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

Preprint ISSN 2198-5855

From large deviations to Wasserstein gradient flows in multiple dimensions

Matthias Erbar[1], Jan Maas[2], D.R. Michiel Renger[3]

submitted: May 22, 2015

[1] University of Bonn
Institute for Applied Mathematics
Endenicher Allee 60
53115 Bonn
Germany
E-Mail: erbar@iam.uni-bonn.de

[2] Institute of Science and Technology Austria
(IST Austria)
Am Campus 1
3400 Klosterneuburg
Austria
E-Mail: jan.maas@ist.ac.at

[3] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: michiel.renger@wias-berlin.de

No. 2110
Berlin 2015

2010 *Mathematics Subject Classification.* 35A15, 35Q84, 46N55, 60F10.

Key words and phrases. Large deviations, gradient flows, Wasserstein calculus, Gamma convergence.

This work has been initiated when MR visited ME and JM at the University of Bonn. The authors gratefully acknowledge support by the German Research Foundation through the Hausdorff Center for Mathematics at the University of Bonn, and the Collaborative Research Centers 1060 “The Mathematics of Emergent Effects” and 1114 “Scaling Cascades in Complex Systems”.

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

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

ABSTRACT. We study the large deviation rate functional for the empirical measure of independent Brownian particles with drift. In one dimension, it has been shown by Adams, Dirr, Peletier and Zimmer [ADPZ11] that this functional is asymptotically equivalent (in the sense of Γ-convergence) to the Jordan–Kinderlehrer–Otto functional arising in the Wasserstein gradient flow structure of the Fokker–Planck equation. In higher dimensions, part of this statement (the lower bound) has been recently proved by Duong, Laschos and Renger, but the upper bound remained open, since the proof in [DLR13] relies on regularity properties of optimal transport maps that are restricted to one dimension. In this note we present a new proof of the upper bound, thereby generalising the result of [ADPZ11] to arbitrary dimensions.

1. INTRODUCTION

In the recent paper [ADPZ11], Adams, Dirr, Peletier and Zimmer unveiled a fundamental connection between two seemingly unrelated aspects of the diffusion equation. They connected the *large deviation rate functional* for the empirical measure of a system of independent diffusing particles to the *entropy gradient flow structure* of the diffusion equation in the Wasserstein space of probability measures. Let us informally describe these two concepts and their connection here, before giving rigorous statements in Section 2.

Large deviations for independently diffusing particles. We consider n indistinguishable particles evolving according to the stochastic differential equations

$$\mathrm{d}X_i(t) = \nabla\Psi(X_i(t))\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_i(t)\ , \tag{1}$$

where $(W_1(t),\ldots,W_n(t))_{t\geq 0}$ is a collection of independent standard $\mathbb{R}^d$-valued Brownian motions. We assume that $\Psi : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable and that its Hessian is uniformly bounded from below. Let $\rho_t^{(n)} := n^{-1}\sum_{i=1}^n \delta_{X_i(t)}$ denote the empirical measure of $\big(X_i(t)\big)_{i=1}^n$. If the initial values $X_i(0)$ are chosen deterministically such that $\rho_0^{(n)}$ converges to some fixed measure $\rho_0 \in \mathcal{P}(\mathbb{R}^d)$, then, for each $t \geq 0$, the empirical measure $\rho_t^{(n)}$ converges almost surely to the unique solution of the Fokker-Planck equation

$$\partial_t\rho_t = \Delta\rho_t + \mathrm{div}(\rho_t\nabla\Psi)\ , \tag{2}$$

with initial condition ρ_0. Under suitable growth conditions on Ψ, a Sanov-type theorem implies that the random measures $(\rho_t^{(n)})_n$ satisfy a large deviation principle of the form

$$\mathbb{P}[\rho_t^n \approx \bar\rho] \sim \exp\big(-nI_t(\bar\rho|\rho_0)\big)\ ,$$

where the rate functional is given by

$$I_t(\bar\rho|\rho_0) := \inf_{\gamma\in\Gamma(\rho_0,\bar\rho)} H(\gamma|\rho_{0,t})\ , \tag{3}$$

see [Léo07, Proposition 3.2] and [PRV13, Theorem A.1]. Here, $\rho_{0,t} \in \mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$ denotes the joint law of a solution (X_0, X_t) to (1) with random initial condition $X_0 \sim \rho_0$ (independent of the Brownian motion), $H(\cdot|\cdot)$ denotes the relative entropy, and $\Gamma(\rho_0,\bar\rho)$ is the set of probability measures $\gamma \in \mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$ with marginals ρ_0 and $\bar\rho$.

In this paper we are interested in the short-time behaviour of the rate functional $I_t(\cdot|\rho_0)$ and its relation to the Wasserstein gradient structure of the Fokker-Planck equation.

The Wasserstein gradient structure of the Fokker-Planck equation. A seminal result by Jordan-Kinderlehrer-Otto [JKO98] asserts that the Fokker-Planck equation (2) can be regarded as the gradient flow equation of the relative entropy

$$\mathcal{F}(\rho) := \begin{cases} \displaystyle\int_{\mathbb{R}^d} \rho(x) \log \rho(x) \, \mathrm{d}x + \int_{\mathbb{R}^d} \Psi(x) \rho(x) \, \mathrm{d}x & \rho(\mathrm{d}x) = \rho(x) \, \mathrm{d}x \ , \\ +\infty & \text{otherwise} \ , \end{cases}$$

in the Wasserstein space of probability measures $(\mathcal{P}_2(\mathbb{R}^d), W_2)$. This result can be rigorously interpreted in different ways, e.g., using the theory of gradient flows in metric spaces, or using an infinite-dimensional Riemannian structure on the space of probability measures; see [AGS08] for details. Here we present the original interpretation from [JKO98], in terms of the convergence of a discrete "minimizing movement" scheme. For $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and $t > 0$, define $J_t(\cdot|\rho_0) : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ by

$$J_t(\bar{\rho}|\rho_0) := \mathcal{F}(\bar{\rho}) + \frac{1}{2t} W_2(\rho_0, \bar{\rho})^2 \ , \quad \text{and set} \quad S_t\rho_0 := \underset{\bar{\rho} \in \mathcal{P}(\mathbb{R}^d)}{\arg\min} \, J_t(\bar{\rho}|\rho_0) \ . \tag{4}$$

