distribution P_0 has a unique strong solution, when it is posed on an arbitrary filtered probability space. This implies that there exists a probability measure $\mathbb{P}$ on Ω and a $\mathbb{P}$-Brownian motion $(B_t)_{0\le t\le T}$ such that

the SDE (1.1) holds. We write $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ for the right-continuous augmentation of the canonical filtration.

For each time $t \in [0,T]$, we denote by $\mathrm{P}_t := \mathbb{P} \circ X_t^{-1}$ the distribution of X_t under $\mathbb{P}$, and by p_t the corresponding probability density function on $\mathbb{R}^n$. The density functions $(p_t)_{0\le t\le T}$ then solve the granular media equation (1.11).

2.2. Regularity assumptions

Assumptions 2.1. The following regularity assumptions will be used frequently.

(i) The functions $V, W\colon \mathbb{R}^n \to [0,\infty)$ are smooth and have Lipschitz continuous gradients with Lipschitz constants $\|\nabla V\|_{\mathrm{Lip}}$, $\|\nabla W\|_{\mathrm{Lip}}$. All derivatives of V and W grow at most exponentially as $|x|$ tends to infinity, and the first derivatives are of linear growth. The latter condition means that there exists a constant $C > 0$ such that

$$\forall\, x \in \mathbb{R}^n\colon \qquad |\nabla V(x)| \le C(1+|x|), \qquad |\nabla W(x)| \le C(1+|x|). \tag{2.1}$$

Furthermore, the function W is even (in other words, symmetric), i.e., $W(x) = W(-x)$ for all $x \in \mathbb{R}^n$.

(ii) The probability distribution P_0 is an element of the space $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$ and the corresponding probability density function $\mathbb{R}^n \ni x \mapsto p_0(x)$ is strictly positive. Moreover, the initial free energy $\mathscr{F}(p_0)$ is finite.

These assumptions ensure that the equation (1.1) belongs to a broad class of strongly solvable McKean–Vlasov SDEs. We relegate the proof of the following result to Subsection 5.1.

Lemma 2.2. *Suppose Assumptions 2.1 hold. Then on an arbitrary filtered probability space, the McKean–Vlasov SDE (1.1) has a pathwise unique, strong solution $(X_t)_{0\le t\le T}$ satisfying*

$$\mathbb{E}\left[\sup_{0\le t\le T} |X_t|^2\right] < \infty. \tag{2.2}$$

Moreover, its marginal distributions $(\mathrm{P}_t)_{0\le t\le T}$ belong to $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$, and the corresponding curve of probability density functions $(p_t)_{0\le t\le T}$ is a classical solution of the granular media equation (1.11).

2.3. Probabilistic representations of gradient flow functionals

To set up our framework, the first step is to express the free energy as well as the relative entropy dissipation functional in probabilistic terms. To this end, we introduce the generalized potential $\Psi\colon \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n) \to [0,\infty)$ and its close relative $\Psi^{\uparrow}$ given by

$$\Psi(x,\mu) := V(x) + \tfrac{1}{2}(W * \mu)(x), \qquad \Psi^{\uparrow}(x,\mu) := V(x) + (W * \mu)(x) \tag{2.3}$$

for $(x,\mu) \in \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n)$. Furthermore, we define the density functions

$$q(x,\mu) := \mathrm{e}^{-\Psi(x,\mu)}, \qquad q^{\uparrow}(x,\mu) := \mathrm{e}^{-\Psi^{\uparrow}(x,\mu)}, \qquad q^{\downarrow}(x) := \mathrm{e}^{-V(x)} \tag{2.4}$$

and the corresponding generalized likelihood ratio functions

$$\ell_t(x,\mu) \coloneqq \frac{p_t(x)}{q(x,\mu)}\,, \qquad \ell_t^{\uparrow}(x,\mu) \coloneqq \frac{p_t(x)}{q^{\uparrow}(x,\mu)}\,, \qquad \ell_t^{\downarrow}(x) \coloneqq \frac{p_t(x)}{q^{\downarrow}(x)} \tag{2.5}$$

for $t \in [0,T]$. Note that if $W \equiv 0$, these three likelihood ratio functions coincide.

For each time $t \in [0,T]$, we introduce σ-finite measures on the Borel sets of $\mathbb{R}^n$, given by

$$\mathrm{Q}_t(A) \coloneqq \int_A q(x,\mathrm{P}_t)\,\mathrm{d}x\,, \qquad \mathrm{Q}_t^{\uparrow}(A) \coloneqq \int_A q^{\uparrow}(x,\mathrm{P}_t)\,\mathrm{d}x\,, \qquad A \in \mathscr{B}(\mathbb{R}^n). \tag{2.6}$$

Intuitively, these measures are (unnormalized) time-dependent Gibbs distributions. If $W \equiv 0$, they coincide with the true Gibbs distribution of the Langevin–Smoluchowski equation (1.6), which is also its stationary distribution (when normalized to a probability measure).

With these definitions, we can now write the gradient flow functionals $\mathscr{F}$ and $\mathscr{D}$, introduced in (1.2) and (1.4), in probabilistic terms: the relative entropy (defined in (1.14)) of P_t with respect to Q_t and the relative Fisher information (defined in (1.15)) of P_t with respect to $\mathrm{Q}_t^{\uparrow}$ can be expressed respectively as

$$H(\mathrm{P}_t \,|\, \mathrm{Q}_t) = \mathbb{E}_{\mathbb{P}}\big[\log \ell_t(X_t,\mathrm{P}_t)\big]\,, \qquad I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow}) = \mathbb{E}_{\mathbb{P}}\big[|\nabla \log \ell_t^{\uparrow}(X_t,\mathrm{P}_t)|^2\big]\,; \tag{2.7}$$

and we have the relations $H(\mathrm{P}_t \,|\, \mathrm{Q}_t) = \mathscr{F}(\mathrm{P}_t)$ as well as $I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow}) = \mathscr{D}(\mathrm{P}_t)$. In particular, the relative entropy $H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ can be written as the $\mathbb{P}$-expectation of the *relative entropy process*

$$\log \ell_t(X_t,\mathrm{P}_t) = \log p_t(X_t) + V(X_t) + \tfrac{1}{2}(W * \mathrm{P}_t)(X_t)\,, \qquad 0 \le t \le T. \tag{2.8}$$

The dynamics of this stochastic process, together with its perturbed counterpart to be introduced in Subsection 3.2 below, will be our main objects of interest.

Remark 2.3. If the reference measure Q_t in (2.6) is a *probability* measure, then the expression (2.7) matches the classical definition of relative entropy given in (1.14). In the general case when Q_t is a σ-finite measure, the definition (1.14) is also valid under the condition that P_t has finite second moment, with the only difference that the range of the function $t \mapsto H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ is extended from $[0,\infty]$ to $(-\infty,\infty]$; we refer to [39, Appendix C] or [19, Section 3] for the details.

3. Main results

3.1. Trajectorial dissipation of relative entropy for McKean–Vlasov diffusions

Our first main result is the semimartingale decomposition of the relative entropy process (2.8). It describes the dissipation of relative entropy along every trajectory of a particle undergoing the McKean–Vlasov dynamics (1.1). In the same spirit as the trajectorial approaches of [29] and [40], we shall study the dynamics of the relative entropy process in the backward direction of time. Concretely, we consider for arbitrary, fixed $T \in (0,\infty)$ the time-reversed canonical process

$$\overline{X}_s \coloneqq X_{T-s}\,, \qquad 0 \le s \le T \tag{3.1}$$

on the filtered probability space $(\Omega, \mathbb{G}, \mathbb{P})$, where $\mathbb{G} = (\mathcal{G}_s)_{0\le s\le T}$ is the $\mathbb{P}$-augmented filtration generated by $(\overline{X}_s)_{0\le s\le T}$.

In order to formulate Theorem 3.1 below, we introduce the time-reversed *Fisher information process*

$$\begin{aligned}\overline{I}_s \coloneqq &\Big(\big|\nabla \log \overline{\ell}_s^{\downarrow}\big|^2 + \tfrac{1}{2}\big|\nabla(W * \overline{\mathrm{P}}_s)\big|^2 + \Big\langle \tfrac{1}{2}\nabla(W * \overline{\mathrm{P}}_s),\, 2\nabla \log \overline{\ell}_s^{\downarrow} + \nabla V\Big\rangle\Big)(\overline{X}_s)\\ &- \mathbb{E}_{\tilde{\mathbb{P}}}\Big[\Big\langle \tfrac{1}{2}\nabla W(\overline{X}_s - \overline{Y}_s), \Big(2\nabla \log \overline{\ell}_s^{\downarrow} - \nabla V + \nabla(W * \overline{\mathrm{P}}_s)\Big)(\overline{Y}_s)\Big\rangle\Big]\end{aligned} \tag{3.2}$$

for $0 \le s \le T$. Here, the process $(\overline{Y}_s)_{0\le s\le T}$ is defined on another probability space $(\tilde{\Omega}, \tilde{\mathbb{G}}, \tilde{\mathbb{P}})$ such that the tuple $(\tilde{\Omega}, \tilde{\mathbb{G}}, \tilde{\mathbb{P}}, (\overline{Y}_s)_{0\le s\le T})$ is an exact copy of $(\Omega, \mathbb{G}, \mathbb{P}, (\overline{X}_s)_{0\le s\le T})$. A bar over a letter means that time is reversed as in (3.1).

We also define the time-reversed *cumulative Fisher information process* as the time integral

$$\overline{F}_s \coloneqq \int_0^s \overline{I}_u \,\mathrm{d}u, \qquad 0 \le s \le T. \tag{3.3}$$

This process will act as the compensator in the semimartingale decomposition of the relative entropy process (2.8). Its relation with the relative Fisher information (2.7) will be given in (3.7) below.

Theorem 3.1. *Suppose Assumptions 2.1 hold. On $(\Omega, \mathbb{G}, \mathbb{P})$, the time-reversed relative entropy process*

$$\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) = \log \ell_{T-s}(X_{T-s}, \mathrm{P}_{T-s}), \qquad 0 \le s \le T \tag{3.4}$$

admits the semimartingale decomposition

$$\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) - \log \overline{\ell}_0(\overline{X}_0, \overline{\mathrm{P}}_0) = \overline{M}_s + \overline{F}_s. \tag{3.5}$$

Here $(\overline{M}_s)_{0\le s\le T}$ is the $L^2(\mathbb{P})$-bounded martingale

$$\overline{M}_s \coloneqq \int_0^s \Big\langle \nabla \log \overline{\ell}_u(\overline{X}_u, \overline{\mathrm{P}}_u),\, \sqrt{2}\,\mathrm{d}\overline{B}_u \Big\rangle, \tag{3.6}$$

with $(\overline{B}_s)_{0\le s\le T}$ a $\mathbb{P}$-Brownian motion of the backward filtration $\mathbb{G}$, and the compensator $(\overline{F}_s)_{0\le s\le T}$ satisfies

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s\big] = \int_0^s I\big(\overline{\mathrm{P}}_u \,\big|\, \overline{\mathrm{Q}}_u^{\uparrow}\big)\,\mathrm{d}u = \mathbb{E}_{\mathbb{P}}\bigg[\int_0^s \big|\nabla \log \overline{\ell}_u^{\uparrow}(\overline{X}_u, \overline{\mathrm{P}}_u)\big|^2\,\mathrm{d}u\bigg] < \infty. \tag{3.7}$$

3.1.1. Examples

We give two concrete examples to illustrate Theorem 3.1.

Example 3.2. We set $n = 1$ and specialize Theorem 3.1 to the case of quadratic confinement potential $V(x) = \frac{x^2}{2}$ and no interaction potential $W \equiv 0$. The initial position X_0 is chosen to be independent of $(B_t)_{0\le t\le T}$ and to be normally distributed with mean 0 and variance $\sigma_0^2 > 0$. In this case, the SDE of (1.1) becomes

$$\mathrm{d}X_t = -X_t\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \qquad 0 \le t \le T \tag{3.8}$$

and its solution is given by the Ornstein–Uhlenbeck process

$$X_t = \mathrm{e}^{-t} X_0 + \sqrt{2}\int_0^t \mathrm{e}^{u-t}\,\mathrm{d}B_u, \qquad 0 \le t \le T \tag{3.9}$$

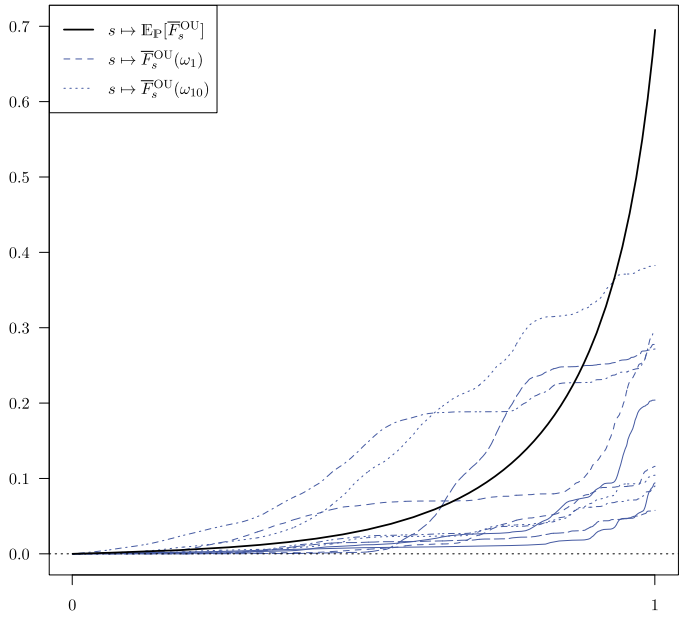

Figure 1. Simulations of the cumulative Fisher information process (3.11) for the Ornstein–Uhlenbeck diffusion (3.8).

with probability density function

$$p_t(x) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\big(-\tfrac{x^2}{2\sigma_t^2}\big), \qquad \sigma_t^2 := 1 + \mathrm{e}^{-2t}(\sigma_0^2 - 1). \tag{3.10}$$

Recalling (2.3) – (2.5) and using (3.10), we see that in this setting the cumulative Fisher information process of (3.3) is explicitly given by

$$\overline{F}_s^{\mathrm{OU}} = \int_0^s \Big(\nabla \log \overline{\ell}_u^{\downarrow}(\overline{X}_u)\Big)^2 \,\mathrm{d}u = \int_0^s \Big(\nabla \log \overline{p}_u(\overline{X}_u) + \overline{X}_u\Big)^2 \,\mathrm{d}u = \int_0^s \big(1 - \tfrac{1}{\overline{\sigma}_u^2}\big)^2 \overline{X}_u^2 \,\mathrm{d}u \tag{3.11}$$

for $0 \le s \le T$. In particular, the non-negativity of the integrand in (3.11) implies that the relative entropy decreases along almost every trajectory.

Now we set $T = 1$ and $\sigma_0^2 = 0.1$. The blue lines in Figure 1 represent ten simulated trajectories $s \mapsto \overline{F}_s^{\mathrm{OU}}(\omega_i)$, for $i = 1, \dots, 10$. The thick black line plots the expected path $s \mapsto \mathbb{E}_{\mathbb{P}}[\overline{F}_s^{\mathrm{OU}}]$ of all possible trajectories.

