OPTIMAL CONTROL FOR A MIXED FLOW OF HAMILTONIAN AND GRADIENT TYPE IN SPACE OF PROBABILITY MEASURES

JIN FENG AND ANDRZEJ ŚWIĘCH, WITH APPENDIX B BY ATANAS STEFANOV

Abstract. In this paper we investigate an optimal control problem in the space of measures on $\mathbb{R}^2$. The problem is motivated by a stochastic interacting particle model which gives the 2-D Navier-Stokes equations in their vorticity formulation as mean-field equation. We prove that the associated Hamilton-Jacobi-Bellman equation, in the space of probability measures, is well-posed in an appropriately defined viscosity solution sense.

1. Introduction

We consider a system of controlled partial differential equations

$$\partial_t \rho + \operatorname{div}(\rho u) = \nu \Delta \rho + m, \tag{1.1}$$

$$u := u_\rho = -K^\perp * \rho. \tag{1.2}$$

In the above, $x = (x_1, x_2) \in \mathbb{R}^2$ and $\nu > 0$,

$$N(x) := -\frac{1}{2\pi} \log |x|, \quad K(x) := \nabla N(x) = -\frac{1}{2\pi} \frac{x}{|x|^2}, \quad x \in \mathbb{R}^2 \setminus \{0\}, \tag{1.3}$$

$$K^\perp(x) := JK(x) = \nabla^\perp N(x), \text{ where } \nabla^\perp = (\partial_{x_2}, -\partial_{x_1}) = J\nabla,$$

$$J := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Let $\mathcal{P}(\mathbb{R}^d)$ denote the space of probability measures on $\mathbb{R}^d$ and $\mathcal{P}_p(\mathbb{R}^d)$, the space of probability measures on $\mathbb{R}^d$ with finite p-th moments. We will introduce a weighted Sobolev space $H_{-1,\rho}(\mathbb{R}^2)$ in (1.12) which can be viewed as the tangent space of $\mathcal{P}_2(\mathbb{R}^2)$. The control variable m satisfies $m(t) \in H_{-1,\rho(t)}(\mathbb{R}^2)$, which means that the control only push along tangent directions. This ensures that the dynamic is kept on the space of probability measures. The meaning of a solution to (1.1)-(1.2) is made precise in Definition 1.1. The existence and some regularity properties of solutions are established in Section 2. Using the tools of calculus on the space of probability measures, we will later see (Lemma 3.8) that (1.1) can be rewritten

2000 *Mathematics Subject Classification.* Primary: 35R15, 49L25, 49J20; Secondary: 35Q30, 42B20, 49L20, 60F17, 60H10.

Key words and phrases. Viscosity solutions, mass transport theory, controlled gradient and Hamiltonian flows, Navier-Stokes in 2D.

The research of Jin Feng was partially supported by US ARO grant W911NF-08-1-0064 and by NSF grant DMS 0806434; the research of Andrzej Święch was partially supported by NSF grant DMS 0856485; the research of Atanas Stefanov was partially supported by NSF grant DMS 0701802; Jin Feng and Andrzej Święch have also been supported by the American Institute of Mathematics through a SQuaRE program. The authors thank the referee for useful remarks and references.

as an abstract evolution equation which is a mixture of a Hamiltonian flow, a gradient flow, and the control variable m, i.e.

$$\partial_t \rho = J\text{-grad}_\rho e - \nu \text{grad}_\rho s + m, \tag{1.4}$$

where $s : \mathcal{P}_2(\mathbb{R}^2) \mapsto \mathbb{R} \cup \{+\infty\}$ is the entropy functional (1.15), and $e : \mathcal{P}_2(\mathbb{R}^2) \mapsto \mathbb{R} \cup \{+\infty\}$ is an energy functional (1.17).

To define our optimal control problem, we prescribe an infinitesimal running cost

$$L(\rho, m) := \frac{1}{4\nu} \|m\|^2_{-1,\rho}. \tag{1.5}$$

We refer to (1.11) for the definition of $\| \cdot \|_{-1,\rho}$ which is defined for all Schwartz distributions. Writing the distributional derivative $\dot{\rho} = \partial_t \rho$, provided $\rho u \in L^1_{loc}(\mathbb{R}^2)$, and identifying m with $\dot{\rho}$ through (1.1), we have

$$L(\rho, \dot{\rho}) = \frac{1}{4\nu} \|\dot{\rho} + \text{div}(\rho u) - \nu \Delta \rho\|^2_{-1,\rho}.$$

In this article, by action integral, we mean

$$A_T[\rho(\cdot)] \ := \ \int_0^T L(\rho(s), \dot{\rho}(s)) ds. \tag{1.6}$$

If $u(t)\rho(t) \in L^1_{loc}(\mathbb{R}^2)$ for $t \in (0, T]$ a.e., then A_T is well defined and takes values in $[0, +\infty]$.