Since this minimisation problem has a unique minimiser, $S_t\rho_0$ is well defined. The JKO-functional J_t can be used to construct an iterative discretization scheme: it was shown in [JKO98] that

$$\rho_t := \lim_{n \to \infty} \big(S_{t/n}\big)^n \rho_0$$

exists for each $t > 0$ and satisfies the Fokker-Planck equation (2).

Relating I_t and J_t. The main result of [ADPZ11] unveils a relation between the large deviation principle and the Wasserstein gradient flow structure. Roughly speaking, it asserts that the functionals I_t and $\frac{1}{2}J_t$ are asymptotically equivalent as $t \to 0$. More precisely, it was shown that

$$I_t(\cdot|\rho_0) - \frac{1}{4t} W_2(\cdot, \rho_0)^2 \to \frac{1}{2}\mathcal{F}(\cdot) - \frac{1}{2}\mathcal{F}(\rho) \qquad \text{as } t \to 0 \ , \tag{5}$$

in the sense of Γ-convergence with respect to the narrow topology. This result provides an appealing microscopic explanation for the emergence of the Wasserstein gradient flow structure at the macroscopic level.

The proof of this theorem in [ADPZ11] required two strong technical assumptions. Firstly, the result was limited to one space dimension. Secondly, the proof required highly restrictive regularity assumptions on the involved measures.

In a subsequent paper [DLR13], Duong, Laschos and Renger were able to remove the strong regularity assumptions. Their approach is based on a different representation of the rate functional I_t due to Dawson and Gärtner [DG87] (see also [FK06]), that we shall describe in Section 2. The proof of the lower bound in the Γ-convergence result in [DLR13] is valid in arbitrary dimensions. However, the remaining part of the argument (the construction of a recovery sequence) is restricted to one dimension, since it relies on regularity estimates for optimal transport maps which are known to be false in multiple dimensions.

In this note we shall provide a different argument for the construction of a recovery sequence that works in arbitrary dimensions. Combined with the result from [DLR13], this completes the proof of (5) in arbitrary dimensions. We refer to Theorem 2.2 below for a precise statement.

Structure of the paper. In Section 2 we give a detailed statement of the main convergence result. In Section 3 we collect well-known results about Wasserstein gradient flows that will be used in the proof. Section 4 contains the proof of the convergence result. For completeness, we also include the proof of the lower bound taken from [DLR13]. In the appendix we provide a short proof of the equivalence of different formulations of the Benamou-Brenier formula.

2. Statement of the main results

In this section we shall rigorously introduce the three objects appearing in the main result of this paper: the Wasserstein metric W_2, the relative entropy functional $\mathcal{F}$, and the large deviation rate functional I_τ.

The Wasserstein metric. Let $\mathcal{P}_2(\mathbb{R}^d) := \{\rho \in \mathcal{P}(\mathbb{R}^d) : \int |x|^2 \rho(\mathrm{d}x) < \infty\}$ denote the set of probability measures with finite second moment. The L^2-Wasserstein distance between $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$ is defined by

$$W_2(\rho_0, \rho_1) := \inf_{\pi \in \Gamma(\rho_0, \rho_1)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, \pi(\mathrm{d}x, \mathrm{d}y) \right)^{1/2} ,$$

where the infimum is taken over all couplings π of ρ_0 and ρ_1, i.e., $\Gamma(\rho_0, \rho_1)$ denotes the collection of all $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with $\pi(\cdot \times \mathbb{R}^d) = \rho_0(\cdot)$ and $\pi(\mathbb{R}^d \times \cdot) = \rho_1(\cdot)$.

The relative entropy. Throughout this paper we assume that $\Psi : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable and λ-convex for some $\lambda \in \mathbb{R}$, i.e., $\mathrm{Hess}\, \Psi(x) \geq \lambda \,\mathrm{Id}$ for all $x \in \mathbb{R}^d$. The relative entropy functional $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\mathcal{F}(\rho) := \begin{cases} \displaystyle\int_{\mathbb{R}^d} f(x) \log f(x) \, \mathrm{d}x + \int_{\mathbb{R}^d} \Psi(x) f(x) \, \mathrm{d}x & \text{if } \rho(\mathrm{d}x) = f(x)\, \mathrm{d}x \ , \\ +\infty & \text{otherwise} \ . \end{cases}$$

This functional is well-defined, since the assumption on the second moment implies that the negative parts of $f \log f$ and Ψf are integrable with respect to the Lebesgue measure. If ρ is absolutely continuous with respect to the Lebesgue measure, then $\mathcal{F}$ can be written as a relative entropy

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^d} g(x) \log g(x) \, \mathrm{d}\nu(x) \ ,$$

where ν is the Borel measure on $\mathbb{R}^d$ defined by $\nu(\mathrm{d}x) = e^{-\Psi(x)} \, \mathrm{d}x$, and $\rho(\mathrm{d}x) = g(x)\nu(\mathrm{d}x)$.