Example 3.3. We set again $n = 1$ and now consider the case of no confinement potential $V \equiv 0$, quadratic interaction potential $W(x) = \frac{x^2}{2}$, and a centered Gaussian initial position X_0 with variance $\sigma_0^2 > 0$, which is independent of the Brownian motion $(B_t)_{t \ge 0}$. In this case, for any $t \ge 0$, the drift term of the SDE of (1.1) is

$$-\nabla(W * \mathrm{P}_t)(X_t) = -\int_{\mathbb{R}^n} \nabla W(X_t - y) p_t(y) \,\mathrm{d}y = -\big(X_t - \mathbb{E}[X_t]\big). \tag{3.12}$$

In particular, the drift term depends on the distribution P_t only through its mean. Substituting it into (1.1), this SDE reduces to

$$\mathrm{d}X_t = -\big(X_t - \mathbb{E}[X_t]\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t\,, \qquad 0 \le t \le T. \tag{3.13}$$

This type of nonlinear, self-interacting SDE has been studied since [6], where it was shown that its solution is also given by the Ornstein–Uhlenbeck process of (3.9). Therefore, similar computations as in Example 3.2 show that in this setting the cumulative Fisher information process is given by

$$\overline{F}_s^{\mathrm{NL}} = \int_0^s \Big(\big(\tfrac{1}{\overline{\sigma}_u^4} + \tfrac{1}{2} - \tfrac{1}{\overline{\sigma}_u^2}\big)\,\overline{X}_u^2 + \mathbb{E}_{\tilde{\mathbb{P}}}\Big[\tfrac{1}{2}(\overline{X}_u - \overline{Y}_u)\big(\tfrac{2}{\overline{\sigma}_u^2}\overline{Y}_u - \overline{Y}_u\big) \Big] \Big)\,\mathrm{d}u \tag{3.14}$$

for $0 \le s \le T$. Using the fact that $(\overline{X}_u)_\#(\tilde{\mathbb{P}}) = (\overline{Y}_u)_\#(\tilde{\mathbb{P}}) = \mathcal{N}(0,\overline{\sigma}_u^2)$, which we have from (3.10), we can compute the expectation appearing in (3.14) and obtain

$$\overline{F}_s^{\mathrm{NL}} = \int_0^s \Big(\big(\tfrac{1}{\overline{\sigma}_u^4} + \tfrac{1}{2} - \tfrac{1}{\overline{\sigma}_u^2}\big)\,\overline{X}_u^2 + \frac{\overline{\sigma}_u^2}{2} - 1 \Big)\,\mathrm{d}u. \tag{3.15}$$

Clearly, $\overline{F}_s^{\mathrm{OU}} \neq \overline{F}_s^{\mathrm{NL}}$. In particular, the integrand in (3.15) is non-negative if and only if

$$\overline{X}_u^2 \ge \big(\tfrac{1}{\overline{\sigma}_u^4} + \tfrac{1}{2} - \tfrac{1}{\overline{\sigma}_u^2}\big)^{-1}\big(1 - \tfrac{\overline{\sigma}_u^2}{2}\big).$$

In other words, as opposed to Example 3.2, relative entropy only decreases along a trajectory if $\overline{X}_u$ is far from its mean. However, after taking expectations in (3.11) and (3.15), we see that the expected rate of relative entropy dissipation in both cases is equal to

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s^{\mathrm{OU}}\big] = \mathbb{E}_{\mathbb{P}}\big[\overline{F}_s^{\mathrm{NL}}\big] = \int_0^s \big(\overline{\sigma}_u - \tfrac{1}{\overline{\sigma}_u}\big)^2\,\mathrm{d}u\,, \qquad 0 \le s \le T. \tag{3.16}$$

Now we set again $T = 1$ and $\sigma_0^2 = 0.1$. In the same vein as in Figure 1, we plot in Figure 2 the paths of ten simulated trajectories $s \mapsto \overline{F}_s^{\mathrm{NL}}(\omega_i)$, for $i = 1,\dots,10$. We observe that some of the red lines describing the paths of these trajectories indeed take negative values. In other words, the cumulative Fisher information process of (3.15), and hence its integrand, can both be negative. Finally, the thick black line in Figure 2 follows the expected path $s \mapsto \mathbb{E}_{\mathbb{P}}[\overline{F}_s^{\mathrm{NL}}]$ of all possible trajectories. According to (3.16), this is the same black line as in Figure 1.

3.1.2. Consequences of Theorem 3.1

We now return to the general statement of Theorem 3.1 and deduce some direct consequences. By averaging the trajectorial result of Theorem 3.1 according to the path measure $\mathbb{P}$, we derive the well-known *relative entropy identity* (3.17) and the *dissipation of relative entropy* (3.18) below. A sketch of proof for the latter result was first given in [13, Proposition 2.1].

Corollary 3.4. *Suppose Assumptions 2.1 hold. For all $t, t_0 \in [0,T]$, we have the relative entropy identity*

$$H(\mathrm{P}_t \,|\, \mathrm{Q}_t) - H(\mathrm{P}_{t_0} \,|\, \mathrm{Q}_{t_0}) = -\int_{t_0}^t I(\mathrm{P}_u \,|\, \mathrm{Q}_u^{\uparrow})\,du. \tag{3.17}$$

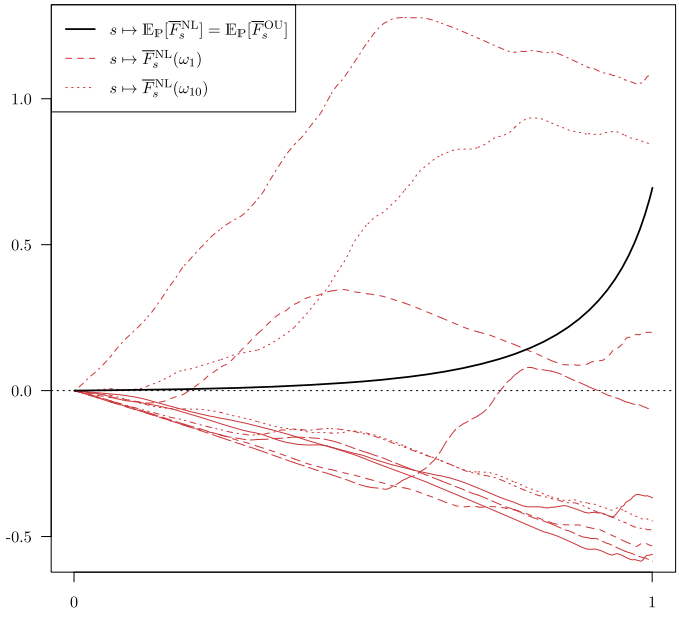

Figure 2. Simulations of the cumulative Fisher information process (3.15) for the nonlinear, self-interacting diffusion (3.13).

In particular, the relative entropy function $t \mapsto H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ is monotonically decreasing. Furthermore, for Lebesgue-a.e. $t \in [0,T]$, the relative Fisher information $I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow})$ is finite, and the rate of relative entropy dissipation is given by

$$\frac{d}{dt} H(\mathrm{P}_t \,|\, \mathrm{Q}_t) = -I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow}). \tag{3.18}$$

Proof. The identity (3.17) follows by taking expectations with respect to the probability measure $\mathbb{P}$ in (3.5), recalling the definitions of (2.7), using (3.7), and invoking the fact that the $\mathbb{P}$-expectation of the martingale (3.6) vanishes. Finally, applying the Lebesgue differentiation theorem to the monotone function $t \mapsto H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ gives (3.18). □

Remark 3.5. The relation (3.18) describes the temporal dissipation of relative entropy at the *ensemble* level. It asserts that the rate of decay of the relative entropy $t \mapsto H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ is given by the relative Fisher information $I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow})$.

Finally, let us place ourselves again on the filtered probability space $(\Omega, \mathbb{G}, \mathbb{P})$ as in Theorem 3.1 and emphasize that this trajectorial result is valid along almost every trajectory $s \mapsto \overline{X}_s(\omega)$ of the underlying McKean–Vlasov process. As a consequence, by taking conditional expectations, we can generalize (3.18) and deduce the following trajectorial rate of relative entropy dissipation.

Corollary 3.6. *Suppose Assumptions 2.1 hold and $\int_0^T \mathbb{E}_{\mathbb{P}}[|\overline{I}_u|]\,\mathrm{d}u < \infty$. For $\mathbb{P}$-a.e. $\omega \in \Omega$ there exists a Lebesgue null set $N_\omega \subseteq [0,T]$ such that for any $s_0 \in [0,T] \setminus N_\omega$ we have*

$$\lim_{s \downarrow s_0} \frac{\mathbb{E}_{\mathbb{P}}\big[\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) \,|\, \mathcal{G}_{s_0}\big](\omega) - \log \overline{\ell}_{s_0}\big(\overline{X}_{s_0}(\omega), \overline{\mathrm{P}}_{s_0}\big)}{s - s_0} = \overline{I}_{s_0}(\omega). \tag{3.19}$$

Remark 3.7. Recalling (3.7), we observe that $\mathbb{E}_{\mathbb{P}}[\overline{I}_{s_0}] = I(\overline{\mathrm{P}}_{s_0} \,|\, \overline{\mathrm{Q}}^{\uparrow}_{s_0})$. Therefore the limiting assertion (3.19) can indeed be viewed as a trajectorial version of the deterministic relative entropy dissipation identity (3.18).

Proof. We let $0 \le s_0 \le s \le T$. By (3.5), (3.3) and Fubini's theorem, we have for $\mathbb{P}$-a.e. $\omega \in \Omega$

$$\mathbb{E}_{\mathbb{P}}\big[\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) \,|\, \mathcal{G}_{s_0}\big](\omega) - \log \overline{\ell}_{s_0}\big(\overline{X}_{s_0}(\omega), \overline{\mathrm{P}}_{s_0}\big) = \int_{s_0}^{s} \mathbb{E}_{\mathbb{P}}\big[\overline{I}_u \,|\, \mathcal{G}_{s_0}\big](\omega)\,\mathrm{d}u. \tag{3.20}$$

Furthermore, Jensen's inequality gives

$$\mathbb{E}_{\mathbb{P}}\bigg[\int_{s_0}^{s} \big|\mathbb{E}_{\mathbb{P}}\big[\overline{I}_u \,|\, \mathcal{G}_{s_0}\big]\big|\,\mathrm{d}u\bigg] \le \int_{s_0}^{s} \mathbb{E}_{\mathbb{P}}\big[|\overline{I}_u|\big]\,\mathrm{d}u < \infty, \tag{3.21}$$

which implies

$$\int_{s_0}^{s} \big|\mathbb{E}_{\mathbb{P}}\big[\overline{I}_u \,|\, \mathcal{G}_{s_0}\big](\omega)\big|\,\mathrm{d}u < \infty \qquad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{3.22}$$

By the Lebesgue differentiation theorem, for every such ω there exists a Lebesgue null set $N_\omega \subseteq [0,T]$ so that the limiting assertion

$$\lim_{s \downarrow s_0} \frac{\int_{s_0}^{s} \mathbb{E}_{\mathbb{P}}\big[\overline{I}_u \,|\, \mathcal{G}_{s_0}\big](\omega)\,\mathrm{d}u}{s - s_0} = \mathbb{E}_{\mathbb{P}}\big[\overline{I}_{s_0} \,|\, \mathcal{G}_{s_0}\big](\omega) = \overline{I}_{s_0}(\omega) \tag{3.23}$$

holds for every $s_0 \in [0,T] \setminus N_\omega$. Finally, combining (3.20) and (3.23) proves (3.19). □

3.2. Gradient flow structure of the granular media equation

In this subsection we apply the trajectorial approach of Subsection 3.1 in order to formulate the gradient flow property of the granular media equation (1.11). To this end, we consider a function $\boldsymbol{\beta}\colon \mathbb{R}^n \to \mathbb{R}$, which will be treated as a perturbation potential. We denote by $V^{\boldsymbol{\beta}} := V + \boldsymbol{\beta}$ the perturbed confinement potential and invoke the following regularity assumptions.

Assumptions 3.8. The function $\boldsymbol{\beta}\colon \mathbb{R}^n \to \mathbb{R}$ is smooth and compactly supported, and we require that Assumptions 2.1 are still satisfied if we replace V by $V^{\boldsymbol{\beta}}$.

Note that Assumptions 3.8 allow us to apply Lemma 2.2 to the "perturbed" McKean–Vlasov SDE

$$\mathrm{d}X_t = -\big(\nabla V^{\boldsymbol{\beta}}(X_t) + \nabla(W * \mathrm{P}^{\boldsymbol{\beta}}_t)(X_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B^{\boldsymbol{\beta}}_t, \qquad t_0 \le t \le T \tag{3.24}$$

starting at time $t_0 \ge 0$, with X_{t_0} having initial distribution $\mathrm{P}^{\boldsymbol{\beta}}_{t_0} = \mathrm{P}_{t_0}$. Therefore, by analogy with Subsection 2.1, we can construct a probability measure $\mathbb{P}^{\boldsymbol{\beta}}$ on $\Omega := C([t_0,T];\mathbb{R}^n)$, under which the canonical process $(X_t)_{t_0 \le t \le T}$ satisfies the SDE (3.24), with $(B^{\boldsymbol{\beta}}_t)_{t_0 \le t \le T}$ being a $\mathbb{P}^{\boldsymbol{\beta}}$-Brownian motion.

For each time $t \in [t_0,T]$, we denote by $\mathrm{P}^{\boldsymbol{\beta}}_t := \mathbb{P}^{\boldsymbol{\beta}} \circ X_t^{-1}$ the probability distribution and by $p^{\boldsymbol{\beta}}_t$ the probability density function of X_t under $\mathbb{P}^{\boldsymbol{\beta}}$. The "perturbed" curve of density functions $(p^{\boldsymbol{\beta}}_t)_{t_0 \le t \le T}$ then satisfies the *perturbed granular media equation*

$$\begin{cases} \partial_t p^{\boldsymbol{\beta}}_t(x) = \operatorname{div}\Big(\nabla p^{\boldsymbol{\beta}}_t(x) + p^{\boldsymbol{\beta}}_t(x)\nabla V^{\boldsymbol{\beta}}(x) + p^{\boldsymbol{\beta}}_t(x)\nabla(W * p^{\boldsymbol{\beta}}_t)(x)\Big), & (t,x) \in (t_0,T) \times \mathbb{R}^n, \\ p^{\boldsymbol{\beta}}_{t_0}(x) = p_{t_0}(x), & x \in \mathbb{R}^n. \end{cases} \tag{3.25}$$

By analogy with (2.3), we define the perturbed potentials

$$\Psi^{\boldsymbol{\beta}}(x,\mu) := V^{\boldsymbol{\beta}}(x) + \tfrac{1}{2}(W*\mu)(x), \qquad \Psi^{\boldsymbol{\beta}\uparrow}(x,\mu) := V^{\boldsymbol{\beta}}(x) + (W*\mu)(x), \qquad \Psi^{\boldsymbol{\beta}\downarrow} := V^{\boldsymbol{\beta}} \tag{3.26}$$

for $(x,\mu) \in \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n)$. In parallel to (2.5), we introduce the *perturbed likelihood ratio functions*

$$\ell_t^{\boldsymbol{\beta}}(x,\mu) := \frac{p_t^{\boldsymbol{\beta}}(x)}{q(x,\mu)}, \qquad \ell_t^{\boldsymbol{\beta}\uparrow}(x,\mu) := \frac{p_t^{\boldsymbol{\beta}}(x)}{q^{\uparrow}(x,\mu)}, \qquad \ell_t^{\boldsymbol{\beta}\downarrow}(x) := \frac{p_t^{\boldsymbol{\beta}}(x)}{q^{\downarrow}(x)} \tag{3.27}$$

for $t \in [t_0,T]$. Finally, we define the σ-finite measures

$$\mathrm{Q}_t^{\boldsymbol{\beta}}(A) := \int_A q(x,\mathrm{P}_t^{\boldsymbol{\beta}})\,\mathrm{d}x, \qquad \mathrm{Q}_t^{\boldsymbol{\beta}\uparrow}(A) := \int_A q^{\uparrow}(x,\mathrm{P}_t^{\boldsymbol{\beta}})\,\mathrm{d}x, \qquad A \in \mathscr{B}(\mathbb{R}^n). \tag{3.28}$$