Next, we introduce a class of admissible paths, or equivalently controls by the above identification,

$$\begin{aligned} \mathcal{K} := \Big\{ & \rho(\cdot) : \rho \in C([0, \infty); \mathcal{P}_2(\mathbb{R}^2)), \exists p : [0, \infty) \times \mathbb{R}^2 \mapsto \mathbb{R} \text{ such that} \\ & \exists 0 = t_0 < t_1 < ... < t_n < ..., t_n \to +\infty \text{ such that } p \in C_c^\infty([t_n, t_{n+1}] \times \mathbb{R}^2), \\ & n = 0, 1, ..., (\rho, m) \text{ solves } (1.1) - (1.2) \text{ with } m = -\nabla \cdot (\rho \nabla p) \Big\}. \end{aligned} \tag{1.7}$$

Above, by writing $p \in C_c^\infty([t_n, t_{n+1}] \times \mathbb{R}^2), n = 0, 1, ...,$ we mean that $p \in C^\infty((t_n, t_{n+1}) \times \mathbb{R}^2)$ and it extends to a function in $C_c^\infty([t_n, t_{n+1}] \times \mathbb{R}^2)$ for every n. However p may be discontinuous at the points t_n. We set

$$\mathcal{K}_{\rho_0} = \Big\{ \rho(\cdot) \in \mathcal{K}; \quad \rho(0) = \rho_0 \Big\}. \tag{1.8}$$

We are interested in variational (or control) problems of the type

$$\begin{aligned} f(\rho_0) \ &:= \ R_\alpha h(\rho_0) \\ &= \ \sup\{ \int_0^\infty e^{-\alpha^{-1} s} \Big(\alpha^{-1} h(\rho(s)) - L(\rho(s), \dot{\rho}(s)) \Big) ds : \rho \in \mathcal{K}_{\rho_0} \} \\ &= \ \sup\{ \int_0^\infty \alpha^{-1} e^{-\alpha^{-1} s} \Big(h(\rho(s)) - A_s(\rho) \Big) ds : \rho \in \mathcal{K}_{\rho_0} \}, \end{aligned} \tag{1.9}$$

where $\alpha > 0$ and h is bounded from above. We call f the value function.

1.1. **Notation.** For $\rho, \gamma \in \mathcal{P}_2(\mathbb{R}^2)$, we set

$$\Gamma(\rho,\gamma) := \{\pi(dx,dy) \in \mathcal{P}(\mathbb{R}^2 \times \mathbb{R}^2) : \pi(dx,\mathbb{R}^2) = \rho(dx), \pi(\mathbb{R}^2, dy) = \gamma(dy)\}.$$

The Wasserstein order 2-metric d on $\mathcal{P}_2(\mathbb{R}^2)$ is defined as

$$d^2(\rho,\gamma) := \inf \Big\{ \int_{\mathbb{R}^2\times\mathbb{R}^2} |x-y|^2 \pi(dx,dy) : \pi \in \Gamma(\rho,\gamma) \Big\}. \tag{1.10}$$

$(\mathcal{P}_2(\mathbb{R}^2), d)$ is a non-locally compact, complete separable metric space (e.g. Chapter 7 of Ambrosio, Gigli and Savaré [3]). We define

$$\|m\|^2_{-1,\rho} := \sup_{\varphi \in C_c^\infty(\mathbb{R}^2)} \{2\langle \varphi, m\rangle - \int_{\mathbb{R}^2} |\nabla \varphi|^2 d\rho\}, \quad \forall m \in \mathcal{D}'(\mathbb{R}^2). \tag{1.11}$$

and

$$H_{-1,\rho}(\mathbb{R}^2) := \{m \in \mathcal{D}'(\mathbb{R}^2) : \|m\|_{-1,\rho} < \infty\}. \tag{1.12}$$

Appendix D in Feng and Kurtz [11] discusses some properties of this space and its relation with the 2-Wasserstein space $(\mathcal{P}_2(\mathbb{R}^2), d)$.

For a function $f : \mathcal{P}_2(\mathbb{R}^d) \mapsto (-\infty, +\infty]$, its effective domain is

$$D(f) := \{\rho \in \mathcal{P}_2(\mathbb{R}^d) : f(\rho) < +\infty\}.$$

Let $E := \mathcal{P}_2(\mathbb{R}^2)$, then (E, d) is a complete separable metric space and this will be our state space throughout the paper.

We define moment functionals

$$M_p(\rho) := \int_{\mathbb{R}^2} |x|^p d\rho, \quad p > 0, \tag{1.13}$$

and the Fisher information functional

$$I(\rho) := \int_{\mathbb{R}^2} \frac{|\nabla \rho|^2}{\rho} dx = 4 \int_{\mathbb{R}^2} |\nabla \sqrt{\rho}|^2 dx, \tag{1.14}$$

$I(\rho) = +\infty$ if ρ does not have a Lebesgue density. The equivalence of the two different expressions above for I, as well as a number of other properties, can be found in Appendix D.6 in [11]. We also define the entropy functional $s : \mathcal{P}_2(\mathbb{R}^2) \mapsto \overline{\mathbb{R}}$,

$$s(\rho) := \int_{\mathbb{R}^2} \rho \log \rho \, dx, \tag{1.15}$$

and $s(\rho) = +\infty$ when ρ does not have Lebesgue density. To see that the above is well defined, we note the following estimate (see e.g. a remark on page 9 of Jordan, Kinderlehrer and Otto [18]):

$$\int_{\mathbb{R}^2} |0 \vee \log \rho| \rho(dx) \le C \int_{\mathbb{R}^2} e^{-|x|/2} dx + \int_{\mathbb{R}^2} |x| \rho(dx). \tag{1.16}$$