We also introduce the relative Fisher information $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to [0, +\infty]$ defined by

$$\mathcal{G}(\rho) = \begin{cases} \displaystyle\int_{\{g>0\}} \frac{|\nabla g(x)|^2}{g(x)} \, \mathrm{d}\nu(x) & \text{if } \rho(\mathrm{d}x) = g(x)\nu(\mathrm{d}x), \ g \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^d) \ , \\ +\infty & \text{otherwise} \ . \end{cases}$$

The large deviation rate functional. The definition of the rate functional I_τ involves a weighted Sobolev norm of negative order 1. Let $\mathcal{D} = C_c^\infty(\mathbb{R}^d)$ be the space of test functions and let $\mathcal{D}'$ be the dual space of distributions. Given $\rho \in \mathcal{P}(\mathbb{R}^d)$, we define the weighted $\dot{H}^{-1}(\rho)$-norm of $s \in \mathcal{D}'$ by the duality formula

$$\|s\|_{-1,\rho}^2 := \sup_{f\in\mathcal{D}} \frac{\langle s, f\rangle^2}{\displaystyle\int_{\mathbb{R}^d} |\nabla f|^2 \,\mathrm{d}\rho} ,$$

where the supremum runs over all smooth test functions $f \in \mathcal{D}$ for which the denominator does not vanish. Using the identity $b^2/a^2 = \sup_{t\in\mathbb{R}} 2tb - t^2a^2$, one obtains the equivalent formula

$$\|s\|_{-1,\rho}^2 = \sup_{f\in\mathcal{D}} \left\{ 2\langle s, f\rangle - \int |\nabla f|^2 \,\mathrm{d}\rho \right\} .$$

For fixed $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\tau > 0$, the functional $I_\tau(\cdot|\rho_0) : \mathcal{P}_2(\mathbb{R}^d) \to [0, +\infty]$ is defined by

$$I_\tau(\bar\rho|\rho_0) := \inf_{(\rho_t)_t\in\mathrm{AC}(\rho_0,\bar\rho)} \frac{1}{4\tau} \int_0^1 \big\|\partial_t\rho_t - \tau\Delta\rho_t - \tau\,\mathrm{div}(\rho_t\nabla\Psi)\big\|_{-1,\rho_t}^2 \,\mathrm{d}t , \tag{6}$$

where $\mathrm{AC}(\rho_0, \rho_1)$ denotes the set of absolutely continuous curves $(\rho_t)_{t\in[0,1]}$ in $\big(\mathcal{P}_2(\mathbb{R}^d), W_2\big)$ with boundary conditions $\rho|_{t=0} = \rho_0$ and $\rho|_{t=1} = \rho_1$. Intuitively, $I_\tau(\bar\rho|\rho_0)$ is the value of an optimal control problem, which requires to interpolate between ρ_0 and $\bar\rho$ in such a way that deviations from the Fokker-Planck equation

$$\partial_t\rho_t = \tau\Delta\rho_t + \tau\,\mathrm{div}(\rho_t\nabla\Psi)$$

are minimised.

Remark 2.1. Under suitable growth conditions on the potential Ψ, the term inside the infimum of (6) is the large deviation rate functional for trajectories $[0, \tau] \to \mathcal{P}(\mathbb{R}^d)$ of the empirical process of independent particles, see [DG87]. It then follows from the contraction principle that (6) is the large deviation rate functional for the empirical measure at time τ, i.e., it coincides with (3). (Note that the time interval $[0, \tau]$ has been rescaled to $[0, 1]$ in (6).) In this paper we will not be concerned with the exact conditions under which these expressions coincide, but rather take (6) as the object of study. For more details, see [DLR13, Section 4]. □

Now we are ready to state the main theorem of this paper:

Theorem 2.2 (Main result)**.** *Let $\Psi \in C^2(\mathbb{R}^d)$ be λ-convex for some $\lambda \in \mathbb{R}$. Then, for every $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\mathcal{G}(\rho_0) < \infty$, we have*

$$I_\tau(\,\cdot \mid \rho_0) - \frac{W_2^2(\rho_0,\,\cdot\,)}{4\tau} \xrightarrow[\tau\to0]{M} \frac{1}{2}\mathcal{F}(\,\cdot\,) - \frac{1}{2}\mathcal{F}(\rho_0) \tag{7}$$

in the sense of Mosco convergence. More precisely:

(i) *For any narrowly converging sequence $\rho_1^\tau \rightharpoonup \rho_1$ in $P_2(\mathbb{R}^d)$,*

$$\liminf_{\tau\to0} \left(I_\tau(\rho_1^\tau \mid \rho_0) - \frac{W_2^2(\rho_0, \rho_1^\tau)}{4\tau} \right) \ \ge\ \frac{1}{2}\mathcal{F}(\rho_1) - \frac{1}{2}\mathcal{F}(\rho_0) . \tag{8}$$

(ii) *For any $\rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a sequence $\rho_1^\tau \in \mathcal{P}_2(\mathbb{R}^d)$ converging to ρ_1 in the Wasserstein metric such that*

$$\limsup_{\tau \to 0} \left(I_\tau(\rho_1^\tau \mid \rho_0) - \frac{W_2^2(\rho_0, \rho_1^\tau)}{4\tau} \right) \leq \frac{1}{2}\mathcal{F}(\rho_1) - \frac{1}{2}\mathcal{F}(\rho_0) \,. \tag{9}$$

As discussed in the introduction, this theorem was first proved in dimension 1 in [ADPZ11] under more restrictive conditions on the measures ρ_0 and ρ_1. Part (i) has been extended to arbitrary dimensions in [DLR13]. The novel contribution of our paper is a proof of (ii) in arbitrary dimensions.