They are the perturbed versions of the measures Q_t and $\mathrm{Q}_t^{\uparrow}$ defined in (2.6). The relative entropy of $\mathrm{P}_t^{\boldsymbol{\beta}}$ with respect to $\mathrm{Q}_t^{\boldsymbol{\beta}}$ is then given by

$$H\big(\mathrm{P}_t^{\boldsymbol{\beta}} \,\big|\, \mathrm{Q}_t^{\boldsymbol{\beta}}\big) = \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\big[\log \ell_t^{\boldsymbol{\beta}}(X_t,\mathrm{P}_t^{\boldsymbol{\beta}})\big] = \mathscr{F}(\mathrm{P}_t^{\boldsymbol{\beta}}) \tag{3.29}$$

and the relative Fisher information of $\mathrm{P}_t^{\boldsymbol{\beta}}$ with respect to $\mathrm{Q}_t^{\boldsymbol{\beta}\uparrow}$ equals

$$I\big(\mathrm{P}_t^{\boldsymbol{\beta}} \,\big|\, \mathrm{Q}_t^{\boldsymbol{\beta}\uparrow}\big) = \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\Big[\big|\nabla \log \ell_t^{\boldsymbol{\beta}\uparrow}(X_t,\mathrm{P}_t^{\boldsymbol{\beta}})\big|^2\Big] = \mathscr{D}(\mathrm{P}_t^{\boldsymbol{\beta}}). \tag{3.30}$$

The following trajectorial result, Theorem 3.9 below, provides the semimartingale decomposition of the *perturbed relative entropy process*

$$\log \ell_t^{\boldsymbol{\beta}}(X_t,\mathrm{P}_t^{\boldsymbol{\beta}}) = \log p_t^{\boldsymbol{\beta}}(X_t) + V(X_t) + \tfrac{1}{2}(W*\mathrm{P}_t^{\boldsymbol{\beta}})(X_t), \qquad t_0 \le t \le T. \tag{3.31}$$

In line with its unperturbed counterpart, Theorem 3.1, we shall formulate this result in the reverse direction of time. We first introduce the perturbed analogues of (3.2) and (3.3): the *time-reversed perturbed Fisher information process* is defined as

$$\overline{I}_s^{\boldsymbol{\beta}} := \Big(\big|\nabla \log \overline{\ell}_s^{\boldsymbol{\beta}\downarrow}\big|^2 + \tfrac{1}{2}\big|\nabla(W*\overline{\mathrm{P}}_s^{\boldsymbol{\beta}})\big|^2 + \Big\langle \tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}_s^{\boldsymbol{\beta}}),\, 2\nabla \log \overline{\ell}_s^{\boldsymbol{\beta}\downarrow} + \nabla V^{\boldsymbol{\beta}}\Big\rangle\Big)(\overline{X}_s) \tag{3.32}$$

$$- \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\Big[\Big\langle \tfrac{1}{2}\nabla W(\overline{X}_s - \overline{Y}_s),\, \Big(2\nabla \log \overline{\ell}_s^{\boldsymbol{\beta}\downarrow} - \nabla V + \nabla(W*\overline{\mathrm{P}}_s^{\boldsymbol{\beta}}) + \nabla\boldsymbol{\beta}\Big)(\overline{Y}_s)\Big\rangle\Big] \tag{3.33}$$

$$+ \Big(\langle \nabla V, \nabla\boldsymbol{\beta}\rangle - \Delta\boldsymbol{\beta}\Big)(\overline{X}_s) \tag{3.34}$$

for all $0 \le s \le T - t_0$, where $(\overline{Y}_s)_{0\le s\le T-t_0}$ is a copy of the process $(\overline{X}_s)_{0\le s\le T-t_0}$ on a copy $(\tilde{\Omega},\tilde{\mathbb{G}},\tilde{\mathbb{P}}^{\boldsymbol{\beta}})$ of the original probability space $(\Omega,\mathbb{G},\mathbb{P}^{\boldsymbol{\beta}})$; the *perturbed cumulative Fisher information process* is defined as

$$\overline{F}_s^{\boldsymbol{\beta}} := \int_0^s \overline{I}_u^{\boldsymbol{\beta}}\,\mathrm{d}u, \qquad 0 \le s \le T - t_0. \tag{3.35}$$

Theorem 3.9. *Suppose Assumptions 3.8 hold. On $(\Omega,\mathbb{G},\mathbb{P}^{\boldsymbol{\beta}})$, the time-reversed perturbed relative entropy process*

$$\log \overline{\ell}_s^{\boldsymbol{\beta}}(\overline{X}_s,\overline{\mathrm{P}}_s^{\boldsymbol{\beta}}) = \log \ell_{T-s}^{\boldsymbol{\beta}}(X_{T-s},\mathrm{P}_{T-s}^{\boldsymbol{\beta}}), \qquad 0 \le s \le T - t_0 \tag{3.36}$$

admits the semimartingale decomposition

$$\log \overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s, \overline{\mathrm{P}}^{\boldsymbol{\beta}}_s) - \log \overline{\ell}^{\boldsymbol{\beta}}_0(\overline{X}_0, \overline{\mathrm{P}}^{\boldsymbol{\beta}}_0) = \overline{M}^{\boldsymbol{\beta}}_s + \overline{F}^{\boldsymbol{\beta}}_s. \tag{3.37}$$

Here $(\overline{M}^{\boldsymbol{\beta}}_s)_{0\le s\le T-t_0}$ *is the* $L^2(\mathbb{P}^{\boldsymbol{\beta}})$*-bounded martingale*

$$\overline{M}^{\boldsymbol{\beta}}_s \coloneqq \int_0^s \Big\langle \nabla \log \overline{\ell}^{\boldsymbol{\beta}}_u(\overline{X}_u, \overline{\mathrm{P}}^{\boldsymbol{\beta}}_u),\, \sqrt{2}\,\mathrm{d}\overline{B}^{\boldsymbol{\beta}}_u \Big\rangle, \tag{3.38}$$

with $(\overline{B}^{\boldsymbol{\beta}}_s)_{0\le s\le T-t_0}$ *a* $\mathbb{P}^{\boldsymbol{\beta}}$*-Brownian motion of the backward filtration* $\mathbb{G}$*, and the compensator (3.35) satisfies*

$$\mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\big[\overline{F}^{\boldsymbol{\beta}}_s\big] = \int_0^s \Big(I\big(\overline{\mathrm{P}}^{\boldsymbol{\beta}}_u \,\big|\, \overline{\mathrm{Q}}^{\boldsymbol{\beta}\uparrow}_u\big) + \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\Big[\Big(\Big\langle \nabla V + \nabla(W * \overline{\mathrm{P}}^{\boldsymbol{\beta}}_u), \nabla\boldsymbol{\beta}\Big\rangle - \Delta\boldsymbol{\beta}\Big)(\overline{X}_u)\Big]\Big)\,\mathrm{d}u < \infty. \tag{3.39}$$

With the dynamics of the time-reversed perturbed relative entropy process at hand, we repeat the same procedure which was carried out for the unperturbed case. Taking expectations with respect to the probability measure $\mathbb{P}^{\boldsymbol{\beta}}$, we arrive at the perturbed relative entropy identity (3.40), and applying the Lebesgue differentiation theorem gives the perturbed relative entropy production identity (3.41).

Corollary 3.10. *Suppose Assumptions 3.8 hold. For all* $0 \le t_0 \le t \le T$*, we have the perturbed relative entropy identity*

$$\begin{aligned} &H\big(\mathrm{P}^{\boldsymbol{\beta}}_t \,\big|\, \mathrm{Q}^{\boldsymbol{\beta}}_t\big) - H\big(\mathrm{P}^{\boldsymbol{\beta}}_{t_0} \,\big|\, \mathrm{Q}^{\boldsymbol{\beta}}_{t_0}\big) \\ &\qquad = -\int_{t_0}^t \Big(I\big(\mathrm{P}^{\boldsymbol{\beta}}_u \,\big|\, \mathrm{Q}^{\boldsymbol{\beta}\uparrow}_u\big) + \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\Big[\Big(\Big\langle \nabla V + \nabla(W * \mathrm{P}^{\boldsymbol{\beta}}_u), \nabla\boldsymbol{\beta}\Big\rangle - \Delta\boldsymbol{\beta}\Big)(X_u)\Big]\Big)\,du. \end{aligned} \tag{3.40}$$

For Lebesgue-a.e. $t_0 \in [0,T]$*, the perturbed rate of relative entropy dissipation is given by*

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|^+_{t=t_0} H\big(\mathrm{P}^{\boldsymbol{\beta}}_t \,\big|\, \mathrm{Q}^{\boldsymbol{\beta}}_t\big) = -\Big(I\big(\mathrm{P}_{t_0} \,\big|\, \mathrm{Q}_{t_0}\big) + \mathbb{E}_{\mathbb{P}}\Big[\Big(\Big\langle \nabla V + \nabla(W * \mathrm{P}_{t_0}), \nabla\boldsymbol{\beta}\Big\rangle - \Delta\boldsymbol{\beta}\Big)(X_{t_0})\Big]\Big). \tag{3.41}$$

Similarly, we have the following trajectorial rate of relative entropy dissipation for the perturbed diffusion.

Corollary 3.11. *Suppose Assumptions 3.8 hold and* $\int_0^{T-t_0} \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}[|\overline{I}^{\boldsymbol{\beta}}_u|]\,\mathrm{d}u < \infty$*. For* $\mathbb{P}^{\boldsymbol{\beta}}$*-a.e.* $\omega \in \Omega$ *there exists a Lebesgue null set* $N^{\boldsymbol{\beta}}_\omega \subseteq [0, T-t_0]$ *such that for any* $s_0 \in [0, T-t_0] \setminus N^{\boldsymbol{\beta}}_\omega$ *we have*

$$\lim_{s\downarrow s_0} \frac{\mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\big[\log \overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s, \overline{\mathrm{P}}^{\boldsymbol{\beta}}_s) \,\big|\, \mathcal{G}_{s_0}\big](\omega) - \log \overline{\ell}^{\boldsymbol{\beta}}_{s_0}\big(\overline{X}_{s_0}(\omega), \overline{\mathrm{P}}^{\boldsymbol{\beta}}_{s_0}\big)}{s - s_0} = \overline{I}^{\boldsymbol{\beta}}_{s_0}(\omega). \tag{3.42}$$

Proof. The proof proceeds almost verbatim as the proof of Corollary 3.6. The only difference is that we now use the semimartingale decomposition (3.37) and the $\mathbb{P}^{\boldsymbol{\beta}}$-martingale property of the process (3.38) in Theorem 3.9. □

We now turn to the computation of the rate of change of the Wasserstein distance along the curve of probability distributions $(\mathrm{P}^{\boldsymbol{\beta}}_t)_{t_0\le t\le T}$. To this end, we set

$$v^{\boldsymbol{\beta}}_t(x) \coloneqq -\big(\nabla \log p^{\boldsymbol{\beta}}_t + \nabla V^{\boldsymbol{\beta}} + \nabla(W * p^{\boldsymbol{\beta}}_t)\big)(x), \qquad (t,x) \in [t_0,T]\times\mathbb{R}^n, \tag{3.43}$$

so that the perturbed granular media equation (3.25) can be viewed as a *continuity equation*

$$\partial_t p_t^{\boldsymbol{\beta}}(x) + \operatorname{div}\left(v_t^{\boldsymbol{\beta}}(x)\, p_t^{\boldsymbol{\beta}}(x)\right) = 0, \qquad (t,x) \in (t_0,T)\times\mathbb{R}^n, \tag{3.44}$$

with $v_t^{\boldsymbol{\beta}}(\cdot)$ as the corresponding *velocity field*. We recall the definition of the tangent space (see Definition 8.4.1 in [3])

$$\operatorname{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n) := \overline{\{\nabla\varphi\colon\, \varphi \in C_c^\infty(\mathbb{R}^n;\mathbb{R})\}}^{L^2(\mu)} \tag{3.45}$$

of $\mathscr{P}_2(\mathbb{R}^n)$ at the point $\mu \in \mathscr{P}_2(\mathbb{R}^n)$, and impose the following additional assumptions.

Assumptions 3.12. In addition to Assumptions 3.8, we suppose that

$$v_t(\cdot) \in \operatorname{Tan}_{\mathrm{P}_t} \mathscr{P}_2(\mathbb{R}^n) \qquad \text{for Lebesgue-a.e. } t \in [0,T], \tag{3.46}$$

where $v_t(\cdot)$ is obtained by taking $\boldsymbol{\beta} \equiv 0$ and $t_0 = 0$ in (3.43).

Remark 3.13. For example, we know from [3, Theorem 10.4.13] that the condition (3.46) is satisfied if, in addition to Assumptions 3.8, V is uniformly convex, i.e., $\operatorname{Hess}(V) \geq \kappa_V I_n$ for some real constant κ_V, and W is a convex function satisfying the *doubling condition*

$$\exists\, C_W > 0 \text{ such that } \forall\, x,y \in \mathbb{R}^n\colon \qquad W(x+y) \leq C_W\left(1 + W(x) + W(y)\right). \tag{3.47}$$

The proof of the following result is based on the general theory of Wasserstein metric derivatives of absolutely continuous curves in $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$; for a thorough discussion, we refer to Chapter 8 in [3].

Lemma 3.14. *Suppose Assumptions 3.12 hold. For Lebesgue-a.e. $t_0 \in [0,T]$, the Wasserstein metric derivative of the perturbed curve $(\mathrm{P}_t^{\boldsymbol{\beta}})_{t_0\leq t\leq T}$ is equal to*

$$\lim_{t\downarrow t_0} \frac{W_2\left(\mathrm{P}_t^{\boldsymbol{\beta}}, \mathrm{P}_{t_0}^{\boldsymbol{\beta}}\right)}{t - t_0} = \left\|v_{t_0}^{\boldsymbol{\beta}}(X_{t_0})\right\|_{L^2(\mathbb{P})} = \left\|\nabla \log \ell_{t_0}^{\uparrow}(X_{t_0},\mathrm{P}_{t_0}) + \nabla\boldsymbol{\beta}(X_{t_0})\right\|_{L^2(\mathbb{P})}. \tag{3.48}$$

Proof. Without loss of generality we can set $\boldsymbol{\beta} \equiv 0$. Note that from (3.7) we have

$$\int_0^T \int_{\mathbb{R}^n} |v_t(x)|^2 \,\mathrm{d}p_t(x)\,\mathrm{d}t = \mathbb{E}_{\mathbb{P}}\left[\int_0^T \left|\nabla \log \ell_t^{\uparrow}(X_t,\mathrm{P}_t)\right|^2 \mathrm{d}t\right] < \infty, \tag{3.49}$$

which implies that $v_t(\cdot) \in L^2(\mathrm{P}_t)$ for Lebesgue-a.e. $t \in [0,T]$. Therefore we can apply Theorem 8.3.1 and Proposition 8.4.5 of [3] to the absolutely continuous curve $(\mathrm{P}_t)_{0\leq t\leq T}$, which yields (3.48). □

We now have all the ingredients to formulate the gradient flow property of the granular media equation. The Wasserstein metric slope of the free energy functional $\mathscr{F}$ along the McKean–Vlasov curve $(\mathrm{P}_t)_{t_0\leq t\leq T}$ is defined as

$$\left|\partial\mathscr{F}\right|_{W_2}(\mathrm{P}_{t_0}) := \lim_{t\downarrow t_0} \frac{H(\mathrm{P}_t \,|\, \mathrm{Q}_t) - H(\mathrm{P}_{t_0} \,|\, \mathrm{Q}_{t_0})}{W_2(\mathrm{P}_t, \mathrm{P}_{t_0})}. \tag{3.50}$$

In order to show that this is the slope of steepest descent, we will compare it with the slope

$$\left|\partial\mathscr{F}^{\boldsymbol{\beta}}\right|_{W_2}\left(\mathrm{P}_{t_0}^{\boldsymbol{\beta}}\right) := \lim_{t\downarrow t_0} \frac{H\left(\mathrm{P}_t^{\boldsymbol{\beta}} \,\middle|\, \mathrm{Q}_t^{\boldsymbol{\beta}}\right) - H\left(\mathrm{P}_{t_0}^{\boldsymbol{\beta}} \,\middle|\, \mathrm{Q}_{t_0}^{\boldsymbol{\beta}}\right)}{W_2\left(\mathrm{P}_t^{\boldsymbol{\beta}}, \mathrm{P}_{t_0}^{\boldsymbol{\beta}}\right)} \tag{3.51}$$

along the perturbed curve $(\mathrm{P}_t^{\boldsymbol{\beta}})_{t_0 \le t \le T}$.