Next, we define an internal energy functional on $\mathcal{P}_2(\mathbb{R}^d)$ by

$$e(\rho) := \frac{1}{2}\langle N * \rho, \rho\rangle = \frac{1}{2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} N(x-y)\rho(dx)\rho(dy). \tag{1.17}$$

A cautionary note is necessary here. Consider a more general situation of $\rho \in \mathcal{M}(\mathbb{R}^2)$, the space of finite signed measures. Formally, it seems that $K = \nabla(-\Delta)^{-1}$ and

$$e(\rho) = \frac{1}{2}\langle(-\Delta)^{-1}\rho, \rho\rangle = \frac{1}{2}\langle(-\Delta)^{-1}\rho, (-\Delta)(-\Delta)^{-1}\rho\rangle = \frac{1}{2}\|\nabla(-\Delta)^{-1}\rho\|_2^2 = \frac{1}{2}\int_{\mathbb{R}^2} |u(x)|^2 dx.$$

The above symbolic calculation is not true because in dimension two, $(-\Delta)^{-1}\rho$ may not decay at infinity sufficiently fast to allow integration by parts. For instance, if ρ has compact support, (3.15) in Majda and Bertozzi [20] gives the following estimate

$$u(x) = \frac{1}{2\pi}\frac{(-x_2, x_1)}{|x|^2}\rho(\mathbb{R}^2) + O(|x|^{-2}), \quad \text{for large } |x|.$$

This implies that $\int_{\mathbb{R}^2} |u(x)|^2 dx = \infty$ if $\rho(\mathbb{R}^2) \neq 0$. However, at least when $\rho \in \mathcal{P}_2(\mathbb{R}^2)$, $e(\rho) \in (-\infty, +\infty]$ is always well-defined. This follows easily from the inequality $N(z) \geq -C(1+|z|)$ for some $C > 0$.

Finally, the velocity field $u = (u_1, u_2)$ defined by (1.2) is a vector. In the following, we denote the 2×2 matrix $Du = (\nabla u_1, \nabla u_2)$.

1.2. **Relation with 2-D incompressible Navier-Stokes equation.** In this section we discuss, rather informally, the motivation behind our optimal control problem and its connection to a particle model.

Consider the 2-D Navier-Stokes equations

$$\partial_t u + (u \cdot \nabla)u + \nabla p = \nu \Delta u, \quad \text{div } u = 0, \quad \lim_{|x|\to\infty} u(x) = 0, \tag{1.18}$$

where $u = u(t, x) = (u_1(t, x_1, x_2), u_2(t, x_1, x_2))$. Taking

$$\mu = \text{curl}_x u = \partial_{x_2} u_1 - \partial_{x_1} u_2,$$

we arrive at its vorticity formulation

$$\partial_t \mu + \text{div}(\mu u) = \nu \Delta \mu, \quad u = -K^\perp * \mu. \tag{1.19}$$

The above is essentially (1.1) with m set to zero. In fact, a refined model at a particle level can be introduced for which (1.19) is just the mean-field (i.e. law of large number) limit equation for a collection of stochastic vortex particles interacting in a particular manner. Ideas of this kind, in the deterministic setting (i.e. with $\nu = 0$), can be found in Chorin [7], Lions [19] for the study of incompressible Euler equation (see also Majda and Bertozzi [20]). In particular, Lions [19] outlines a heuristic procedure to motivate the formulation of a class of variational problems associated with large time coherent structures for such flows.

The stochastic interacting particle system was first introduced in [21] and was later studied in [22, 23, 24, 25]. Various results about convergence of laws of the empirical measures (as the number of particles goes to infinity) to the law of the solution of the vorticity equation, and the propagation of chaos property, have been proved in the above papers. We refer the readers there for details and further references.

The stochastic particle particle system is constructed in the following way. Let there be n_1 particles $(X_1, \ldots, X_{n_1})$ modeling vortices rotating counter-clockwise, and let there be n_2 particles $(Y_1, \ldots, Y_{n_2})$ modeling vortices rotating clockwise, and let $n = n_1 + n_2$ be the total number of particles. As we move to the macroscopic level, we lose track of individual particles and only see collective effects given by number density of two types of vortices, which is represented by the probability measure

$$\rho_n(t, dx, dy) = \frac{1}{n_1 \times n_2} \sum_{i,j} \delta_{(X_i(t), Y_j(t))}(dx, dy). \tag{1.20}$$

Two useful functionals of ρ_n are

$$\rho_{n,+}(t, dx) = n_1^{-1} \sum_{i=1}^{n_1} \delta_{X_i(t)}(dx), \quad \rho_{n,-}(t, dy) = n_2^{-1} \sum_{i=1}^{n_2} \delta_{Y_i(t)}(dy). \tag{1.21}$$

They are the marginal probabilities of $\rho_n(t, dx, dy)$. The dynamic of all particles is defined through a system of stochastic differential equations

$$dX_i = u_{\rho_n}(X_i)dt + \sqrt{2\nu}dB_i, \tag{1.22}$$

$$dY_j = u_{\rho_n}(Y_j)dt + \sqrt{2\nu}dW_j, \tag{1.23}$$

where the density dependent vector field

$$\begin{aligned} u_{\rho_n}(z) &= -J\nabla N_n * (\frac{n_1}{n}\rho_{n,+} - \frac{n_2}{n}\rho_{n,-})(z) \\ &= -n^{-1} \sum_{i=1}^{n_1} J\nabla N_n(z - X_i) + n^{-1} \sum_{j=1}^{n_2} J\nabla N_n(z - Y_j). \end{aligned}$$

In the above, $(B_1, \dots, B_{n_1}, W_1, \dots, W_{n_2})$ is an $\mathbb{R}^{2n}$-dimensional standard Brownian motion and N_n is some Lipschitz smooth potential approximating the Newtonian potential N in (1.3) with the property $\nabla N_n(0) = 0$.