Remark 2.3. The right-hand side in (8) and (9) is well-defined in $\mathbb{R} \cup \{+\infty\}$, since our assumptions on ρ_0 imply that $\mathcal{F}(\rho_0) < \infty$. This is a consequence of the HWI-inequality by Otto and Villani [OV00] (see also [Vil09, Corollary 20.13]), which asserts that $\mathcal{G}(\rho) \leq W_2(\rho, \nu)\sqrt{\mathcal{F}(\rho)} - \frac{\lambda}{2} W_2^2(\rho, \nu)$. □

3. Ingredients of the proof

The Benamou-Brenier formula. It will be convenient to work with the dynamic characterization of the Wasserstein distance due to Benamou–Brenier [BB00]. It asserts that, for $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$,

$$W_2^2(\rho_0, \rho_1) = \inf_{(\rho_t)_t \in \mathrm{AC}(\rho_0, \rho_1)} \left\{ \int_0^1 \|\partial_t \rho_t\|_{-1,\rho_t}^2 \,\mathrm{d}t \right\} \,, \tag{10}$$

Recall that for any absolutely continuous curve $(\rho_t)_{t \in [0,1]}$ with respect to W_2, the metric derivative

$$|\dot{\rho}_t| := \lim_{h \to 0} \frac{W_2(\rho_{t+h}, \rho_t)}{h}$$

exists for a.e. t, see, e.g., [AGS08, Theorem 1.1.2]. In view of (10), we have the identity

$$|\dot{\rho}_t| = \|\partial_t \rho_t\|_{-1,\rho_t} \,. \tag{11}$$

We refer to Appendix A for an equivalent formulation of the Benamou-Brenier formula which is commonly used in the literature on optimal transport and to [AGS08, Theorem 8.3.1] for a proof of (10), (11) in this formulation.

Relative entropy, Fisher information, and heat flow. A seminal result by McCann [McC97] asserts that the λ-convexity of Ψ implies *displacement λ-convexity* of $\mathcal{F}$, see also [Vil03, Theorem 5.15]. This means that for any constant speed W_2-geodesic $(\rho_t)_{t \in [0,1]} \subseteq \mathcal{P}_2(\mathbb{R}^d)$ and any $t \in [0, 1]$, we have

$$\mathcal{F}(\rho_t) \leq (1-t)\mathcal{F}(\rho_0) + t\mathcal{F}(\rho_1) - \frac{\lambda}{2} t(1-t) W_2^2(\rho_0, \rho_1) \,. \tag{12}$$

In particular, $\mathcal{F}$ is finite along geodesics as soon as it is finite at the endpoints. The fact that the relative Fisher-information does *not* enjoy this property is the source of several complications in [DLR13].

The semigroup associated to the Fokker–Planck equation (2) will be denoted by $(P_t)_{t \geq 0}$. More precisely, for $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ we set $P_t \rho := \rho_t$, where $(\rho_t)_t$ is the unique distributional solution to the Fokker–Planck equation (2) with $\rho_0 = \rho$. This solution can be obtained using, e.g., the metric theory of gradient flows for (generalised) λ-convex functionals, see [AGS08, Thm. 11.2.8].

In the following result we collect some well-known results on the behaviour of the semigroup $(P_t)_{t\geq 0}$.

Lemma 3.1. *The following assertions hold:*

1 *The curve $t \mapsto P_t\rho$ is continuous on $[0,\infty)$ and locally absolutely continuous on $(0,\infty)$ with respect to W_2.*

2 *For all $\rho, \sigma \in \mathcal{P}_2(\mathbb{R}^d)$ and all $t \geq 0$ we have the contraction estimate:*

$$W_2(P_t\rho, P_t\sigma) \leq e^{-\lambda t} W_2(\rho, \sigma) \tag{13}$$

Moreover, for any curve $(\rho_s)_s$ that is absolutely continuous with respect to W_2 we have

$$\|\partial_s(P_t\rho_s)\|_{-1,P_t\rho_s} \leq e^{-\lambda t}\|\partial_s\rho_s\|_{-1,\rho_s} \ . \tag{14}$$

3 *For all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ and $t > 0$ we have*

$$\mathcal{F}(P_t\rho) < \infty \ , \quad \mathcal{G}(P_t\rho) < \infty \ , \tag{15}$$

as well as the bounds

$$\mathcal{F}(P_t\rho) \leq \mathcal{F}(\rho) \ , \quad \mathcal{G}(P_t\rho) \leq e^{-2\lambda t}\mathcal{G}(\rho) \ . \tag{16}$$

Finally, for any W_2-geodesic $(\rho_s)_{s\in[0,1]}$ with $\mathcal{F}(\rho_0), \mathcal{F}(\rho_1) < \infty$, we have as $t \searrow 0$:

$$\mathcal{F}(P_t\rho_s) \ \nearrow \ \mathcal{F}(\rho_s) \quad \textit{uniformly in } s \in [0,1] \ . \tag{17}$$

Proof. For part (1) and the properties (13), (15) and (16), see [AGS08, Theorems 11.2.1 and 11.2.8]. The estimate (14) follows immediately from (13) and (11). It remains to prove the statement (17), which is less standard. Note first that by (12) we have that $s \mapsto \mathcal{F}(\rho_s)$ is continuous and bounded. Our aim is to show that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\mathcal{F}(\rho_s) - \mathcal{F}(P_t\rho_s) < \varepsilon$ whenever $t < \delta$ and $s \in [0,1]$. Assume the contrary, i.e., that there exist $\varepsilon > 0$ and sequences $t_k \to 0$ and $(s_k) \subset [0,1]$ such that for all k,

$$\mathcal{F}(\rho_{s_k}) - \mathcal{F}(P_{t_k}\rho_{s_k}) \ \geq \ \varepsilon \ . \tag{18}$$