Theorem 3.15. *Suppose Assumptions 3.12 hold. Then the following assertions hold for Lebesgue-a.e. $t_0 \in [0,T]$: the random variables*

$$\mathcal{L}_{t_0}^{\uparrow} := \nabla \log \ell_{t_0}^{\uparrow}(X_{t_0}, \mathrm{P}_{t_0}) \qquad \text{and} \qquad \mathcal{B}_{t_0} := \nabla \boldsymbol{\beta}(X_{t_0}) \tag{3.52}$$

are elements of $L^2(\mathbb{P})$, and the Wasserstein metric slope of the free energy functional $\mathscr{F}$ along the McKean–Vlasov curve $(\mathrm{P}_t)_{t_0 \le t \le T}$ is given by

$$\big|\partial \mathscr{F}\big|_{W_2}(\mathrm{P}_{t_0}) = -\big\|\mathcal{L}_{t_0}^{\uparrow}\big\|_{L^2(\mathbb{P})}. \tag{3.53}$$

If $\mathcal{L}_{t_0}^{\uparrow} + \mathcal{B}_{t_0} \neq 0$, the metric slope along the perturbed curve $(\mathrm{P}_t^{\boldsymbol{\beta}})_{t_0 \le t \le T}$ is equal to

$$\big|\partial \mathscr{F}^{\boldsymbol{\beta}}\big|_{W_2}\big(\mathrm{P}_{t_0}^{\boldsymbol{\beta}}\big) = -\left\langle \mathcal{L}_{t_0}^{\uparrow}, \frac{\mathcal{L}_{t_0}^{\uparrow} + \mathcal{B}_{t_0}}{\big\|\mathcal{L}_{t_0}^{\uparrow} + \mathcal{B}_{t_0}\big\|_{L^2(\mathbb{P})}} \right\rangle_{L^2(\mathbb{P})}. \tag{3.54}$$

In particular,

$$\big|\partial \mathscr{F}\big|_{W_2}(\mathrm{P}_{t_0}) \le \big|\partial \mathscr{F}^{\boldsymbol{\beta}}\big|_{W_2}\big(\mathrm{P}_{t_0}^{\boldsymbol{\beta}}\big) \tag{3.55}$$

with equality if and only if $\mathcal{L}_{t_0}^{\uparrow} + \mathcal{B}_{t_0}$ is a positive multiple of $\mathcal{L}_{t_0}^{\uparrow}$.

Proof. The equality (3.53) follows from (3.18) and by taking $\boldsymbol{\beta} \equiv 0$ in (3.48). For the proof of (3.54), we first observe that from (3.41) and (3.48) we obtain the equality

$$\big|\partial \mathscr{F}^{\boldsymbol{\beta}}\big|_{W_2}\big(\mathrm{P}_{t_0}^{\boldsymbol{\beta}}\big) = -\frac{\big\|\mathcal{L}_{t_0}^{\uparrow}\big\|_{L^2(\mathbb{P})}^2 + \mathbb{E}_{\mathbb{P}}\Big[\Big(\Big\langle \nabla V + \nabla(W * \mathrm{P}_{t_0}), \nabla \boldsymbol{\beta} \Big\rangle - \Delta \boldsymbol{\beta}\Big)(X_{t_0})\Big]}{\big\|\mathcal{L}_{t_0}^{\uparrow} + \mathcal{B}_{t_0}\big\|_{L^2(\mathbb{P})}} \tag{3.56}$$

for Lebesgue-a.e. $t_0 \in [0,T]$. Integrating by parts and recalling the notations in (2.3) – (2.5), we find that the expectation in the numerator of (3.56) is equal to

$$\int_{\mathbb{R}^n} \Big\langle \log \nabla \ell_{t_0}^{\uparrow}(x), \nabla \boldsymbol{\beta}(x) \Big\rangle \, p_{t_0}(x) \, \mathrm{d}x = \Big\langle \mathcal{L}_{t_0}^{\uparrow}, \mathcal{B}_{t_0} \Big\rangle_{L^2(\mathbb{P})}. \tag{3.57}$$

Now (3.55) follows by the Cauchy–Schwarz inequality. □

3.3. A trajectorial proof of the HWBI inequality

In this subsection, we show how our trajectorial approach can be adapted to give a simple proof of the HWBI inequality. While the techniques that will be used are similar, the setting of this section is independent from the rest of the paper. In particular, we shall impose convexity assumptions on the potentials V, W.

We fix two probability measures ν_0 and ν_1 in $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$. By Brenier's theorem [9], there exists a convex function $\varphi \colon \mathbb{R}^n \to \mathbb{R}$ such that

$$W_2^2(\nu_0, \nu_1) = \int_{\mathbb{R}^n} |x - \nabla \varphi(x)|^2 \, \mathrm{d}\nu_0(x). \tag{3.58}$$

The displacement interpolation of McCann [46] between ν_0 and ν_1 is given by

$$\nu_t := (T_t)_{\#}\nu_0, \qquad T_t(x) := (1-t)x + t\nabla\varphi(x), \qquad 0 \le t \le 1. \tag{3.59}$$

In particular, since the endpoints ν_0 and ν_1 belong to $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$, each ν_t has a probability density function ρ_t; see, e.g., [58, Remarks 5.13 (i)].

As before, we consider a confinement potential V and an interaction potential W. For each $t \in [0,1]$, we then define by analogy with (2.6), the σ-finite measures

$$\mu_t(A) := \int_A q(x,\nu_t)\,\mathrm{d}x, \qquad \mu_t^{\uparrow}(A) := \int_A q^{\uparrow}(x,\nu_t)\,\mathrm{d}x, \qquad A \in \mathscr{B}(\mathbb{R}^n), \tag{3.60}$$

where we recall the definitions of the density functions q and $q^{\uparrow}$ in (2.4). In parallel to the likelihood ratio functions in (2.5), we define

$$r_t(x,\nu) := \frac{\rho_t(x)}{q(x,\nu)}, \qquad r_t^{\uparrow}(x,\nu) := \frac{\rho_t(x)}{q^{\uparrow}(x,\nu)}, \qquad (t,x,\nu) \in [0,1]\times\mathbb{R}^n\times\mathscr{P}_2(\mathbb{R}^n). \tag{3.61}$$

Then the relative entropy of ν_t with respect to μ_t is given by

$$H(\nu_t \,|\, \mu_t) = \int_{\mathbb{R}^n} \rho_t(x) \log r_t(x,\nu_t)\,\mathrm{d}x \tag{3.62}$$

and the relative Fisher information of ν_t with respect to $\mu_t^{\uparrow}$ is equal to

$$I\big(\nu_t \,|\, \mu_t^{\uparrow}\big) = \int_{\mathbb{R}^n} |\nabla \log r_t^{\uparrow}(x,\nu_t)|^2\, \rho_t(x)\,\mathrm{d}x. \tag{3.63}$$

We impose the following regularity conditions for Proposition 3.17, noting that the strong assumptions placed on ρ_0 and ρ_1 are only temporary and will be removed in Assumptions 3.18 of Theorem 3.19.

Assumptions 3.16. The functions $V, W\colon \mathbb{R}^n \to [0,\infty)$ are smooth and W is symmetric. The probability density functions ρ_0 and ρ_1 are smooth, compactly supported and strictly positive in the interior of their respective supports.

Proposition 3.17. *Suppose Assumptions 3.16 hold. Along the displacement interpolation $(\nu_t)_{0\le t\le 1}$, the rate of relative entropy dissipation at time $t = 0$, with respect to the "reference curve of probability measures" $(\mu_t)_{0\le t\le 1}$, is given by*

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}^{+} H(\nu_t \,|\, \mu_t) = \int_{\mathbb{R}^n} \Big\langle \nabla \log r_0^{\uparrow}(x,\nu_0), \nabla\varphi(x) - x \Big\rangle\, \rho_0(x)\,\mathrm{d}x. \tag{3.64}$$

Combining Proposition 3.17 with the displacement convexity results of McCann [46], we obtain the following generalization of the HWBI inequality. Equivalent versions of this inequality can be found in [18, Theorem 4.1] and [25, Theorem D.50].

Assumptions 3.18. The functions $V, W\colon \mathbb{R}^n \to [0,\infty)$ are smooth and W is symmetric. Furthermore, V and W are uniformly convex, i.e., there exist real constants κ_V and κ_W such that

$$\mathrm{Hess}(V) \ge \kappa_V I_n, \qquad \mathrm{Hess}(W) \ge \kappa_W I_n. \tag{3.65}$$

Theorem 3.19. *Suppose Assumptions 3.18 hold and the relative entropy $H(\nu_1 \,|\, \mu_1)$ is finite. Then*

$$H(\nu_0 \,|\, \mu_0) - H(\nu_1 \,|\, \mu_1) \le -\int_{\mathbb{R}^n} \left\langle \nabla \log r_0^{\uparrow}(x, \nu_0), \nabla\varphi(x) - x \right\rangle \rho_0(x)\,\mathrm{d}x \tag{3.66}$$

$$- \tfrac{\kappa_V + \kappa_W}{2}\, W_2^2(\nu_0, \nu_1) + \tfrac{\kappa_W}{2}\, |b(\nu_0) - b(\nu_1)|^2. \tag{3.67}$$

Remark 3.20. By the Cauchy–Schwarz inequality, the right-hand side of (3.66) can be bounded from above by

$$\sqrt{\int_{\mathbb{R}^n} \left|\nabla \log r_0^{\uparrow}(x, \nu_0)\right|^2 \rho_0(x)\,\mathrm{d}x} \sqrt{\int_{\mathbb{R}^n} |\nabla\varphi(x) - x|^2 \rho_0(x)\,\mathrm{d}x} = \sqrt{I(\nu_0 \,|\, \mu_0^{\uparrow})}\, W_2(\nu_0, \nu_1), \tag{3.68}$$

and we obtain the usual form of the HWBI inequality (1.16); see also [2, Theorem 4.2].

4. Proofs of the main results

This section is devoted to the proofs of the results stated in Section 3. We shall first prove the main trajectorial results: Theorem 3.1 and its "perturbed" counterpart, Theorem 3.9.

4.1. The proofs of Theorem 3.1 and Theorem 3.9

Since Theorem 3.1 follows immediately from Theorem 3.9 by setting the perturbation $\boldsymbol{\beta}\colon \mathbb{R}^n \to \mathbb{R}$ to be the zero function, we start with the general setting of Theorem 3.9. We first recall a classical result concerning the time reversal of diffusions.

Lemma 4.1 ([32, Theorem 2.1], [39, Theorems G.2, G.5]). *Suppose Assumptions 3.8 hold. On the probability space $(\Omega, \mathbb{G}, \mathbb{P}^{\boldsymbol{\beta}})$, the process*

$$\overline{B}_s^{\boldsymbol{\beta}} := B_{T-s}^{\boldsymbol{\beta}} - B_T^{\boldsymbol{\beta}} - \sqrt{2} \int_0^s \nabla \log \overline{p}_u^{\boldsymbol{\beta}}(\overline{X}_u^{\boldsymbol{\beta}})\,\mathrm{d}u, \qquad 0 \le s \le T - t_0 \tag{4.1}$$

is a Brownian motion. Moreover, the time-reversed canonical process $(\overline{X}_s)_{0 \le s \le T - t_0}$ satisfies

$$\mathrm{d}\overline{X}_s = \left(2\nabla \log \overline{\ell}_s^{\boldsymbol{\beta}\downarrow} - \nabla V + \nabla(W * \overline{\mathrm{P}}_s^{\boldsymbol{\beta}}) + \nabla\boldsymbol{\beta}\right)(\overline{X}_s)\,\mathrm{d}s + \sqrt{2}\,\mathrm{d}\overline{B}_s^{\boldsymbol{\beta}}. \tag{4.2}$$

By means of Lemma 4.1, the first step in the proof of Theorem 3.9 is to compute the dynamics of the time-reversed perturbed relative entropy process (3.36). For the reader's convenience, we recall the following characterization of the L-derivative in [10, pp. 383].

Definition 4.2. Let $f\colon \mathscr{P}_2(\mathbb{R}^n) \to \mathbb{R}$ and $\mu_0 \in \mathscr{P}_2(\mathbb{R}^n)$. On a probability space $(\Omega, \mathbb{F}, \mathbb{P})$, let X_0 be a random variable with distribution μ_0. We define $\partial_\mu f(\mu_0)\colon \mathbb{R}^n \to \mathbb{R}^n$ as the L-*derivative* of f at μ_0, if for any $\mu \in \mathscr{P}_2(\mathbb{R}^n)$ and any random variable X with distribution μ,

$$f(\mu) = f(\mu_0) + \mathbb{E}_{\mathbb{P}}\Big[\big\langle \partial_\mu f(\mu_0)(X_0),\, X - X_0 \big\rangle\Big] + o\big(\|X - X_0\|_{L^2(\mathbb{P})}\big).$$

Remark 4.3. The above characterization of the L-derivative depends neither on the choice of the probability space $(\Omega,\mathbb{F},\mathbb{P})$, nor of the random variables X and X_0 used to represent μ and μ_0, respectively. Moreover, if the L-derivative exists, it is uniquely defined up to μ_0-equivalence. We refer to Proposition 5.25 and Remark 5.26 in [10] for the details.

Proposition 4.4. *Suppose Assumptions 3.8 hold. On $(\Omega,\mathbb{G},\mathbb{P}^{\boldsymbol{\beta}})$, the time-reversed perturbed relative entropy process (3.36) satisfies*

$$\mathrm{d}\log\overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s,\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)=\Big\langle\nabla\log\overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s,\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s),\sqrt{2}\,\mathrm{d}\overline{B}^{\boldsymbol{\beta}}_s\Big\rangle+\Big(\big|\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s\big|^2+\tfrac{1}{2}\big|\nabla(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)\big|^2\Big)(\overline{X}_s)\,\mathrm{d}s \tag{4.3}$$

$$+\Big(\Big\langle\tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s),2\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s+\nabla V^{\boldsymbol{\beta}}\Big\rangle+\langle\nabla V,\nabla\boldsymbol{\beta}\rangle-\Delta\boldsymbol{\beta}\Big)(\overline{X}_s)\,\mathrm{d}s \tag{4.4}$$

$$-\mathbb{E}_{\tilde{\mathbb{P}}^{\boldsymbol{\beta}}}\Big[\Big\langle\tfrac{1}{2}\nabla W(\overline{X}_s-\overline{Y}_s),\Big(2\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s-\nabla V+\nabla(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)+\nabla\boldsymbol{\beta}\Big)(\overline{Y}_s)\Big\rangle\Big]\,\mathrm{d}s, \tag{4.5}$$

where $(\overline{Y}_s)_{0\le s\le T-t_0}$ is a copy of the process $(\overline{X}_s)_{0\le s\le T-t_0}$ on a copy $(\tilde{\Omega},\tilde{\mathbb{G}},\tilde{\mathbb{P}}^{\boldsymbol{\beta}})$ of the original probability space $(\Omega,\mathbb{G},\mathbb{P}^{\boldsymbol{\beta}})$.