By Ito's formula, for each $\varphi \in C_c^2(\mathbb{R}^2 \times \mathbb{R}^2)$,

$$d\langle \varphi, \rho_n(t)\rangle = \langle \nabla_x \varphi(x, y) \cdot u_{\rho_n}(x) + \nabla_y \varphi(x, y) \cdot u_{\rho_n}(y), \rho_n\rangle + \nu\langle \Delta_{(x,y)}\varphi, \rho_n\rangle dt + dM_n^\varphi(t)$$

where M_n^φ is a martingale with co-quadratic variation

$$d[M^{\varphi_1}, M^{\varphi_2}](t) = 2\nu \frac{1}{n_1 \times n_2} \langle \nabla_{(x,y)}\varphi_1 \cdot \nabla_{(x,y)}\varphi_2, \rho_n\rangle dt. \tag{1.24}$$

Define

$$\begin{aligned} A_n f(\rho) &= \langle \nabla_x \frac{\delta f}{\delta \rho}(x, y) \cdot u_\rho(x) + \nabla_y \frac{\delta f}{\delta \rho}(x, y) \cdot u_\rho(y), \rho\rangle + \nu\langle \Delta_{(x,y)} \frac{\delta f}{\delta \rho}, \rho\rangle \\ &\quad + (n_1 \times n_2)^{-1}\nu \sum_{l,m=1}^{k} \partial_{lm}\psi(\langle \varphi_1, \rho\rangle, \dots, \langle \varphi_k, \rho\rangle)\langle \nabla\varphi_l \nabla\varphi_m, \rho\rangle, \end{aligned} \tag{1.25}$$

for smooth test functions

$$f(\rho) = \psi(\langle \varphi_1, \rho\rangle, \dots, \langle \varphi_k, \rho\rangle), \psi \in C^2(\mathbb{R}^k), \varphi_i \in C_c^\infty(\mathbb{R}^2 \times \mathbb{R}^2), k = 1, 2, \dots. \tag{1.26}$$

Therefore, $\rho_n(t)$ as a probability measure valued process, with trajectories in $C([0, \infty); \mathcal{P}(\mathbb{R}^2 \times \mathbb{R}^2))$, solves the *martingale problem* given by

$$f(\rho_n(t)) - \int_0^t A_n f(\rho_n(s))ds = \text{martingale}. \tag{1.27}$$

The process $\rho_n(\cdot)$ is Markovian. Indeed, the form of (1.25) suggests that the projection of ρ_n to "lower dimensional" sub-space of probability marginals (ρ_n^+, ρ_n^-), where

$$\rho_n^+(t,dx) := \rho_n(t,dx,\mathbb{R}^2), \quad \rho_n^-(t,dy) := \rho_n(t,\mathbb{R}^2,dy)$$

forms a system of Markov processes as well. Its martingale problem generator can be identified through

$$\begin{aligned} (1.28) \qquad & A_n g(\rho_+,\rho_-) \\ = \quad & \langle \nabla \frac{\delta g}{\delta \rho_+} \cdot u_\rho, \rho_+ \rangle + \langle \nabla \frac{\delta g}{\delta \rho_-} \cdot u_\rho, \rho_- \rangle + \nu \langle \Delta \frac{\delta g}{\delta \rho_+}, \rho_+ \rangle + \nu \langle \Delta \frac{\delta g}{\delta \rho_-}, \rho_- \rangle \\ & + n_1^{-1} \nu \sum_{l,m=1}^{k} \partial_{lm} \psi(\langle \varphi_1, \rho_+ \rangle, \ldots, \langle \phi_k, \rho_- \rangle) \langle \nabla \varphi_l \nabla \varphi_m, \rho_+ \rangle \\ & + n_2^{-1} \nu \sum_{l,m=k+1}^{2k} \partial_{lm} \psi(\langle \varphi_1, \rho_+ \rangle, \ldots, \langle \phi_k, \rho_- \rangle) \langle \nabla \phi_l \nabla \phi_m, \rho_- \rangle, \end{aligned}$$

for smooth test functions

$$(1.29) \qquad g(\rho) = \psi(\langle \varphi_1, \rho_+ \rangle, \ldots, \langle \varphi_k, \rho_+ \rangle, \langle \phi_1, \rho_- \rangle, \ldots, \langle \phi_k, \rho_- \rangle),$$

where $\psi \in C^2(\mathbb{R}^{2k}), \varphi_i, \phi_j \in C_c^\infty(\mathbb{R}^2), k = 1, 2, \ldots$. Assuming $n^{-1} n_i \to \Lambda_i$, and considering

$$\mu_n := \frac{n_1}{n} \rho_{n,+} - \frac{n_2}{n} \rho_{n,-},$$

then $\mu_n \to \mu$ at least formally, where μ solves (1.19). To simplify, we consider the special situation of counter-clockwise rotation only, i.e. if $\Lambda_- = 0$, the limiting martingale problem for (1.27) becomes the Schwartz distributional formulation of (1.1)-(1.2) with $m = 0$. This corresponds to letting $n = n_1$ and eliminating everything involving ρ_- component in the above g and $A_n g$.