By compactness we can assume that $s_k \to s_0$ as $k \to \infty$ for some $s_0 \in [0,1]$. We claim that $P_{t_k}\rho_{s_k} \to \rho_{s_0}$ in W_2-distance as $k \to \infty$. Indeed, again by (13) the triangle inequality yields

$$\begin{aligned} W_2(\rho_{s_0}, P_{t_k}\rho_{s_k}) \ &\leq \ W_2(\rho_{s_0}, P_{t_k}\rho_{s_0}) + W_2(P_{t_k}\rho_{s_0}, P_{t_k}\rho_{s_k}) \\ &\leq \ W_2(\rho_{s_0}, P_{t_k}\rho_{s_0}) + e^{-\lambda t_k} W_2(\rho_{s_0}, \rho_{s_k}) \ , \end{aligned}$$

and the claim follows from the continuity of P_t at $t = 0$ and the continuity of the curve (ρ_s). Passing to the limit $k \to \infty$ in (18), using the continuity of $s \mapsto \mathcal{F}(\rho_s)$ and the lower semicontinuity of $\mathcal{F}$ with respect to W_2, we obtain the following contradiction:

$$0 \ = \ \mathcal{F}(\rho_{s_0}) - \mathcal{F}(\rho_{s_0}) \ \geq \ \limsup_{k\to\infty} \Big(\mathcal{F}(\rho_{s_k}) - \mathcal{F}(P_{t_k}\rho_{s_k})\Big) \ \geq \ \varepsilon \ ,$$

which completes the proof. □

We conclude this section by stating some useful identities for the derivative of the entropy. In fact, for any absolutely continuous curve $(\rho_t)_{t\in[0,1]}$ with $\mathcal{F}(\rho_t)\in\mathbb{R}$ for all t and $\int_0^1 \mathcal{G}(\rho_t)\,\mathrm{d}t<\infty$ we have that $t\mapsto\mathcal{F}(\rho_t)$ is absolutely continuous with

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\rho_t) \;=\; -\big\langle \partial_t\rho_t, \Delta\rho_t + \mathsf{div}(\rho_t\nabla\Psi)\big\rangle_{-1,\rho_t} \tag{19}$$

for a.e. $t\in[0,1]$, see [DLR13, Lemma 2.3]. In particular, if ρ_t satisfies the Fokker-Planck equation we have

$$-\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\rho_t) = \|\Delta\rho_t + \mathsf{div}(\rho_t\nabla\Psi)\|^2_{-1,\rho_t} = \mathcal{G}(\rho_t)\ , \tag{20}$$

where the second equality follows from (25).

4. Proof of the main result

4.1. **Upper bound.** In this section we prove existence of the recovery sequence, i.e., statement (ii) of Theorem 2.2. For this purpose we define the set $Q := \big\{\rho\in\mathcal{P}_2(\mathbb{R}^d) : \mathcal{G}(\rho)<\infty\big\}$. Note that $\mathcal{F}(\rho)<\infty$ for all $\rho\in Q$ in view of Remark 2.3. Below we will prove the following two claims:

Claim 4.1. *For all $\rho_0,\rho_1\in Q$ we have as $\tau\to 0$,*

$$I_\tau(\rho_1\mid\rho_0) - \frac{1}{4\tau}W_2^2(\rho_0,\rho_1) \to \frac{1}{2}\mathcal{F}(\rho_1) - \frac{1}{2}\mathcal{F}(\rho_0)\ . \tag{21}$$

Claim 4.2. *For every $\rho\in\mathcal{P}_2(\mathbb{R}^d)$ there exists a sequence $(\rho^n)_n\subseteq Q$ such that $W_2^2(\rho^n,\rho)\to 0$ and $\mathcal{F}(\rho^n)\to\mathcal{F}(\rho)$.*

The existence of the recovery sequence then follows from a straightforward diagonal argument, see [DLR13, Proposition 6.2] for details.

Proof of Claim 4.1: We only need to prove the limsup inequality for the left-hand side of (21), since the liminf inequality will be proved in Section 4.2 below. If $\rho_0=\rho_1$ the claim is immediate, so we take distinct measures $\rho_0,\rho_1\in Q$, and take a geodesic $(\rho_t)_{t\in[0,1]}$ connecting ρ_0 and ρ_1. We will approximate this curve by running the semigroup for a small time $\varepsilon=\varepsilon(\tau)>0$, which will be determined below. A careful choice of ε as a function of τ is crucial for our argument. We thus consider the curve $(\rho_t^\varepsilon)_{t\in[0,1]}$ defined by

$$\rho_t^\varepsilon \;=\; \begin{cases} P_t\rho_0\ , & 0\le t\le\varepsilon\ ,\\ P_\varepsilon\rho_{\frac{t-\varepsilon}{1-2\varepsilon}}\ , & \varepsilon\le t\le 1-\varepsilon\ ,\\ P_{1-t}\rho_1\ , & 1-\varepsilon\le t\le 1\ . \end{cases}$$

For the sake of brevity, we shall write $\mathcal{L}\rho = \Delta\rho + \operatorname{div}(\rho\nabla\Psi)$. Using the definition of $I_\tau(\rho_1 \mid \rho_0)$ and the second identity (20), we obtain

$$\begin{aligned}
&I_\tau(\rho_1 \mid \rho_0) - \frac{W_2^2(\rho_0,\rho_1)}{4\tau}\\
&\le \frac{1}{4\tau}\left(\int_0^1 \|\partial_t\rho_t^\varepsilon - \tau\mathcal{L}\rho_t^\varepsilon\|_{-1,\rho_t^\varepsilon}^2 \,\mathrm{d}t - W_2^2(\rho_0,\rho_1)\right)\\
&= \frac{1}{4\tau}\left(\int_0^1 \|\partial_t\rho_t^\varepsilon\|_{-1,\rho_t^\varepsilon}^2 \,\mathrm{d}t - W_2^2(\rho_0,\rho_1)\right) - \frac{1}{2}\int_0^1 \langle\partial_t\rho_t^\varepsilon, \mathcal{L}\rho_t^\varepsilon\rangle_{-1,\rho_t^\varepsilon}\,\mathrm{d}t + \frac{\tau}{4}\int_0^1 \mathcal{G}(\rho_t^\varepsilon)\,\mathrm{d}t\ .
\end{aligned}$$