Proof. Applying a generalized version of Itô's formula for McKean–Vlasov diffusions [10, Proposition 5.102] and using the backward dynamics in (4.2), we obtain

$$\mathrm{d}\log\overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s,\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)=\Big\langle\nabla\log\overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s,\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s),\sqrt{2}\,\mathrm{d}\overline{B}^{\boldsymbol{\beta}}_s\Big\rangle+\Big(\partial_s\log\overline{\ell}^{\boldsymbol{\beta}}_s+\Delta\log\overline{\ell}^{\boldsymbol{\beta}}_s\Big)(\overline{X}_s,\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)\,\mathrm{d}s \tag{4.6}$$

$$+\Big\langle\nabla\log\overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s,\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s),\Big(2\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s-\nabla V+\nabla(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)+\nabla\boldsymbol{\beta}\Big)(\overline{X}_s)\Big\rangle\,\mathrm{d}s \tag{4.7}$$

$$+\mathbb{E}_{\tilde{\mathbb{P}}^{\boldsymbol{\beta}}}\Big[\Big\langle\Big(\partial_\mu\log\overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s,\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)\Big),\Big(2\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s-\nabla V+\nabla(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)+\nabla\boldsymbol{\beta}\Big)\Big\rangle(\overline{Y}_s)\Big]\,\mathrm{d}s \tag{4.8}$$

$$+\mathbb{E}_{\tilde{\mathbb{P}}^{\boldsymbol{\beta}}}\Big[\mathrm{trace}\Big(\partial_y\partial_\mu\log\overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s,\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)(\overline{Y}_s)\Big)\Big]\,\mathrm{d}s, \tag{4.9}$$

where $(\overline{Y}_s)_{0\le s\le T-t_0}$ is a copy of the process $(\overline{X}_s)_{0\le s\le T-t_0}$ on a copy $(\tilde{\Omega},\tilde{\mathbb{G}},\tilde{\mathbb{P}}^{\boldsymbol{\beta}})$ of the original probability space $(\Omega,\mathbb{G},\mathbb{P}^{\boldsymbol{\beta}})$. The L-derivative appearing in (4.8) and (4.9) is calculated to be

$$\Big(\partial_\mu\log\overline{\ell}^{\boldsymbol{\beta}}_s(x,\mu)\Big)(y)=\tfrac{1}{2}\big(\partial_\mu(W*\mu)(x)\big)(y)=-\tfrac{1}{2}\nabla W(x-y) \tag{4.10}$$

for $(x,\mu,y)\in\mathbb{R}^n\times\mathscr{P}_2(\mathbb{R}^n)\times\mathbb{R}^n$, see [10, Section 5.2.2, Example 1] for the computation of the L-derivative of a function which is linear in the distribution variable. Consequently, we have

$$\mathrm{trace}\Big(\partial_y\partial_\mu\log\overline{\ell}^{\boldsymbol{\beta}}_s(x,\mu)(y)\Big)=-\tfrac{1}{2}\mathrm{trace}\big(\partial_y\nabla W(x-y)\big)=\tfrac{1}{2}\Delta W(x-y). \tag{4.11}$$

Putting (4.10) and (4.11) into (4.8) and (4.9), respectively, as well as using the identities

$$\partial_s\log\overline{\ell}^{\boldsymbol{\beta}}_s(x,\mu)=\partial_s\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s(x), \tag{4.12}$$

$$\nabla\log\overline{\ell}^{\boldsymbol{\beta}}_s(x,\mu)=\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s(x)+\tfrac{1}{2}\nabla(W*\mu)(x), \tag{4.13}$$

$$\Delta\log\overline{\ell}^{\boldsymbol{\beta}}_s(x,\mu)=\Delta\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s(x)+\tfrac{1}{2}\Delta(W*\mu)(x), \tag{4.14}$$

we obtain

$$\mathrm{d}\log\overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s,\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)=\Big\langle\nabla\log\overline{\ell}^{\boldsymbol{\beta}}_s(\overline{X}_s,\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s),\sqrt{2}\,\mathrm{d}\overline{B}^{\boldsymbol{\beta}}_s\Big\rangle+\Big(\partial_s\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s+\Delta\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s\Big)(\overline{X}_s)\,\mathrm{d}s \tag{4.15}$$

$$+\Big\langle\Big(\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s+\tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)\Big),\Big(2\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s-\nabla V+\nabla(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)+\nabla\boldsymbol{\beta}\Big)\Big\rangle(\overline{X}_s)\,\mathrm{d}s \tag{4.16}$$

$$-\mathbb{E}_{\bar{\mathbb{P}}^{\boldsymbol{\beta}}}\Big[\Big\langle\tfrac{1}{2}\nabla W(\overline{X}_s-\overline{Y}_s),\Big(2\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s-\nabla V+\nabla(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)+\nabla\boldsymbol{\beta}\Big)(\overline{Y}_s)\Big\rangle\Big]\,\mathrm{d}s \tag{4.17}$$

$$+\tfrac{1}{2}\Big(\Delta(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)(\overline{X}_s)+\mathbb{E}_{\bar{\mathbb{P}}^{\boldsymbol{\beta}}}\Big[\Delta W(\overline{X}_s-\overline{Y}_s)\Big]\Big)\,\mathrm{d}s. \tag{4.18}$$

Regarding the expression of (4.18), we observe that $\Delta(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)(\overline{X}_s)=\mathbb{E}_{\bar{\mathbb{P}}^{\boldsymbol{\beta}}}[\Delta W(\overline{X}_s-\overline{Y}_s)]$. Finally, elementary computations based on (2.4), (3.25) and (3.27) show that the perturbed log-likelihood ratio function $(s,x)\mapsto\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s(x)$ of (3.27) satisfies

$$\begin{aligned}\partial_s\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s&=\Big\langle\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s,\nabla V-\nabla(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)-\nabla\boldsymbol{\beta}\Big\rangle-\big|\nabla\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s\big|^2-\Delta\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_s\\&\quad+\Big\langle\nabla V,\nabla(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)+\nabla\boldsymbol{\beta}\Big\rangle-\Delta(W*\overline{\mathrm{P}}^{\boldsymbol{\beta}}_s)-\Delta\boldsymbol{\beta}\end{aligned} \tag{4.19}$$

on $(0,T-t_0)\times\mathbb{R}^n$, with terminal condition $\log\overline{\ell}^{\boldsymbol{\beta}\downarrow}_{T-t_0}=\log\overline{\ell}^{\downarrow}_{T-t_0}$. Inserting (4.19) into (4.15), we obtain (4.3) – (4.5). □

Setting the perturbation $\boldsymbol{\beta}$ to be the zero function, we obtain the following result.

Corollary 4.5. *Suppose Assumptions 2.1 hold. On $(\Omega,\mathbb{G},\mathbb{P})$, the time-reversed relative entropy process (3.4) satisfies*

$$\mathrm{d}\log\overline{\ell}_s(\overline{X}_s,\overline{\mathrm{P}}_s)=\Big\langle\nabla\log\overline{\ell}_s(\overline{X}_s,\overline{\mathrm{P}}_s),\sqrt{2}\,\mathrm{d}\overline{B}_s\Big\rangle \tag{4.20}$$

$$+\Big(\big|\nabla\log\overline{\ell}^{\downarrow}_s\big|^2+\tfrac{1}{2}\big|\nabla(W*\overline{\mathrm{P}}_s)\big|^2+\Big\langle\tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}_s),2\nabla\log\overline{\ell}^{\downarrow}_s+\nabla V\Big\rangle\Big)(\overline{X}_s)\,\mathrm{d}s \tag{4.21}$$

$$-\mathbb{E}_{\bar{\mathbb{P}}}\Big[\Big\langle\tfrac{1}{2}\nabla W(\overline{X}_s-\overline{Y}_s),\Big(2\nabla\log\overline{\ell}^{\downarrow}_s-\nabla V+\nabla(W*\overline{\mathrm{P}}_s)\Big)(\overline{Y}_s)\Big\rangle\Big]\,\mathrm{d}s. \tag{4.22}$$

Here, the process

$$\overline{B}_s:=B_{T-s}-B_T-\sqrt{2}\int_0^s\nabla\log\overline{p}_u(\overline{X}_u)\,\mathrm{d}u,\qquad 0\le s\le T \tag{4.23}$$

is a $\mathbb{P}$-Brownian motion with respect to the backward filtration $\mathbb{G}$, and $(\overline{Y}_s)_{0\le s\le T}$ is a copy of the process $(\overline{X}_s)_{0\le s\le T}$ on a copy $(\tilde{\Omega},\tilde{\mathbb{G}},\tilde{\mathbb{P}})$ of the original probability space $(\Omega,\mathbb{G},\mathbb{P})$.

Before turning to the final part of the proof of Theorem 3.1, we state a classical result based on the general theory of the Cameron–Martin–Maruyama–Girsanov transformation [44]. The connection between relative entropy (the left-hand side of (4.25) below) and energy (the right-hand side of (4.25)) is the foundation of Föllmer's entropy approach to the time reversal of diffusion processes on Wiener space [26–28]. We denote by $\mathbb{W}_x$ the Wiener measure on $\Omega=C([0,T];\mathbb{R}^n)$ with starting point $x\in\mathbb{R}^n$, and define by

$$\mathbb{W}_{x,2}(A):=\mathbb{W}_x\Big(\omega\in\Omega\colon(\sqrt{2}X)(\omega)\in A\Big),\qquad A\in\mathscr{B}(\Omega) \tag{4.24}$$

the Wiener measure with starting point x and variance 2.

Lemma 4.6. *The relative entropy of $\mathbb{P}$ with respect to $\mathbb{W}_{\mathrm{P}_0,2} := \int_{\mathbb{R}^n} \mathbb{W}_{x,2}\,\mathrm{d}\mathrm{P}_0(x)$ is given by*

$$H\big(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}\big) = \mathbb{E}_{\mathbb{P}}\bigg[\int_0^T \big|\nabla V(X_t) + \nabla(W * \mathrm{P}_t)(X_t)\big|^2\,\mathrm{d}t\bigg] < \infty. \tag{4.25}$$

Proof. Recalling (2.3), the drift of the McKean–Vlasov dynamics (1.1) can be expressed as

$$-\nabla\Psi^{\uparrow}(x,\mathrm{P}_t) = -\big(\nabla V(x) + \nabla(W * \mathrm{P}_t)(x)\big)\,, \qquad (t,x) \in [0,T]\times\mathbb{R}^n. \tag{4.26}$$

For any $t \in [0,T]$, using the elementary inequality $(a+b)^2 \le 2a^2 + 2b^2$, we have

$$\mathbb{E}_{\mathbb{P}}\Big[|\nabla\Psi^{\uparrow}(X_t,\mathrm{P}_t)|^2\Big] \le 2\,\mathbb{E}_{\mathbb{P}}\Big[|\nabla V(X_t)|^2\Big] + 2\,\mathbb{E}_{\mathbb{P}}\Big[|\nabla(W*\mathrm{P}_t)(X_t)|^2\Big]. \tag{4.27}$$

Using the linear growth condition (2.1) from Assumptions 2.1 (i), we find

$$\mathbb{E}_{\mathbb{P}}\Big[|\nabla V(X_t)|^2\Big] \le 2C^2\Big(1 + \mathbb{E}_{\mathbb{P}}\big[|X_t|^2\big]\Big) \le 2C^2\Big(1 + \mathbb{E}_{\mathbb{P}}\Big[\sup_{0\le t\le T}|X_t|^2\Big]\Big). \tag{4.28}$$

Similarly, by Jensen's inequality and (2.1), we obtain

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\Big[|\nabla(W*\mathrm{P}_t)(X_t)|^2\Big] &\le \int_{\mathbb{R}^n\times\mathbb{R}^n} |\nabla W(x-y)|^2\,p_t(y)\,p_t(x)\,\mathrm{d}y\,\mathrm{d}x && (4.29)\\
&\le 2C^2\bigg(1 + \int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2\,p_t(y)\,p_t(x)\,\mathrm{d}y\,\mathrm{d}x\bigg) && (4.30)\\
&\le 2C^2\bigg(1 + 2\int_{\mathbb{R}^n\times\mathbb{R}^n} (|x|^2+|y|^2)\,p_t(y)\,p_t(x)\,\mathrm{d}y\,\mathrm{d}x\bigg) && (4.31)\\
&= 2C^2\Big(1 + 4\,\mathbb{E}_{\mathbb{P}}\big[|X_t|^2\big]\Big) \le 8C^2\Big(1 + \mathbb{E}_{\mathbb{P}}\Big[\sup_{0\le t\le T}|X_t|^2\Big]\Big). && (4.32)
\end{aligned}$$

Altogether, we get

$$\mathbb{E}_{\mathbb{P}}\bigg[\int_0^T |\nabla\Psi^{\uparrow}(X_t,\mathrm{P}_t)|^2\,\mathrm{d}t\bigg] \le 20T\Big(1 + \mathbb{E}_{\mathbb{P}}\Big[\sup_{0\le t\le T}|X_t|^2\Big]\Big) < \infty, \tag{4.33}$$

where the finiteness of this expression follows from the uniform second moment property (2.2) of Lemma 2.2. From [44, Section 7.6.4] we now conclude that $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{W}_{\mathrm{P}_0,2}$, and the Radon–Nikodym derivatives are given by

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}_{\mathrm{P}_0,2}}\bigg|_{\mathcal{F}_t} = \exp\bigg(-2\int_0^t \big\langle\nabla\Psi^{\uparrow}(X_u,\mathrm{P}_u),\,\sqrt{2}\,\mathrm{d}B_u\big\rangle + \int_0^t |\nabla\Psi^{\uparrow}(X_u,\mathrm{P}_u)|^2\,\mathrm{d}u\bigg), \qquad 0\le t\le T. \tag{4.34}$$

The integrability property (4.33) implies that the $\mathbb{P}$-expectation of the stochastic integral in (4.34) vanishes, and we obtain

$$H\big(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}\big) = \mathbb{E}_{\mathbb{P}}\bigg[\log\bigg(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}_{\mathrm{P}_0,2}}\bigg)\bigg] = \mathbb{E}_{\mathbb{P}}\bigg[\int_0^T |\nabla\Psi^{\uparrow}(X_t,\mathrm{P}_t)|^2\,\mathrm{d}t\bigg] < \infty, \tag{4.35}$$

which shows (4.25). □

Denoting n-dimensional Lebesgue measure by λ, we consider on $\Omega = C([0,T];\mathbb{R}^n)$ the σ-finite measure $\mathbb{W}_{\lambda,2} \coloneqq \int_{\mathbb{R}^n} \mathbb{W}_{x,2}\,\mathrm{d}\lambda(x)$, which is known as the law of the *reversible Brownian motion* on $\mathbb{R}^n$ with variance 2; see [19,42]. The fundamental property of reversible Brownian motion is that it is invariant under time reversal. This property can be formalized as follows. Let $R\colon \Omega \to \Omega$ be the pathwise time reversal operator on Ω, given by $X_s \circ R = X_{T-s}$ for $s \in [0,T]$. For any measure μ on Ω, we denote its time reversal by $\overline{\mu} \coloneqq R_{\#}\mu$. Then we have the invariance property $\overline{\mathbb{W}}_{\lambda,2} = \mathbb{W}_{\lambda,2}$. Let us also consider the probability measure $\mathbb{W}_{\mathrm{P}_T,2} \coloneqq \int_{\mathbb{R}^n} \mathbb{W}_{x,2}\,\mathrm{dP}_T(x)$ and its time reversal given by $\overline{\mathbb{W}}_{\mathrm{P}_T,2} = \int_{\mathbb{R}^n} \overline{\mathbb{W}}_{x,2}\,\mathrm{dP}_T(x)$. Then, as already noted in [26, Remarks 3.7], we have the following result.