Therefore, at least from a mathematical point of view, model (1.22)-(1.23) contains more information than the usual Navier-Stokes equation (1.19), which only expresses the (already averaged) mean-field behavior. One example where such an additional information can be useful is in computing a path space level entropy (rescaled by particle numbers) in the sense of Boltzmann. The rigorous justification of such computation belongs to the probability theory of large deviations. To be precise, we want to identify a function $\mathbb{S}$ which takes values in $[0, +\infty]$ and is defined over Borel sets A in an appropriately defined path space, such that

$$(1.30) \; -\mathbb{S}(A^\circ) \le \liminf_{n\to\infty} n^{-1} \log \mathbb{P}(\rho_n(\cdot) \in A^\circ) \le \limsup_{n\to\infty} n^{-1} \log \mathbb{P}(\rho_n(\cdot) \in \bar{A}) \le -\mathbb{S}(\bar{A}).$$

In particular, $\mathbb{S}$ has a "density" $\mathbb{I}$ which is known as the action functional (or rate function):

$$\mathbb{S}(A) = \inf_{\rho(\cdot)\in A} \mathbb{I}(\rho(\cdot)),$$

with

$$(1.31) \qquad \mathbb{I}(\rho(\cdot)) = -\lim_{\epsilon\to 0+} \lim_{n\to\infty} n^{-1} \log \mathbb{P}(\rho_n(\cdot) \in B_\epsilon(\rho(\cdot))) = \int_0^T L(\rho(t), \dot{\rho}(t)) dt.$$

We note that $\mathbb{S}$ is really a quantity arising from the nonlinear scaling behavior of the stochastic processes (1.21), it contain information about the limit process and cannot be obtained from the limiting deterministic 2-D incompressible Navier-Stokes equations or their vorticity formulation alone. Justification of the limit (1.30) and identification of the rate function

$\mathbb{I}$ (and hence $\mathbb{S}$) are related to variational problems of optimal control nature, which are (1.1)-(1.5) in this specific context. They can be studied by establishing well posed-ness of the Hamilton-Jacobi-Bellman (HJB) equation

$$(I - \alpha H)f = h \tag{1.32}$$

for $0 < \alpha < \alpha_0$ for some $\alpha_0 > 0$, and for a sufficiently large class of h, where the first order differential operator $Hg(\rho) := H(\rho, \text{grad} g(\rho))$ is and appropriate limit of

$$H_n g = n^{-1} e^{-ng} A_n e^{ng}.$$

See Feng and Kurtz [11] for a general method developed in the context of metric-space valued Markov processes.

The above H is defined for smooth test functions g of the form (1.29). One can then extend H in a viscosity extension sense to the H defined in Section 4. This helps when a convergence theory (H_n converges to H) and a well posed-ness theory for the limiting equation need to be studied in one framework altogether. A general method for doing this is described on pages 111-113, and illustrated in Section 13.3.3 of [11] for a related model. We do not try to investigate if this approach can be applied here. Instead, we are only concerned with the well posed-ness theory for (1.32). We do it by a direct approach which is more in line with a more classical viscosity solution approach. We introduce another class of test functions (Definition 4.1) on which H (from Section 4) is defined. Our notion of viscosity solution applies to discontinuous functions. We prove a comparison principle for the general case of discontinuous sub/super-solutions, and then show, using dynamic programming principle arguments, that the value function is the unique viscosity solution of the HJB equation. This has an advantage of avoiding a rather delicate issue of the continuity of the value function, which we obtain as a byproduct of the proof of comparison principle. Another feature of our argument is the use of Borwein-Preiss [6] variational principle to produce extremal points. This allows us to do perturbed optimization based on the use of d^2 as part of test functions. The whole paper relies heavily on the abstract calculus in the Wasserstein space based on mass transport techniques, whose rigorous theory can be found in [3]. Especially, the chain rule formula (Appendix D) is critical for our purposes. An equation of type (1.32) for which $K^\perp$ (see the definition of H) is replaced by a smooth kernel has been investigated successfully in the space of measures by Feng and Katsoulakis [10] (see also [11]). Here, the singularity of K is the main source of technical difficulties. However in the proof of comparison, many techniques of [10] still apply after more or less extensive modifications. To obtain key estimates we have to proceed through rather involved approximations and make extensive use of Sobolev type inequalities to estimate orders of approximation. We mention that existence of viscosity solutions was not investigated in [10]. In [11], it was indirectly handled using Markov processes properties and large deviation techniques.