We shall estimate these three terms separately. Let $c_\lambda, k_\lambda > 0$ be sufficiently large so that $\frac{e^{-2\lambda\varepsilon}}{1-2\varepsilon} \le 1 + k_\lambda\varepsilon$ and $\int_0^\varepsilon e^{-2\lambda t}\,\mathrm{d}t \le c_\lambda\varepsilon$ for all $\varepsilon \in (0,\frac{1}{4})$. Using the semigroup estimates (16) and (14) and the Benamou-Brenier formula (10), the first term can be bounded by

$$\begin{aligned}
&\int_0^1 \|\partial_t\rho_t^\varepsilon\|_{-1,\rho_t^\varepsilon}^2 \,\mathrm{d}t\\
&\qquad = \int_0^\varepsilon \|\mathcal{L}\rho_t^\varepsilon\|_{-1,\rho_t^\varepsilon}^2\,\mathrm{d}t + \frac{1}{1-2\varepsilon}\int_0^1 \|\partial_t(P_\varepsilon\rho_t)\|_{-1,P_\varepsilon\rho_t}^2\,\mathrm{d}t + \int_{1-\varepsilon}^1 \|\mathcal{L}\rho_t^\varepsilon\|_{-1,\rho_t^\varepsilon}^2\,\mathrm{d}t\\
&\qquad \le \int_0^\varepsilon \mathcal{G}(P_t\rho_0)\,\mathrm{d}t + \frac{e^{-2\lambda\varepsilon}}{1-2\varepsilon}\int_0^1 \|\partial_t\rho_t\|_{-1,\rho_t}^2\,\mathrm{d}t + \int_{1-\varepsilon}^1 \mathcal{G}(P_{1-t}\rho_1)\,\mathrm{d}t\\
&\qquad \le c_\lambda\varepsilon\mathcal{G}(\rho_0) + (1+k_\lambda\varepsilon)W_2^2(\rho_0,\rho_1) + c_\lambda\varepsilon\mathcal{G}(\rho_1)\ .
\end{aligned}$$

To treat the second term, we observe that (19) yields

$$\int_0^1 \langle\partial_t\rho_t^\varepsilon, \mathcal{L}\rho_t^\varepsilon\rangle_{-1,\rho_t^\varepsilon}\,\mathrm{d}t = \mathcal{F}(\rho_0) - \mathcal{F}(\rho_1)\ .$$

For the third term we use (16) to obtain

$$\int_0^1 \mathcal{G}(\rho_t^\varepsilon)\,\mathrm{d}t \le c_\lambda\varepsilon(\mathcal{G}(\rho_0)+\mathcal{G}(\rho_1)) + h(\varepsilon)\ , \quad \text{where} \quad h(\varepsilon) = \int_0^1 \mathcal{G}(P_\varepsilon\rho_t)\,\mathrm{d}t\ .$$

Combining these three bounds, we infer that

$$\begin{aligned}
I_\tau(\rho_1 \mid \rho_0) - \frac{W_2^2(\rho_0,\rho_1)}{4\tau} &\le \frac{1}{2}\mathcal{F}(\rho_1) - \frac{1}{2}\mathcal{F}(\rho_0) + \varepsilon\frac{c_\lambda}{4}\Big(\tau + \frac{1}{\tau}\Big)\big(\mathcal{G}(\rho_0)+\mathcal{G}(\rho_1)\big)\\
&\quad + \frac{k_\lambda\varepsilon}{4\tau}W_2^2(\rho_0,\rho_1) + \frac{\tau}{4}h(\varepsilon)\ .
\end{aligned}$$

We claim that $\varepsilon = \varepsilon(\tau)$ can be chosen as a function of τ such that

$$\frac{\varepsilon(\tau)}{\tau} \to 0 \quad \text{and} \quad \tau h\big(\varepsilon(\tau)\big) \to 0 \quad \text{as } \tau \to 0\ . \tag{22}$$

This yields the limsup inequality in (21). The corresponding liminf inequality will follow from (8).

It thus remains to prove the claim (22). For $\varepsilon > 0$ we set

$$g(\varepsilon) := \sqrt{\varepsilon/h(\varepsilon)}\ .$$

Writing $g(\varepsilon) = \sqrt{\varepsilon e^{2\lambda\varepsilon}/e^{2\lambda\varepsilon}h(\varepsilon)}$, it follows from (16) that g is strictly increasing on $(0,\varepsilon_0)$ for ε_0 sufficiently small. Taking into account that $h(0) > 0$ since $\rho_0 \ne \rho_1$, we note that $\lim_{\varepsilon\to 0} g(\varepsilon) = 0$. To show that g is right-continuous, note that for each $t \in [0,1]$, the function

$G_t : \varepsilon \mapsto \mathcal{G}(P_\varepsilon \rho_t)$ is lower semicontinuous and non-negative, see e.g. [AGS08, Proposition 10.4.14]. Fatou's lemma implies that $h := \int_0^1 G_t \, \mathrm{d}t$ is lower semicontinuous as well. Hence g is upper semicontinuous and thus right-continuous, since it is also increasing. It follows from these properties that we can define

$$\varepsilon(\tau) := g^{-1}(\tau) := \inf \big\{ \varepsilon : g(\varepsilon) > \tau \big\}$$

as the generalised inverse of g. We shall show that this function has the desired properties.