Lemma 4.7. *We have the relative entropy relations*

$$H\big(\mathbb{P}\,|\,\mathbb{W}_{\lambda,2}\big) = H\big(\mathrm{P}_0\,|\,\lambda\big) + H\big(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}\big) \tag{4.36}$$

and

$$H\big(\mathbb{P}\,|\,\overline{\mathbb{W}}_{\lambda,2}\big) = H\big(\mathrm{P}_T\,|\,\lambda\big) + H\big(\mathbb{P}\,|\,\overline{\mathbb{W}}_{\mathrm{P}_T,2}\big). \tag{4.37}$$

Furthermore, all these relative entropies are finite.

Proof. For any $x \in \mathbb{R}^n$, we let $\mathbb{P}_x(\,\cdot\,) \coloneqq \mathbb{P}(\,\cdot\,|\,X_0 = x)$ denote (a version of) the conditional probability measure $\mathbb{P}$ given $X_0 = x$. By the chain rule for relative entropy [19, Theorem 2] we have

$$H\big(\mathbb{P}\,|\,\mathbb{W}_{\lambda,2}\big) = H\big(\mathrm{P}_0\,|\,\lambda\big) + \int_{\mathbb{R}^n} H\big(\mathbb{P}_x\,|\,\mathbb{W}_{x,2}\big)\,\mathrm{dP}_0(x) \tag{4.38}$$

and at the same time

$$H\big(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}\big) = H\big(\mathrm{P}_0\,|\,\mathrm{P}_0\big) + \int_{\mathbb{R}^n} H\big(\mathbb{P}_x\,|\,\mathbb{W}_{x,2}\big)\,\mathrm{dP}_0(x) = \int_{\mathbb{R}^n} H\big(\mathbb{P}_x\,|\,\mathbb{W}_{x,2}\big)\,\mathrm{dP}_0(x), \tag{4.39}$$

implying the first identity (4.36). Regarding the finite entropy assertions, we recall (1.2), (1.3) and observe that

$$H\big(\mathrm{P}_0\,|\,\lambda\big) = \int_{\mathbb{R}^n} p_0(x)\log p_0(x)\,\mathrm{d}x = \mathcal{U}(\mathrm{P}_0) \le \mathscr{F}(\mathrm{P}_0) < \infty, \tag{4.40}$$

where the finiteness follows from Assumptions 2.1 (ii). Furthermore, from Lemma 4.6 we know that $H(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}) < \infty$.

By the same arguments as above, (4.37) follows again by the chain rule for relative entropy. Using the invariance property $\overline{\mathbb{W}}_{\lambda,2} = \mathbb{W}_{\lambda,2}$, and as we already know that $H(\mathbb{P}\,|\,\mathbb{W}_{\lambda,2}) < \infty$, it follows that $H(\mathbb{P}\,|\,\overline{\mathbb{W}}_{\lambda,2}) < \infty$. Let us recall now that $\mathrm{P}_T \in \mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$ by Lemma 2.2. On the one hand, since P_T has finite second moment, $H(\mathrm{P}_T\,|\,\lambda)$ cannot take the value $-\infty$ as noted in Remark 2.3. On the other hand, the absolute continuity of P_T implies that $H(\mathrm{P}_T\,|\,\lambda)$ cannot take the value $+\infty$. Therefore, we conclude that $H(\mathbb{P}\,|\,\overline{\mathbb{W}}_{\mathrm{P}_T,2}) < \infty$. □

We have assembled now all the ingredients needed for the proof of Theorem 3.1.

Proof of Theorem 3.1. Recalling the definition of the stochastic integral process $(\overline{M}_s)_{0\le s\le T}$ in (3.6) and of the cumulative Fisher information process $(\overline{F}_s)_{0\le s\le T}$ in (3.3), we see that the stochastic differential of (4.20) – (4.22) can be expressed as claimed in (3.5).

Since $(\overline{M}_s)_{0\le s\le T}$ is a stochastic integral process, it is a continuous local martingale. In order to show that it is an $L^2(\mathbb{P})$-bounded martingale, it suffices to show the integrability condition

$$\mathbb{E}_{\mathbb{P}}\Big[\big\langle \overline{M},\overline{M}\big\rangle_T\Big] = \mathbb{E}_{\mathbb{P}}\bigg[2\int_0^T \big|\nabla \log \overline{\ell}_u(\overline{X}_u,\overline{\mathrm{P}}_u)\big|^2\,\mathrm{d}u\bigg] < \infty; \tag{4.41}$$

see, e.g. [51, Corollary IV.1.25]. On $(\Omega,\mathbb{G},\mathbb{P})$, the time-reversed canonical process $(\overline{X}_s)_{0\le s\le T}$ has backward dynamics

$$\mathrm{d}\overline{X}_s = \overline{\vartheta}_s(\overline{X}_s)\,\mathrm{d}s + \sqrt{2}\,\mathrm{d}\overline{B}_s\,, \qquad 0\le s\le T \tag{4.42}$$

with initial distribution P_T, where the drift term is given by

$$\overline{\vartheta}_s(x) \coloneqq \Big(2\nabla \log \overline{\ell}_s^{\downarrow} - \nabla V + \nabla(W*\overline{\mathrm{P}}_s)\Big)(x) = 2\nabla \log \overline{\ell}_s(x,\overline{\mathrm{P}}_s) - \nabla V(x) \in \mathbb{R}^n \tag{4.43}$$

for $(s,x)\in[0,T]\times\mathbb{R}^n$. Therefore, in order to prove (4.41), it suffices to show the two integrability conditions

$$\mathbb{E}_{\mathbb{P}}\bigg[\int_0^T |\nabla V(\overline{X}_u)|^2\,\mathrm{d}u\bigg] < \infty \qquad \text{and} \qquad \mathbb{E}_{\mathbb{P}}\bigg[\int_0^T |\overline{\vartheta}_u(\overline{X}_u)|^2\,\mathrm{d}u\bigg] < \infty. \tag{4.44}$$

The first condition is a direct consequence of (4.28). From [26, Lemma 2.6] we conclude that the expectation of the second condition is bounded by the relative entropy $H(\mathbb{P}\,|\,\overline{\mathbb{W}}_{\mathrm{P}_T,2})$, which is finite on account of Lemma 4.7.

In order to complete the proof of Theorem 3.1, it remains to show (3.7). To begin with, we take expectation with respect to $\mathbb{P}$ in (3.3) and invoke Fubini's theorem to interchange the $\mathbb{P}$-expectation and the time integral. Applying once more Fubini's theorem, we swap the $\mathbb{P}$-expectation with the $\tilde{\mathbb{P}}$-expectation appearing in (3.2). Next, we recall Assumptions 2.1 (i) and use the symmetry of the interaction potential, which implies that $\nabla W(-x) = -\nabla W(x)$ for all $x\in\mathbb{R}^n$. Furthermore, as the distribution of $\overline{Y}_u$ under $\tilde{\mathbb{P}}$ is the same as the distribution of $\overline{X}_u$ under $\mathbb{P}$, we deduce that

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s\big] = \int_0^s \mathbb{E}_{\mathbb{P}}\Big[\Big(\big|\nabla \log \overline{\ell}_u^{\downarrow}\big|^2 + \tfrac{1}{2}\big|\nabla(W*\overline{\mathrm{P}}_u)\big|^2 + \Big\langle \tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}_u),\, 2\nabla \log \overline{\ell}_u^{\downarrow} + \nabla V\Big\rangle\Big)(\overline{X}_u)\Big]\,\mathrm{d}u \tag{4.45}$$

$$+ \int_0^s \mathbb{E}_{\mathbb{P}}\Big[\Big\langle \tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}_u),\, \Big(2\nabla \log \overline{\ell}_u^{\downarrow} - \nabla V + \nabla(W*\overline{\mathrm{P}}_u)\Big)(\overline{X}_u)\Big\rangle\Big]\,\mathrm{d}u \tag{4.46}$$

for $0\le s\le T$. Recalling the definitions in (2.3) – (2.5), we obtain

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s\big] = \int_0^s \mathbb{E}_{\mathbb{P}}\Big[\big|\nabla \log \overline{\ell}_u^{\uparrow}(\overline{X}_u,\overline{\mathrm{P}}_u)\big|^2\Big]\,\mathrm{d}u = \int_0^s I\big(\overline{\mathrm{P}}_u\,\big|\,\overline{\mathrm{Q}}_u^{\uparrow}\big)\,\mathrm{d}u < \infty\,, \qquad 0\le s\le T, \tag{4.47}$$

where the second equality is immediate from (2.7), and the finiteness of the expression in (4.47) is justified as follows. Again, from (2.3) – (2.5) we find

$$\big|\nabla \log \overline{\ell}_s^{\uparrow}(\overline{X}_s,\overline{\mathrm{P}}_s)\big|^2 = \big|\nabla \log \overline{\ell}_s(\overline{X}_s,\overline{\mathrm{P}}_s) + \tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}_s)(\overline{X}_s)\big|^2 \tag{4.48}$$

$$\le 2\big|\nabla \log \overline{\ell}_s(\overline{X}_s,\overline{\mathrm{P}}_s)\big|^2 + \tfrac{1}{2}\big|\nabla(W*\overline{\mathrm{P}}_s)(\overline{X}_s)\big|^2. \tag{4.49}$$

In light of (4.41) and (4.29) – (4.32), we see that the expression in (4.47) is finite, which in turn justifies a posteriori the former applications of Fubini's theorem. □

Remark 4.8. In the above proofs of Lemmas 4.6, 4.7 and Theorem 3.1, the linear growth condition (2.1) from Assumptions 2.1 (i) was crucial. For a different proof in the (linear) setting without interaction (i.e., $W \equiv 0$), we refer to [40], where the confinement potential V (which is called Ψ in [40]) is not necessarily of linear growth, but instead satisfies a weaker coercivity condition. The key in that setting is to apply Lemma 2.48 in [37].

The proof of Theorem 3.9 is now an easy consequence.

Proof of Theorem 3.9. Recalling the definition of the process $(\overline{M}_s^{\boldsymbol{\beta}})_{0\le s\le T}$ in (3.38) and of the perturbed cumulative Fisher information process $(\overline{F}_s^{\boldsymbol{\beta}})_{0\le s\le T}$ in (3.35), we see that the stochastic differential of (4.3) – (4.5) can be expressed as claimed in (3.37).

As in the proof of Theorem 3.1 we will now argue that

$$\mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\Big[\big\langle \overline{M}^{\boldsymbol{\beta}}, \overline{M}^{\boldsymbol{\beta}}\big\rangle_{T-t_0}\Big] = \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\bigg[2\int_0^{T-t_0} \big|\nabla \log \overline{\ell}_u^{\boldsymbol{\beta}}(\overline{X}_u, \overline{\mathrm{P}}_u^{\boldsymbol{\beta}})\big|^2 \,\mathrm{d}u\bigg] < \infty, \tag{4.50}$$

which will then imply that the stochastic integral process $(\overline{M}_s^{\boldsymbol{\beta}})_{0\le s\le T-t_0}$ is an $L^2(\mathbb{P}^{\boldsymbol{\beta}})$-bounded martingale. To this end, we define the density $q^{\boldsymbol{\beta}}(x,\mu) := \mathrm{e}^{-\Psi^{\boldsymbol{\beta}}(x,\mu)}$ for $(x,\mu) \in \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n)$, and consider the "doubly perturbed" likelihood ratio function

$$\ell_t^{\boldsymbol{\beta},\boldsymbol{\beta}}(x,\mu) := \frac{p_t^{\boldsymbol{\beta}}(x)}{q^{\boldsymbol{\beta}}(x,\mu)}, \qquad (t,x) \in [t_0,T]\times\mathbb{R}^n. \tag{4.51}$$

As the Assumptions 2.1 are invariant under the passage from the potential V to $V^{\boldsymbol{\beta}} = V + \boldsymbol{\beta}$, we can apply Theorem 3.1 to the potential $V^{\boldsymbol{\beta}}$ and obtain

$$\mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\bigg[2\int_0^{T-t_0} \big|\nabla \log \overline{\ell}_u^{\boldsymbol{\beta},\boldsymbol{\beta}}(\overline{X}_u, \overline{\mathrm{P}}_u^{\boldsymbol{\beta}})\big|^2 \,\mathrm{d}u\bigg] < \infty. \tag{4.52}$$

Now, since $\ell_t^{\boldsymbol{\beta}}(x,\mu)/\ell_t^{\boldsymbol{\beta},\boldsymbol{\beta}}(x,\mu) = \mathrm{e}^{\boldsymbol{\beta}(x)}$, we observe that the difference

$$\nabla \log \ell_t^{\boldsymbol{\beta}}(x,\mu) - \nabla \log \ell_t^{\boldsymbol{\beta},\boldsymbol{\beta}}(x,\mu) = \nabla\boldsymbol{\beta}(x) \tag{4.53}$$

is a bounded function. Together with (4.52), this implies (4.50).

It remains to check (3.39). A similar calculation as in the proof of Theorem 3.1 leads to the identity

$$\mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\big[\overline{F}_s^{\boldsymbol{\beta}}\big] = \int_0^s \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\Big[\big|\nabla \log \overline{\ell}_u^{\boldsymbol{\beta}\uparrow}(\overline{X}_u, \overline{\mathrm{P}}_u^{\boldsymbol{\beta}})\big|^2 + \Big(\Big\langle \nabla V + \nabla(W * \overline{\mathrm{P}}_u^{\boldsymbol{\beta}}), \nabla\boldsymbol{\beta}\Big\rangle - \Delta\boldsymbol{\beta}\Big)(\overline{X}_u)\Big]\,\mathrm{d}u \tag{4.54}$$

for $0 \le s \le T - t_0$. Repeating the reasoning of the previous paragraph for the function $\ell_t^{\boldsymbol{\beta}\uparrow}$ instead of $\ell_t^{\boldsymbol{\beta}}$, we find that

$$\mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\bigg[\int_0^{T-t_0} \big|\nabla \log \overline{\ell}_u^{\boldsymbol{\beta}\uparrow}(\overline{X}_u^{\boldsymbol{\beta}}, \overline{\mathrm{P}}_u^{\boldsymbol{\beta}})\big|^2 \,\mathrm{d}u\bigg] < \infty. \tag{4.55}$$

Since the function

$$[0,T-t_0]\times\mathbb{R}^n \ni (t,x) \longmapsto \Big\langle \nabla V + \nabla(W * \overline{\mathrm{P}}_t^{\boldsymbol{\beta}}), \nabla\boldsymbol{\beta}\Big\rangle(x) - \Delta\boldsymbol{\beta}(x) \tag{4.56}$$

is bounded, we conclude that the quantity of (4.54) is finite. Finally, recalling the definition (3.30), we arrive at (3.39). □

4.2. The proofs of Proposition 3.17 and Theorem 3.19

Proof of Proposition 3.17. The first step is to view the probability density functions $(\rho_t)_{0\le t\le 1}$, corresponding to the displacement interpolation $(\nu_t)_{0\le t\le 1}$ of (3.59), as a solution to a continuity equation. Recalling the convex function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ of (3.58), we define a function $\mathsf{u}_0\colon \mathbb{R}^n \to \mathbb{R}$ by $\mathsf{u}_0(x) := \varphi(x) - |x|^2/2$; and for each $t \in (0,1]$, we let the function $\mathsf{u}_t : \mathbb{R}^n \to \mathbb{R}$ be defined by the Hopf–Lax formula

$$\mathsf{u}_t(x) := \inf_{y\in\mathbb{R}^n} \Big(\mathsf{u}_0(y) + \frac{|x-y|^2}{2t}\Big). \tag{4.57}$$