Finally, we remark that equations of type (1.1)-(1.2), with $K^\perp$ replaced by various other kernels and with possibly different diffusion terms, have recently appeared in modeling of biological aggregation (swarming, schooling, flocking, etc.) and other pattern formation models that incorporate aggregative and dispersive types of behavior. We refer for instance to [5, 28] and the references therein for more. In these models the motions of individuals are represented by deterministic or stochastic particle systems similar to the one described here in Section 1.2, and the continuum equations are their mean field limits. The general framework developed in this paper could potentially apply to such problems which have similar Hamiltonian/gradient structure. This however has to be investigated on a case by case basis as our analysis of the control problem and the associated HJB equation relies heavily on many special properties of equation (1.1)-(1.2), in particular the orthogonality of J-$\text{grad}_\rho e$ and $\text{grad}_\rho s$.

1.3. **Solution of the control equation.** We make precise sense of what we mean by (1.1-1.2).

Definition 1.1 (Weak Solution)**.** (ρ, m), $\rho : [0,+\infty) \mapsto \mathcal{P}_2(\mathbb{R}^2)$, is said to be a (weak) solution to (1.1) provided that

(1) $\rho \in C([0,\infty), \mathcal{P}(\mathbb{R}^2))$;
(2) for each $0 < t < \infty$, $\rho \in L^2([0,t], L^2(\mathbb{R}^2))$ and $m \in L^2([0,t], H_{-1,\rho}(\mathbb{R}^2))$;
(3) for all $0 \le s < t < \infty$ and $\varphi = \varphi(t,x) \in C_c^\infty([0,\infty) \times \mathbb{R}^2)$,

$$\begin{aligned}(1.33)\quad & \langle \varphi(t), \rho(t)\rangle - \langle \varphi(s), \rho(s)\rangle \\ &= \int_s^t \Big(\int_{x\in\mathbb{R}^2} (\partial_r \varphi(r,x) + \nabla\varphi(r,x)\cdot u(r,x) + \nu\Delta\varphi(t,x))\rho(r,dx) + \langle m(r), \varphi(r)\rangle \Big) dr.\end{aligned}$$

If we restrict (1.1) to the time interval $[0,T]$ we require test functions φ to be in $C_c^\infty([0,T] \times \mathbb{R}^2)$.

We recall that in the definition above $\langle m(r), \varphi(r)\rangle$ is understood in the distribution sense. Since (see [11], Appendix D5) there exists $v(t)$ such that $m(r) = -\nabla\cdot(\rho(r)v(r))$, where $\int |v|^2\rho < +\infty$, we have $\langle m(r), \varphi(r)\rangle = \int \nabla\varphi(r,x)\cdot v(r,x)\rho(r,dx)$.

Remark 1.2. Recall that $u = -K^\perp * \rho$. If for each $0 < t < \infty$, $\rho \in L^2([0,t], L^2(\mathbb{R}^2))$ then by (7.27) we know that $\rho|u| \in L^1([0,t]\times\mathbb{R}^2)$, and it then follows that the term $\int_s^t \int_{\mathbb{R}^2} (\nabla\varphi)\cdot u d\rho dr$ in (1.33) is well defined.

Existence and regularity estimates for solution to (1.1) are established in Section 2.

1.4. **Main result.** Let H, the test functions and the notion of viscosity solution for (1.32) be defined according to Section 4. The main result of this article is the following well posed-ness of (1.32), which follows from the conclusions of Corollary 5.3 and Theorem 6.3.

Theorem 1.3. *Let $\alpha > 0$ and $h \in C_b(E)$ be uniformly continuous on finite level sets of $s + M_2$. Then equation (1.32) has a unique bounded viscosity solution which is given by the value function $f = R_\alpha h$ in (1.9). The solution is uniformly continuous on finite level sets of $s + M_2$.*

2. Existence and regularity for the controlled partial differential equation

Let $p \in C_c^\infty([0,\infty)\times\mathbb{R}^2)$. We define

$$m := -\nabla\cdot(\rho\nabla p), \quad v(t,x) := \nabla_x p(t,x).$$

We construct a solution of (1.1)-(1.2) with initial condition $\rho(0) = \rho_0$ through a stochastic particle method and prove some regularity results. For results on approximations of the uncontrolled version of (1.1)-(1.2) by an interacting particle system we refer to [21, 22, 23, 24, 25] and for general results about existence and uniqueness of solutions of the 2-D vorticity equation we refer to [13, 20] and to [14], where existence of solutions was shown when the initial vorticity was a finite Radon measure.

We introduce a controlled version of (1.22-1.23) but with only counter-clockwise rotating particles

$$dX_i = u_{G_n * \rho_n(t)}(X_i)dt + v(t, X_i)dt + \sqrt{2\nu}dB_i, \quad X_i(0) = X_{0i},\ i = 1, ..., n, \tag{2.1}$$

where $X_{0i}, i = 1, 2, ...,$ are i.i.d. random variables with law ρ_0 defined on the same probability space as the Brownian motions $B_i, i = 1, 2, ...$ Moreover, we choose a specific $N_n := G_n * N$ with G_n defined as follows. Let $J(z) : \mathbb{R}^2 \mapsto \mathbb{R}_+$ be a radial symmetric, C^∞ function with support on $[-1, 1]^2$, $\int J(z)dz = 1$. Let $J_\delta(z) := \delta^{-2}J(\delta^{-1}z)$ and $\delta_n \to 0$ be such that $n\delta_n^2 \to \infty$ (e.g. $\delta_n = n^{-1/3}$). With a slight abuse of notation, we also denote $J_n = J_{\delta_n}$. Finally, $G_n := J_n * J_n$. Therefore

$$u_{G_n * \rho_n(t)}(x) = -(K^\perp * G_n) * \rho_n(x).$$

Recall that solutions of (2.1) define measures

$$\rho_n(t) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(t)}(dx).$$

Define

$$e_n(\rho_n(t)) := \frac{1}{2}\langle N_n * \rho_n(t), \rho_n(t)\rangle = \frac{1}{2n^2} \sum_{i,j=1}^{n} N_n(X_i(t) - X_j(t)). \tag{2.2}$$