Since g is right-continuous, we note that $g(\varepsilon(\tau)/2) \le \tau \le g(\varepsilon(\tau))$, which implies that the expressions in (22) can be estimated from above by

$$\frac{\varepsilon(\tau)}{\tau} \le 2\sqrt{\frac{\varepsilon(\tau)}{2} h\left(\frac{\varepsilon(\tau)}{2}\right)}\,, \qquad \tau h\big(\varepsilon(\tau)\big) \le \sqrt{\varepsilon(\tau) h(\varepsilon(\tau))}\,.$$

It thus suffices to show that $\varepsilon h(\varepsilon) \to 0$ as $\varepsilon \to 0$. To show this, note that $\varepsilon^{-1} \int_0^\varepsilon e^{2\lambda s} \, \mathrm{d}s \ge \min\{1, e^{\lambda/2}\} =: \tilde{k}_\lambda$ for all $\varepsilon \in (0, \frac{1}{4})$. Using (16) and (20) we obtain

$$\tilde{k}_\lambda \varepsilon \mathcal{G}(P_\varepsilon \rho_t) \le \mathcal{G}(P_\varepsilon \rho_t) \int_0^\varepsilon e^{2\lambda(\varepsilon - s)} \, \mathrm{d}s \le \int_0^\varepsilon \mathcal{G}(P_s \rho_t) \, \mathrm{d}s = \mathcal{F}(\rho_t) - \mathcal{F}(P_\varepsilon \rho_t)\,.$$

By (17) the right-hand side converges to 0 as $\varepsilon \to 0$, uniformly for $t \in [0,1]$. It follows that

$$\varepsilon h(\varepsilon) = \varepsilon \int_0^1 \mathcal{G}(P_\varepsilon \rho_t) \, \mathrm{d}t \ \to \ 0$$

as $\varepsilon \to 0$, which completes the proof. □

Proof of Claim 4.2: We approximate $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ by applying the semigroup. The first inequality in (15) yield that $P_\varepsilon \rho \in Q$ for any $\varepsilon > 0$, and Lemma 3.1(1) implies that $P_\varepsilon \rho$ approximates ρ in W_2-distance. Finally, since $\mathcal{F}$ is lower semicontinuous with respect to W_2, the convergence $\mathcal{F}(P_\varepsilon \rho) \to \mathcal{F}(\rho)$ as $\varepsilon \to 0$ follows from (16). □

4.2. **Lower bound.** For completeness, we reproduce here the short proof of statement i) in Theorem 2.2, the lower bound, given in [DLR13, Theorem 5.1].

By definition of the infimum in (6), there exists a sequence of absolutely continuous curves $(\rho_t^\tau)_{t \in [0,1]}$ with $\int_0^1 \mathcal{G}(\rho_t^\tau) \, \mathrm{d}t < \infty$ such that

$$\begin{aligned} I_\tau(\rho_1^\tau \mid \rho_0) + \tau &\ge \frac{1}{4\tau} \int_0^1 \big\| \partial_t \rho_t^\tau - \tau(\Delta \rho_t^\tau + \mathrm{div}(\rho_t^\tau \nabla \Psi)) \big\|_{-1,\rho_t^\tau}^2 \, \mathrm{d}t \\ &= \frac{1}{4\tau} \int_0^1 \| \partial_t \rho_t^\tau \|_{-1,\rho_t^\tau}^2 \, \mathrm{d}t + \frac{1}{2} \int_0^1 \langle \partial \rho_t^\tau, \Delta \rho_t^\tau + \mathrm{div}(\rho_t^\tau \nabla \Psi) \rangle_{-1,\rho_t^\tau} \, \mathrm{d}t \\ &\qquad + \frac{\tau}{4} \int_0^1 \big\| \Delta \rho_t^\tau + \mathrm{div}(\rho_t^\tau \nabla \Psi) \big\|_{-1,\rho_t^\tau}^2 \, \mathrm{d}t \\ &\ge \frac{1}{4\tau} W_2^2(\rho_0, \rho_1^\tau) + \frac{1}{2} \mathcal{F}(\rho_1^\tau) - \frac{1}{2} \mathcal{F}(\rho_0), \end{aligned}$$

where the last line follows from the Benamou-Brenier formula (10) together with the fact that $t \mapsto \mathcal{F}(\rho_t^\tau)$ is absolutely continuous and satisfies (19). The claim (8) then follows from the narrow lower semicontinuity of $\mathcal{F}$.

Appendix A. Equivalent formulations of the Benamou-Brenier formula

The Benamou-Brenier formula in optimal transport asserts that for $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$,

$$W_2^2(\rho_0, \rho_1) = \inf_{(\rho_t)_t \in \mathrm{AC}(\rho_0,\rho_1)} \left\{ \int_0^1 |||\partial_t \rho_t|||^2_{-1,\rho_t} \, \mathrm{d}t \right\} . \tag{23}$$

In this formula, the norm $|||\cdot|||_{-1,\rho}$ is defined by

$$|||s|||^2_{-1,\rho} := \inf_{v \in L^2(\rho;\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} |v(x)|^2 \, \mathrm{d}\rho(x) \ : \ s + \mathrm{div}(\rho v) = 0 \right\} . \tag{24}$$

for $\rho \in \mathcal{P}(\mathbb{R}^d)$ and $s \in \mathcal{D}'$. It can be shown that the infimum in this definition is uniquely attained, and its minimiser can be characterised as follows: a solution $v \in L^2(\rho;\mathbb{R}^d)$ to the "continuity equation" $s + \mathrm{div}(\rho v) = 0$ is optimal in (24) if and only if it belongs to the space of generalised gradient vector fields defined by

$$H_\rho := \overline{\{\nabla\psi : \mathbb{R}^d \to \mathbb{R}^d \mid \psi \in \mathcal{D}\}}^{L^2(\rho;\mathbb{R}^d)} .$$

We refer to [AGS08, Section 8.4] for the proof of these facts. Note in particular that

$$|||\mathrm{div}(\rho\nabla\psi)|||^2_{-1,\rho} = \int_{\mathbb{R}^d} |\nabla\psi(x)|^2 \, \mathrm{d}\rho(x) \tag{25}$$

whenever $\nabla\psi \in L^2(\rho;\mathbb{R}^d)$.