For all $t \in [0,1)$, we denote the gradient of u_t by $\mathsf{v}_t := \nabla \mathsf{u}_t$. For $t = 0$, it is clear that $\mathsf{v}_0 = \nabla\varphi - \mathrm{Id}$ is well-defined. For $t \in (0,1)$, the gradient v_t is defined Lebesgue-a.e. by [58, Theorem 5.51 (i)], and

$$\mathsf{v}_t(x) = \nabla \mathsf{u}_0 \circ (T_t)^{-1}(x), \qquad \text{for all } x \in T_t(\mathbb{R}^n), \tag{4.58}$$

where T_t is defined in (3.59). Note that the inverse of T_t is well-defined because T_t is injective; see [58, Section 5.4.8]. From (4.58) we see that $(\mathsf{v}_t)_{0\le t<1}$ is the *velocity field* associated with the trajectories $(T_t)_{0\le t<1}$, i.e.,

$$T_t(x) = x + \int_0^t \mathsf{v}_s(T_s(x))\,\mathrm{d}s, \qquad 0 \le t < 1. \tag{4.59}$$

By [58, Theorem 5.51 (ii)], the curve of probability density functions $(\rho_t)_{0<t<1}$ satisfies the continuity equation

$$\partial_t \rho_t(x) + \mathrm{div}\big(\rho_t(x)\,\mathsf{v}_t(x)\big) = 0, \qquad (t,x) \in (0,1)\times\mathbb{R}^n. \tag{4.60}$$

On a sufficiently rich probability space $(S,\mathcal{S},\mathsf{P})$, we let $Z_0\colon S \to \mathbb{R}^n$ be a random variable with probability distribution ν_0. For each $0 < t \le 1$, we let $Z_t := T_t(Z_0)$. From (3.59) we see that the random variable Z_t has distribution ν_t, and (4.59) yields the representation

$$Z_t = Z_0 + \int_0^t \mathsf{v}_s(Z_s)\,\mathrm{d}s, \qquad 0 \le t < 1. \tag{4.61}$$

In conjunction with (4.60), we deduce

$$\mathrm{d}\rho_t(Z_t) = \partial_t\rho_t(Z_t) + \big\langle \nabla\rho_t(Z_t), \mathrm{d}Z_t\big\rangle = -\rho_t(Z_t)\,\mathrm{div}\big(\mathsf{v}_t(Z_t)\big)\,\mathrm{d}t, \tag{4.62}$$

and thus

$$\mathrm{d}\log\rho_t(Z_t) = -\mathrm{div}\big(\mathsf{v}_t(Z_t)\big)\,\mathrm{d}t. \tag{4.63}$$

Recalling the definition of the density function q in (2.4), a similar argument as in (4.10) shows that

$$\big(\partial_\nu \log q(x,\nu)\big)(y) = \tfrac{1}{2}\big(\partial_\nu (W*\nu)(x)\big)(y) = -\tfrac{1}{2}\nabla W(x-y) \tag{4.64}$$

for $(x,\nu,y) \in \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n)\times\mathbb{R}^n$. Applying a generalized version of Itô's formula for McKean–Vlasov diffusions [10, Proposition 5.102], and using the dynamics (4.61) as well as the L-derivative (4.64), we obtain

$$\mathrm{d}\log q(Z_t,\nu_t) = -\Big\langle \nabla V + \tfrac{1}{2}\nabla(W*\nu_t), \mathsf{v}_t\Big\rangle(Z_t)\,\mathrm{d}t + \tfrac{1}{2}\mathbb{E}_{\tilde{\mathsf{P}}}\Big[\Big\langle \nabla W(Z_t - \tilde{Z}_t), \mathsf{v}_t(\tilde{Z}_t)\Big\rangle\Big]\,\mathrm{d}t \tag{4.65}$$

for $0 < t < 1$. Here, the process $(\tilde{Z}_t)_{0<t<1}$ is defined on another probability space $(\tilde{S}, \tilde{\mathcal{S}}, \tilde{\mathsf{P}})$ such that the tuple $(S, \mathcal{S}, \mathsf{P}, (Z_t)_{0<t<1})$ is an exact copy of $(\tilde{S}, \tilde{\mathcal{S}}, \tilde{\mathsf{P}}, (\tilde{Z}_t)_{0<t<1})$. Now taking the difference between (4.63) and (4.65) gives the dynamics

$$\log r_t(Z_t, \nu_t) - \log r_0(Z_0, \nu_0) = \int_0^t \Big(\Big\langle \nabla V + \tfrac{1}{2} \nabla (W * \nu_s), \mathsf{v}_s \Big\rangle (Z_s) - \operatorname{div}\big(\mathsf{v}_s(Z_s)\big) \Big) \, \mathrm{d}s \tag{4.66}$$

$$- \tfrac{1}{2} \int_0^t \mathbb{E}_{\tilde{\mathsf{P}}} \Big[\Big\langle \nabla W(Z_s - \tilde{Z}_s), \mathsf{v}_s(\tilde{Z}_s) \Big\rangle \Big] \, \mathrm{d}s \tag{4.67}$$

of the relative entropy process $(\log r_t(Z_t, \nu_t))_{0<t<1}$. Next, let us make two observations. Firstly, integration by parts yields

$$\mathbb{E}_{\mathsf{P}} \Big[\operatorname{div}\big(\mathsf{v}_t(Z_t)\big) \Big] = -\mathbb{E}_{\mathsf{P}} \Big[\Big\langle \nabla \log \rho_t(Z_t), \mathsf{v}_t(Z_t) \Big\rangle \Big] . \tag{4.68}$$

Secondly, by applying Fubini's theorem, and using that W is an even function as well as $(\tilde{Z}_t)_{\#} \tilde{\mathsf{P}} = \nu_t$, we obtain the identity

$$\mathbb{E}_{\mathsf{P}} \Big[\mathbb{E}_{\tilde{\mathsf{P}}} \Big[\Big\langle \nabla W(Z_t - \tilde{Z}_t), \mathsf{v}_t(\tilde{Z}_t) \Big\rangle \Big] \Big] = -\mathbb{E}_{\mathsf{P}} \Big[\Big\langle \nabla (W * \nu_t)(Z_t), \mathsf{v}_t(Z_t) \Big\rangle \Big] . \tag{4.69}$$

Returning to (4.66), (4.67), we take P-expectations and use (4.68), (4.69) to obtain

$$H(\nu_t \,|\, \mu_t) - H(\nu_0 \,|\, \mu_0) = \int_0^t \mathbb{E}_{\mathsf{P}} \Big[\Big\langle \nabla \log \rho_s + \nabla V + \nabla (W * \nu_s), \mathsf{v}_s \Big\rangle (Z_s) \Big] \, \mathrm{d}s \tag{4.70}$$

$$= \int_0^t \mathbb{E}_{\mathsf{P}} \Big[\Big\langle \nabla \log r_s^{\uparrow}(Z_s, \nu_s), \mathsf{v}_s(Z_s) \Big\rangle \Big] \, \mathrm{d}s, \tag{4.71}$$

where for the second equality we recall the notations in (3.61) and (2.4). Finally, letting $t \downarrow 0$, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}^{+} H(\nu_t \,|\, \mu_t) = \int_{\mathbb{R}^n} \Big\langle \nabla \log r_0^{\uparrow}(x, \nu_0), \mathsf{v}_0(x) \Big\rangle \rho_0(x) \, \mathrm{d}x; \tag{4.72}$$

and since $\mathsf{v}_0 = \nabla \varphi - \mathrm{Id}$, we arrive at (3.64). □

Proof of Theorem 3.19. Without loss of generality, we assume that the probability density functions ρ_0 and ρ_1 satisfy the strong regularity Assumptions 3.16. The general case then follows by a density argument. We will not provide the details here, but refer to [58, Chapter 9.4], where this regularization is carried out in the simpler setting of the HWI inequality.

Let us recall the energy functionals $\mathcal{U}, \mathcal{V}, \mathcal{W}$ defined in (1.3), and introduce the functions

$$f(t) \coloneqq \mathcal{U}(\rho_t), \qquad g(t) \coloneqq \mathcal{V}(\rho_t), \qquad h(t) \coloneqq \mathcal{W}(\rho_t), \qquad 0 \le t \le 1, \tag{4.73}$$

where $(\rho_t)_{0 \le t \le 1}$ is the curve of probability density functions corresponding to the displacement interpolation $(\nu_t)_{0 \le t \le 1}$ of (3.59). Then the sum $F \coloneqq f + g + h$ of these functions satisfies the relation $F(t) = H(\nu_t \,|\, \mu_t)$. In light of [58, Theorem 5.15 (i)], the internal energy functional $\mathcal{U}$ is displacement convex, i.e.,

$$f''(t) \ge 0, \qquad 0 \le t \le 1. \tag{4.74}$$

By Assumptions 3.18, the confinement potential $V\colon \mathbb{R}^n \to [0,\infty)$ is κ_V-uniformly convex. Therefore, [58, Theorem 5.15 (ii)] implies that the potential energy functional $\mathcal{V}$ is κ_V-uniformly displacement convex. In other words,

$$g''(t) \geq \kappa_V\, W_2^2(\nu_0,\nu_1), \qquad 0 \leq t \leq 1. \tag{4.75}$$

Again from Assumptions 3.18, the interaction potential $W\colon \mathbb{R}^n \to [0,\infty)$ is assumed to be symmetric and κ_W-uniformly convex. Therefore, a similar argument as in the proof of [58, Theorem 5.15 (iii)] leads to the $\kappa_W(W_2^2(\nu_0,\nu_1) - |b(\nu_0) - b(\nu_1)|^2)$-uniform convexity of h, so

$$h''(t) \geq \kappa_W \Big(W_2^2(\nu_0,\nu_1) - |b(\nu_0) - b(\nu_1)|^2 \Big), \qquad 0 \leq t \leq 1. \tag{4.76}$$

The details of the proof of (4.76) are postponed to Subsection 5.2. By combining the estimates (4.74) – (4.76), we deduce that the relative entropy function $[0,1] \ni t \mapsto F(t) = H(\nu_t \,|\, \mu_t)$ satisfies

$$F''(t) \geq (\kappa_V + \kappa_W)\, W_2^2(\nu_0,\nu_1) - \kappa_W\, |b(\nu_0) - b(\nu_1)|^2. \tag{4.77}$$

Furthermore, from Proposition 3.17 we have

$$F'(0^+) = \int_{\mathbb{R}^n} \Big\langle \nabla \log r_0^{\uparrow}(x,\nu_0), \nabla\varphi(x) - x \Big\rangle\, \rho_0(x)\,\mathrm{d}x. \tag{4.78}$$

In conjunction with (4.77) and (4.78), the Taylor formula $F(1) = F(0) + F'(0^+) + \int_0^1 (1-t)F''(t)\,\mathrm{d}t$ now yields the inequality (3.66) – (3.67). □

5. Proofs of auxiliary results

5.1. Proof of Lemma 2.2

The generalized potential $\Psi^{\uparrow}$ of (2.3) allows us to cast the McKean–Vlasov dynamics of (1.1) in the more compact form

$$\mathrm{d}X_t = -\nabla\Psi^{\uparrow}(X_t, \mathrm{P}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \qquad 0 \leq t \leq T. \tag{5.1}$$

Then, for any two pairs $(x,\mu),(x',\mu') \in \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n)$, using the Lipschitz continuity of ∇V in Assumptions 2.1 (i) yields

$$|\nabla\Psi^{\uparrow}(x,\mu) - \nabla\Psi^{\uparrow}(x',\mu')| \leq \|\nabla V\|_{\mathrm{Lip}}\, |x - x'| + |\nabla(W * \mu)(x) - \nabla(W * \mu')(x')|. \tag{5.2}$$

For the convolution term, using Jensen's inequality and the Lipschitz continuity of ∇W in Assumptions 2.1 (i) leads to

$$|\nabla(W * \mu)(x) - \nabla(W * \mu')(x')| \leq \|\nabla W\|_{\mathrm{Lip}}\, |x - x'| + |\nabla(W * \mu)(x') - \nabla(W * \mu')(x')|. \tag{5.3}$$

For the last term above, by the Kantorovich–Rubinstein theorem [58, Theorem 1.14], we have

$$\begin{aligned} \left| \int_{\mathbb{R}^n} \nabla W(x' - \cdot)\,\mathrm{d}(\mu - \mu') \right| &\leq \|\nabla W\|_{\mathrm{Lip}}\, \sup \left\{ \int_{\mathbb{R}^n} \varphi\,\mathrm{d}(\mu - \mu') \colon \varphi \in L^1(|\mu - \mu'|), \|\varphi\|_{\mathrm{Lip}} \leq 1 \right\} \\ &= \|\nabla W\|_{\mathrm{Lip}}\, W_1(\mu,\mu') \leq \|\nabla W\|_{\mathrm{Lip}}\, W_2(\mu,\mu'), \end{aligned} \tag{5.4}$$

where

$$W_1(\mu,\mu') = \inf_{Y\sim\mu, Z\sim\mu'} \mathbb{E}|Y-Z|, \qquad \mu,\mu' \in \mathscr{P}_1(\mathbb{R}^n) \tag{5.5}$$

denotes the 1-Wasserstein-distance, and the inequality in (5.4) follows from Jensen's inequality. Altogether, we obtain

$$|\nabla\Psi^{\uparrow}(x,\mu) - \nabla\Psi^{\uparrow}(x',\mu')| \le (\|\nabla V\|_{\mathrm{Lip}} + \|\nabla W\|_{\mathrm{Lip}})\,|x-x'| + \|\nabla W\|_{\mathrm{Lip}}\, W_2(\mu,\mu'). \tag{5.6}$$

In particular, this shows that the function $-\nabla\Psi^{\uparrow}$ is Lipschitz continuous on the product metric space $(\mathbb{R}^n,|\cdot|)\times(\mathscr{P}_2(\mathbb{R}^n),W_2)$. In conjunction with Assumptions 2.1 (ii), [10, Theorem 4.21] implies that the McKean–Vlasov SDE (5.1) has a pathwise unique, strong solution satisfying the uniform second moment condition (2.2). Now we can linearize (5.1) by fixing the time-marginals $(\mathrm{P}_t)_{0\le t\le T}$, so that the drift term can be viewed as a function $(t,x)\mapsto\nabla\Psi^{\uparrow}(x,\mathrm{P}_t)$, and (5.1) becomes an ordinary SDE with a time-inhomogeneous drift coefficient.