Lemma 2.1. *Let $M_2(\rho_0) < +\infty$ and $s(\rho_0) < +\infty$. Then $\lim_{n\to\infty} d(\rho_n(0), \rho_0) = 0$ a.s., and*

$$\mathbb{E}[M_2(\rho_n(0))] = M_2(\rho_0), \tag{2.3}$$

$$\lim_{n\to+\infty} \mathbb{E}[e_n(\rho_n(0))] = e(\rho_0). \tag{2.4}$$

Proof. For every φ satisfying $(1 + |x|^2)^{-1}\varphi \in C_b(\mathbb{R}^2)$, we have

$$E[|\varphi(X_{0i})|] \le C[1 + \mathbb{E}[|X_{0i}|^2]] = C[1 + M_2(\rho_0)] < \infty.$$

By strong law of large numbers,

$$\lim_{n\to\infty} \int_{\mathbb{R}^2} \varphi(x)\rho_n(0)(dx) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \varphi(X_{i0}) = \int_{\mathbb{R}^2} \varphi(x)\rho_0(dx) \quad a.s.$$

This implies that ρ_n converges to ρ_0 in the 2-Wasserstein distance almost surely.

Equality (2.3) follows by direct verification

$$\mathbb{E}[M_2(\rho_n(0))] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[|X_{0i}|^2] = M_2(\rho_0).$$

As regards (2.4), using (2.2) we have

$$\mathbb{E}[e_n(\rho_n(0))] = \frac{1}{2n^2} \sum_{i,j=1, i\neq j}^{n} \mathbb{E}[N_n(X_{0i} - X_{0j})] + \frac{1}{2n}N_n(0) = \frac{n^2 - n}{2n^2}e_n(\rho_0) + \frac{1}{2n}N_n(0).$$

Since $|J_\delta| \le C\delta^{-2}$ and is equal to zero when $|x| > \delta$, an easy calculation involving a change to polar coordinates gives us $|N_n(0)| \le -C \log \delta_n$. Moreover, by Lemma 7.4, $e_n(\rho_0) \to e(\rho_0)$ as $n \to +\infty$. Therefore (2.4) follows. $\square$

By the existence theory for stochastic differential equations with Lipschitz coefficients, the above finite system has a unique solution on any finite time interval.

Then by Ito's formula, for each $0 < t \leq T$, $\varphi \in C_c^2([0,T] \times \mathbb{R}^2)$,

$$
\begin{aligned}
(2.5) \qquad d\langle \varphi, \rho_n(t)\rangle &= \langle \partial_t \varphi, \rho_n\rangle dt + \langle -\mathrm{div}(\rho_n(u_{G_n*\rho_n} + v)) + \nu\Delta\rho_n, \varphi\rangle dt + dM_n^\varphi(t) \\
&= \langle \partial_t \varphi + (u_{G_n*\rho_n} + v)\cdot\nabla\varphi + \nu\Delta\varphi, \rho_n\rangle dt + dM_n^\varphi(t),
\end{aligned}
$$

where

$$M_n^\varphi(t) = \sqrt{2\nu} n^{-1} \sum_{i=1}^n \int_0^t \nabla\varphi(r, X_i(r)) \cdot dB_i(r).$$

We recall that for $0 \leq s < t \leq T$, $M_{n,s}^\varphi(t) := M_n^\varphi(t) - M_n^\varphi(s)$ is a martingale for $t \geq s$ with quadratic variation

$$(2.6) \qquad [M_{n,s}^\varphi](t) = 2\nu n^{-1} \int_s^t \int_{\mathbb{R}^2} |\nabla\varphi(r,x)|^2 \rho_n(r,dx)dr.$$

Taking expectation in the integral form of (2.5) we have for $0 \leq s < t \leq T$

$$(2.7) \qquad \mathbb{E}\langle\varphi(t), \rho_n(t)\rangle - \mathbb{E}\langle\varphi(s), \rho_n(s)\rangle = \mathbb{E}\int_s^t [\langle \varphi_r + \nabla\varphi\cdot(u_{G_n*\rho_n} + v) + \nu\Delta\varphi, \rho_n(r)\rangle]\, dr.$$

In the following, when we view $\mathcal{P}(\mathbb{R}^d)$ as a metric space, the metric is always taken to be a fixed one that gives the weak convergence of probability measure topology (i.e. the narrow convergence topology); when we view $\mathcal{P}_p(\mathbb{R}^d)$ as a metric space, however, the metric is always the p-Wasserstein metric.