The following lemma relates the norm $|||\cdot|||_{-1,\rho}$ to the norm $\|\cdot\|_{-1,\rho}$ defined in Section 2.

Lemma A.1. *Let $\rho \in \mathcal{P}(\mathbb{R}^d)$ and $s \in \mathcal{D}'$. Then $\|s\|_{-1,\rho} = |||s|||_{-1,\rho}$.*

Proof. Suppose first that $|||s|||_{-1,\rho} < \infty$, and let $v \in L^2(\rho;\mathbb{R}^d)$ be the unique minimiser in the definition of $|||s|||_{-1,\rho}$. If $|||s|||_{-1,\rho} = 0$, it follows that v vanishes ρ-a.e., hence $\langle s, f\rangle = 0$ for all $f \in \mathcal{D}$, which implies that $\|s\|_{-1,\rho} = 0$. Assume now, without loss of generality, that $|||s|||^2_{-1,\rho} = \int |v|^2 \, \mathrm{d}\rho = 1$. Then,

$$\begin{aligned}
\|s\|_{-1,\rho} &= \sup_{f\in\mathcal{D}} \left\{ \langle -\mathrm{div}(\rho v), f\rangle \ \middle| \ \int_{\mathbb{R}^d} |\nabla f|^2 \, \mathrm{d}\rho = 1 \right\} \\
&= \sup_{f\in\mathcal{D}} \left\{ \int_{\mathbb{R}^d} v \cdot \nabla f \, \mathrm{d}\rho \ \middle| \ \int_{\mathbb{R}^d} |\nabla f|^2 \, \mathrm{d}\rho = 1 \right\} \\
&= \sup_{f\in\mathcal{D}} \left\{ \frac{1}{2} \int_{\mathbb{R}^d} |v|^2 + |\nabla f|^2 - |v - \nabla f|^2 \, \mathrm{d}\rho \ \middle| \ \int_{\mathbb{R}^d} |\nabla f|^2 \, \mathrm{d}\rho = 1 \right\} \\
&= \sup_{f\in\mathcal{D}} \left\{ 1 - \frac{1}{2} \int_{\mathbb{R}^d} |v - \nabla f|^2 \, \mathrm{d}\rho \ \middle| \ \int_{\mathbb{R}^d} |\nabla f|^2 \, \mathrm{d}\rho = 1 \right\} .
\end{aligned}$$

Since $v \in H_\rho$, it follows from this computation that $\|s\|_{-1,\rho} = 1 = |||s|||_{-1,\rho}$.

On the other hand, if $\|s\|_{-1,\rho} < \infty$, it follows from $\langle s, f\rangle \le \|s\|_{-1,\rho} \cdot \|\nabla f\|_{L^2(\rho;\mathbb{R}^d)}$ that the mapping

$$T : \{\nabla f : f \in \mathcal{D}\} \to \mathbb{R}, \qquad \nabla f \mapsto \langle s, f\rangle$$

extends to a bounded linear functional $T : (H_\rho, \|\cdot\|_{L^2(\rho;\mathbb{R}^d)}) \to \mathbb{R}$ of norm $\|s\|_{-1,\rho}$. Hence, the Riesz representation theorem implies that $\langle s, f\rangle = \int_{\mathbb{R}^d} v \cdot \nabla f \, \mathrm{d}\rho$ for some $v \in H_\rho$ with

$\|v\|_{L^2(\rho;\mathbb{R}^d)} = \|s\|_{-1,\rho}$. It follows that $|||s|||_{-1,\rho} \leq \|v\|_{L^2(\rho;\mathbb{R}^d)}$. In view of the first part of the proof, the latter inequality is in fact an equality. □

As a consequence, of this lemma we infer that the Benamou-Brenier formulas in (10) and (23) are equivalent.

References

[ADPZ11] S. Adams, N. Dirr, M. A. Peletier, and J. Zimmer. From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage. *Communications in Mathematical Physics*, 307(3):791–815, 2011.

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics. ETH Zürich. Birkhauser, Basel, Switzerland, 2nd edition, 2008.

[BB00] J.D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[DG87] D.A. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

[DLR13] M. H. Duong, V. Laschos, and M. Renger. Wasserstein gradient flows from large deviations of many-particle limits. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(4):1166–1188, 2013.

[FK06] J. Feng and T.G. Kurtz. *Large deviations for stochastic processes*, volume 131 of *Mathematical surveys and monographs*. American Mathematical Society, Providence, RI, USA, 2006.

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

[Léo07] C. Léonard. A large deviation approach to optimal transport. `arxiv.org/abs/0710.1461v1`, 2007.

[McC97] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.

[OV00] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[PRV13] M.A. Peletier, D.R.M. Renger, and M. Veneroni. Variational formulation of the Fokker-Planck equation with decay: a particle approach. *Communications in Contemporary Mathematics*, 15(5):1350017, 2013.

[Vil03] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, USA, 2003.

[Vil09] C. Villani. *Optimal transport, Old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2009.