The absolute continuity of the time-marginals $(\mathrm{P}_t)_{0\le t\le T}$ is immediate from Lemma 4.6. A standard argument using the classical Itô's formula shows that the curve of probability density functions $(p_t)_{0\le t\le T}$ is a weak solution of the granular media equation (1.11). Finally, we turn to the regularity of this solution. From (5.6), we see that the drift $x\mapsto\nabla\Psi^{\uparrow}(x,\mathrm{P}_t)$ is Lipschitz continuous for every $t\in[0,T]$, and Assumptions 2.1 (i) implies that the drift is also of linear growth. The desired smoothness of $(p_t)_{0\le t\le T}$ now follows from a straightforward adaptation of the theorem in [53], see also Remarks (i) – (ii) therein. □

5.2. Proof of (4.76)

We first rewrite the interaction energy functional $\mathcal{W}$ along the displacement interpolation $(\nu_t)_{0\le t\le 1}$. Using (3.59), for any $t\in[0,1]$, we have

$$h(t) = \tfrac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} W(x-y)\,\nu_t(\mathrm{d}x)\,\nu_t(\mathrm{d}y) \tag{5.7}$$

$$= \tfrac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} W\big(T_t(x)-T_t(y)\big)\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y) \tag{5.8}$$

$$= \tfrac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} W\Big(x-y-t\big(\theta(x)-\theta(y)\big)\Big)\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y), \tag{5.9}$$

where $\theta\colon\mathbb{R}^n\to\mathbb{R}^n$ is defined as $\theta(x) := x-\nabla\varphi(x)$. Now, for any $t_1,t_2,\sigma\in[0,1]$, by the κ_W-uniform convexity of W in Assumptions 3.18, we obtain

$$\sigma h(t_1) + (1-\sigma)h(t_2) - h\big(\sigma t_1 + (1-\sigma)t_2\big) \tag{5.10}$$

$$= \tfrac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} \Big(\sigma W\Big(x-y-t_1\big(\theta(x)-\theta(y)\big)\Big) + (1-\sigma)W\Big(x-y-t_2\big(\theta(x)-\theta(y)\big)\Big) \tag{5.11}$$

$$- W\Big(x-y-\big(\sigma t_1+(1-\sigma)t_2\big)\big(\theta(x)-\theta(y)\big)\Big)\Big)\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y) \tag{5.12}$$

$$\ge \tfrac{1}{4}\kappa_W\sigma(1-\sigma)(t_1-t_2)^2\int_{\mathbb{R}^n\times\mathbb{R}^n} |\theta(x)-\theta(y)|^2\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y). \tag{5.13}$$

Next, we express the integral in (5.13) as

$$\begin{aligned}
&\tfrac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} |\theta(x)-\theta(y)|^2\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y)\\
&\qquad = \int_{\mathbb{R}^n} |\theta(x)|^2\,\nu_0(\mathrm{d}x) - \left|\int_{\mathbb{R}^n} \theta(x)\,\nu_0(\mathrm{d}x)\right|^2\\
&\qquad = \int_{\mathbb{R}^n} |x-\nabla\varphi(x)|^2\,\nu_0(\mathrm{d}x) - \left|\int_{\mathbb{R}^n} x\,\nu_0(\mathrm{d}x) - \int_{\mathbb{R}^n} x\,\nu_1(\mathrm{d}x)\right|^2\\
&\qquad = W_2^2(\nu_0,\nu_1) - |b(\nu_0)-b(\nu_1)|^2.
\end{aligned}$$

Putting this back into (5.13), we deduce that h is uniformly convex, with constant

$$\kappa_W\left(W_2^2(\nu_0,\nu_1) - |b(\nu_0)-b(\nu_1)|^2\right). \tag{5.14}$$

□

Acknowledgments

We are grateful to Ioannis Karatzas and Walter Schachermayer for suggesting this problem and giving us generous advice. We thank Robert Fernholz, Miguel Garrido, Tomoyuki Ichiba, Donghan Kim, Kasper Larsen, and Mete Soner for helpful comments during the INTECH research meetings. Thanks also go to an associate editor and an anonymous referee for their valuable comments and suggestions.

Funding

B. Tschiderer acknowledges support by the Austrian Science Fund (FWF) under grant P28661, by the Vienna Science and Technology Fund (WWTF) through project MA16-021, and additionally appreciates travel support through the National Science Foundation (NSF) under grant NSF-DMS-14-05210. L.C. Yeung acknowledges support under grant NSF-DMS-20-04997.

References

[1] Acciaio, B., Beiglböck, M., Penkner, F., Schachermayer, W. and Temme, J. (2013). A trajectorial interpretation of Doob's martingale inequalities. *Ann. Appl. Probab.* **23** 1494–1505. MR3098440 https://doi.org/10.1214/12-aap878

[2] Agueh, M., Ghoussoub, N. and Kang, X. (2004). Geometric inequalities via a general comparison principle for interacting gases. *Geom. Funct. Anal.* **14** 215–244. MR2053603 https://doi.org/10.1007/s00039-004-0455-x

[3] Ambrosio, L., Gigli, N. and Savaré, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Basel: Birkhäuser. MR2401600

[4] Backhoff, J., Conforti, G., Gentil, I. and Léonard, C. (2020). The mean field Schrödinger problem: Ergodic behavior, entropy estimates and functional inequalities. *Probab. Theory Related Fields* **178** 475–530. MR4146543 https://doi.org/10.1007/s00440-020-00977-8

[5] Beiglböck, M. and Siorpaes, P. (2015). Pathwise versions of the Burkholder-Davis-Gundy inequality. *Bernoulli* **21** 360–373. MR3322322 https://doi.org/10.3150/13-BEJ570

[6] Benachour, S., Roynette, B., Talay, D. and Vallois, P. (1998). Nonlinear self-stabilizing processes. I. Existence, invariant probability, propagation of chaos. *Stochastic Process. Appl.* **75** 173–201. MR1632193 https://doi.org/10.1016/S0304-4149(98)00018-0

[7] Benedetto, D., Caglioti, E., Carrillo, J.A. and Pulvirenti, M. (1998). A non-Maxwellian steady distribution for one-dimensional granular media. *J. Stat. Phys.* **91** 979–990. MR1637274 https://doi.org/10.1023/A:1023032000560

[8] Bolley, F., Gentil, I. and Guillin, A. (2013). Uniform convergence to equilibrium for granular media. *Arch. Ration. Mech. Anal.* **208** 429–445. MR3035983 https://doi.org/10.1007/s00205-012-0599-z

[9] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809 https://doi.org/10.1002/cpa.3160440402

[10] Carmona, R. and Delarue, F. (2018). *Probabilistic Theory of Mean Field Games with Applications. I. Mean Field FBSDEs, Control, and Games*. *Probability Theory and Stochastic Modelling* **83**. Cham: Springer. MR3752669

[11] Carrillo, J.A., Gvalani, R.S., Pavliotis, G.A. and Schlichting, A. (2020). Long-time behaviour and phase transitions for the McKean-Vlasov equation on the torus. *Arch. Ration. Mech. Anal.* **235** 635–690. MR4062483 https://doi.org/10.1007/s00205-019-01430-4

[12] Carrillo, J.A., Jüngel, A., Markowich, P.A., Toscani, G. and Unterreiter, A. (2001). Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities. *Monatsh. Math.* **133** 1–82. MR1853037 https://doi.org/10.1007/s006050170032

[13] Carrillo, J.A., McCann, R.J. and Villani, C. (2003). Kinetic equilibration rates for granular media and related equations: Entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoam.* **19** 971–1018. MR2053570 https://doi.org/10.4171/RMI/376

[14] Carrillo, J.A., McCann, R.J. and Villani, C. (2006). Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.* **179** 217–263. MR2209130 https://doi.org/10.1007/s00205-005-0386-1

[15] Cattiaux, P. (2004). A pathwise approach of some classical inequalities. *Potential Anal.* **20** 361–394. MR2032116 https://doi.org/10.1023/B:POTA.0000009847.84908.6f

[16] Cattiaux, P. and Guillin, A. (2014). Semi log-concave Markov diffusions. In *Séminaire de Probabilités XLVI*. *Lecture Notes in Math.* **2123** 231–292. Cham: Springer. MR3330820 https://doi.org/10.1007/978-3-319-11970-0_9

[17] Cattiaux, P., Guillin, A. and Malrieu, F. (2008). Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Related Fields* **140** 19–40. MR2357669 https://doi.org/10.1007/s00440-007-0056-3

[18] Cordero-Erausquin, D., Gangbo, W. and Houdré, C. (2004). Inequalities for generalized entropy and optimal transportation. In *Recent Advances in the Theory and Applications of Mass Transport*. *Contemp. Math.* **353** 73–94. Providence, RI: Amer. Math. Soc. MR2079071 https://doi.org/10.1090/conm/353/06433

[19] Léonard, C. (2014). Some Properties of Path Measures. In *Sémin. Probab. XLVI*, (C. Donati-Martin, A. Lejay and A. Rouault, eds.). *Lecture Notes in Math.* **2123** 207–230. Switzerland: Springer International Publishing.

[20] Duong, M.H. and Tugaut, J. (2016). Stationary solutions of the Vlasov-Fokker-Planck equation: Existence, characterization and phase-transition. *Appl. Math. Lett.* **52** 38–45. MR3416384 https://doi.org/10.1016/j.aml.2015.08.003

[21] Duong, M.H. and Tugaut, J. (2018). The Vlasov-Fokker-Planck equation in non-convex landscapes: Convergence to equilibrium. *Electron. Commun. Probab.* **23** Paper No. 19, 10. MR3779816 https://doi.org/10.1214/18-ECP116

[22] Eldan, R., Lehec, J. and Shenfeld, Y. (2020). Stability of the logarithmic Sobolev inequality via the Föllmer process. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 2253–2269. MR4116725 https://doi.org/10.1214/19-AIHP1038

[23] Eldan, R. and Mikulincer, D. (2020). Stability of the Shannon-Stam inequality via the Föllmer process. *Probab. Theory Related Fields* **177** 891–922. MR4126934 https://doi.org/10.1007/s00440-020-00967-w

[24] Erbar, M., Fathi, M., Laschos, V. and Schlichting, A. (2016). Gradient flow structure for McKean-Vlasov equations on discrete spaces. *Discrete Contin. Dyn. Syst.* **36** 6799–6833. MR3567821 https://doi.org/10.3934/dcds.2016096

[25] Feng, J. and Kurtz, T.G. (2006). *Large Deviations for Stochastic Processes*. *Mathematical Surveys and Monographs* **131**. Providence, RI: Amer. Math. Soc. MR2260560 https://doi.org/10.1090/surv/131
[26] Föllmer, H. (1985). An entropy approach to the time reversal of diffusion processes. In *Stochastic Differential Systems (Marseille-Luminy, 1984)*. *Lect. Notes Control Inf. Sci.* **69** 156–163. Berlin: Springer. MR0798318 https://doi.org/10.1007/BFb0005070
[27] Föllmer, H. (1986). Time reversal on Wiener space. In *Stochastic Processes—Mathematics and Physics (Bielefeld, 1984)*. *Lecture Notes in Math.* **1158** 119–129. Berlin: Springer. MR0838561 https://doi.org/10.1007/BFb0080212
[28] Föllmer, H. (1988). Random fields and diffusion processes. In *École D'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*. *Lecture Notes in Math.* **1362** 101–203. Berlin: Springer. MR0983373 https://doi.org/10.1007/BFb0086180
[29] Fontbona, J. and Jourdain, B. (2016). A trajectorial interpretation of the dissipations of entropy and Fisher information for stochastic differential equations. *Ann. Probab.* **44** 131–170. MR3456334 https://doi.org/10.1214/14-AOP969
[30] Gardiner, C. (2009). *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed. *Springer Series in Synergetics*. Berlin: Springer. MR2676235
[31] Gentil, I., Léonard, C., Ripani, L. and Tamanini, L. (2020). An entropic interpolation proof of the HWI inequality. *Stochastic Process. Appl.* **130** 907–923. MR4046525 https://doi.org/10.1016/j.spa.2019.04.002
[32] Haussmann, U.G. and Pardoux, É. (1986). Time reversal of diffusions. *Ann. Probab.* **14** 1188–1205. MR0866342
[33] Herrmann, S. and Tugaut, J. (2010). Non-uniqueness of stationary measures for self-stabilizing processes. *Stochastic Process. Appl.* **120** 1215–1246. MR2639745 https://doi.org/10.1016/j.spa.2010.03.009
[34] Hu, K., Ren, Z., Šiška, D. and Szpruch, Ł. (2021). Mean-field Langevin dynamics and energy landscape of neural networks. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 2043–2065. MR4328560 https://doi.org/10.1214/20-aihp1140
[35] Jordan, R. and Kinderlehrer, D. (1996). An extended variational principle. In *Partial Differential Equations and Applications*. *Lecture Notes in Pure and Applied Mathematics* **177** 187–200. New York: Dekker. MR1371591 https://doi.org/10.5006/1.3292113
[36] Jordan, R., Kinderlehrer, D. and Otto, F. (1998). The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29** 1–17. MR1617171 https://doi.org/10.1137/S0036141096303359
[37] Karatzas, I. and Kardaras, C. (2021). *Portfolio Theory and Arbitrage: A Course in Mathematical Finance*. *Grad. Stud. Math.* **214**. Providence, Rhode Island: Amer. Math. Soc.
[38] Karatzas, I., Maas, J. and Schachermayer, W. (2021). Trajectorial dissipation and gradient flow for the relative entropy in Markov chains. *Commun. Inf. Syst.* **21** 481–536. MR4273512 https://doi.org/10.4310/CIS.2021.v21.n4.a1
[39] Karatzas, I., Schachermayer, W. and Tschiderer, B. (2020). Trajectorial Otto calculus. arXiv:1811.08686.
[40] Karatzas, I., Schachermayer, W. and Tschiderer, B. (2021). A trajectorial approach to the gradient flow properties of Langevin-Smoluchowski diffusions. *Teor. Veroyatn. Primen.* **66** 839–888. MR4331222 https://doi.org/10.4213/tvp55
[41] Lehec, J. (2013). Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 885–899. MR3112438 https://doi.org/10.1214/11-aihp464
[42] Léonard, C. (2014). A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** 1533–1574. MR3121631 https://doi.org/10.3934/dcds.2014.34.1533
[43] Lions, P.L. (2007–2008). Théorie des jeux de champ moyen et applications. Lectures at the Collège de France. https://www.college-de-france.fr/site/pierre-louis-lions/course-2007-2008.htm.
[44] Liptser, R.S. and Shiryaev, A.N. (2001). *Statistics of Random Processes. I: General Theory*, expanded ed. *Applications of Mathematics (New York)* **5**. Berlin: Springer. MR1800857
[45] Malrieu, F. (2003). Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.* **13** 540–560. MR1970276 https://doi.org/10.1214/aoap/1050689593
[46] McCann, R.J. (1997). A convexity principle for interacting gases. *Adv. Math.* **128** 153–179. MR1451422 https://doi.org/10.1006/aima.1997.1634
[47] McKean, H.P. Jr. (1966). A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* **56** 1907–1911. MR0221595 https://doi.org/10.1073/pnas.56.6.1907

[48] Mei, S., Montanari, A. and Nguyen, P.-M. (2018). A mean field view of the landscape of two-layer neural networks. *Proc. Natl. Acad. Sci. USA* **115** E7665–E7671. MR3845070 https://doi.org/10.1073/pnas.1806579115
[49] Otto, F. (2001). The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations* **26** 101–174. MR1842429 https://doi.org/10.1081/PDE-100002243
[50] Otto, F. and Villani, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400. MR1760620 https://doi.org/10.1006/jfan.1999.3557
[51] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **293**. Berlin: Springer. MR1725357 https://doi.org/10.1007/978-3-662-06400-9
[52] Risken, H. (1984). *The Fokker-Planck Equation: Methods of Solution and Applications*. *Springer Series in Synergetics* **18**. Berlin: Springer. MR0749386 https://doi.org/10.1007/978-3-642-96807-5
[53] Rogers, L.C.G. (1985). Smooth transition densities for one-dimensional diffusions. *Bull. Lond. Math. Soc.* **17** 157–161. MR0806242 https://doi.org/10.1112/blms/17.2.157
[54] Tugaut, J. (2013). Convergence to the equilibria for self-stabilizing processes in double-well landscape. *Ann. Probab.* **41** 1427–1460. MR3098681 https://doi.org/10.1214/12-AOP749
[55] Tugaut, J. (2014). Phase transitions of McKean-Vlasov processes in double-wells landscape. *Stochastics* **86** 257–284. MR3180036 https://doi.org/10.1080/17442508.2013.775287
[56] Tugaut, J. (2014). Self-stabilizing processes in multi-wells landscape in $\mathbb{R}^d$-invariant probabilities. *J. Theoret. Probab.* **27** 57–79. MR3174216 https://doi.org/10.1007/s10959-012-0435-2
[57] Veretennikov, A.Yu. (2006). On ergodic measures for McKean-Vlasov stochastic equations. In *Monte Carlo and Quasi-Monte Carlo Methods 2004* 471–486. Berlin: Springer. MR2208726 https://doi.org/10.1007/3-540-31186-6_29
[58] Villani, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Providence, RI: Amer. Math. Soc. MR1964483 https://doi.org/10.1090/gsm/058
[59] Villani, C. (2006). Mathematics of granular materials. *J. Stat. Phys.* **124** 781–822. MR2264625 https://doi.org/10.1007/s10955-006-9038-6

Received July 2021 and revised February 2022