Lemma 2.2. *Let $M_2(\rho_0) < +\infty, s(\rho_0) < +\infty$, $T > 0$ and $p \in C_c^\infty([0,\infty) \times \mathbb{R}^2)$ be given. Then there exists $\rho(\cdot) \in C([0,T]; \mathcal{P}(\mathbb{R}^2))$ such that $(\rho, -\nabla\cdot(\rho\nabla p))$ is a weak solution of (1.1) on $[0,T]$ and $\rho(0) = \rho_0$. Moreover for every $0 \leq t \leq T$*

$$(2.8) \qquad M_2(\rho(t)) \leq (M_2(\rho_0) + (\|v\|_\infty^2 + 4\nu + C_2)t)e^t.$$

and

$$(2.9) \qquad \int_0^T \|\rho(r)\|_2^2 dr < +\infty.$$

In particular $M_2(\rho(t))$ is continuous at 0. We also have $\rho(\cdot) \in C([0,T]; \mathcal{P}_q(\mathbb{R}^2))$ for $0 \leq q < 2$.

Proof. We will show that the family $\{\rho_n(\cdot) : n = 1, 2, \ldots\}$ as $C([0,T]; \mathcal{P}(\mathbb{R}^2))$ valued random variables is tight and we will obtain $\rho(\cdot)$ in a limit which will be made precise later.

We consider second moment estimates first. We note that the set

$$K_C = \{\rho \in \mathcal{P}(\mathbb{R}^2) : M_2(\rho) \leq C\}$$

is compact in $\mathcal{P}(\mathbb{R}^2)$ with the weak (narrow) convergence topology. By Ito's formula

$$
\begin{aligned}
(2.10) \qquad dM_2(\rho_n(t)) &= 2\langle x\cdot(u_{G_n*\rho} + v), \rho_n\rangle dt + 4\nu dt + dN_n(t) \\
&\leq M_2(\rho_n(t))dt + \int_{\mathbb{R}^2} |v(t,x)|^2 \rho_n(t,dx)dt + \epsilon_n(t)dt + 4\nu dt + dN_n(t)
\end{aligned}
$$

where

$$(2.11) \qquad \epsilon_n(t) := -\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} (x-y)\cdot K_n^\perp(x-y)\rho_n(t,dx)\rho_n(t,dy)$$

and local martingale $N_n(t)$ has quadratic variation

$$[N_n](t) = 8\nu n^{-1} \int_0^t \int_{\mathbb{R}^2} |x|^2 \rho_n(r, dx) dr.$$

The inequality above follows because, denoting $K_n = G_n * K$,

$$\begin{aligned}
& \int_{\mathbb{R}^2} x \cdot u_{G_n * \rho_n}(x) \rho_n(t, dx) \\
= & -\int_{\mathbb{R}^2} x \cdot (K_n^\perp * \rho_n)(x) \rho_n(t, dx) \\
= & -\int_{\mathbb{R}^2} \int_{\mathbb{R}^2} ((x-y)+y) \cdot K_n^\perp(x-y) \rho_n(t, dx) \rho_n(t, dy) \\
= & \ \epsilon_n(t) - \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} y \cdot K_n^\perp(x-y) \rho_n(t, dx) \rho_n(t, dy) \\
= & \ \epsilon_n(t) - \int_{\mathbb{R}^2} y \cdot u_{G_n * \rho_n}(y) \rho_n(t, dy),
\end{aligned}$$

where we used the fact that $K_n^\perp(-z) = -K_n^\perp(z)$. If we replaced K_n in (2.11) by K, then $\epsilon_n(t)$ would become zero because $z \cdot K^\perp(z) = 0$. This identity does not hold when K is replaced by K_n. However, we have the following estimate. We notice that

$$\begin{aligned}
z \cdot K_n^\perp(z) &= \int_{\mathbb{R}^2} z \cdot K^\perp(z-y) G_n(y) dy \\
&= \int_{\mathbb{R}^2} y \cdot K^\perp(z-y) G_n(y) dy = \int_{|y| \leq 2/n} y \cdot K^\perp(z-y) G_n(y) dy.
\end{aligned}$$

Therefore

$$\begin{aligned}
|\epsilon_n(t)| &\leq \|z \cdot K_n^\perp(z)\|_\infty \leq \sup_{|y| \leq 2/n} |y G_n(y)| \sup_{z \in \mathbb{R}^2} \int_{|y| \leq 2/n} |K^\perp(z-y)| dy \\
&\leq C_1 n \sup_{z \in \mathbb{R}^2} \int_{|y| \leq 2/n} \frac{1}{|z-y|} dy \leq C_1 n \int_{|y| \leq 2/n} \frac{1}{|y|} dy \leq C_2.
\end{aligned}$$

Denote the stopping time

$$\tau_{n,C} = \inf\{t > 0 : M_2(\rho_n(t)) > C\}.$$

Then by the optional sampling theorem and Grownwall's inequality, we have

$$\mathbb{E}[M_2(\rho_n(\tau_{n,C} \wedge T))] \leq (M_2(\rho_0) + (\|v\|_\infty^2 + 4\nu + C_2)T) e^T$$

implying

$$\begin{aligned}
\lim_{C \to \infty} \sup_n \mathbb{P}(\tau_{n,C} \leq T) &\leq \lim_{C \to \infty} \sup_n \mathbb{P}(M_2(\rho_n(\tau_{n,C} \wedge T)) \geq C) \\
&\leq \lim_{C \to \infty} \sup_n C^{-1} \mathbb{E}[M_2(\rho_n(\tau_{n,C} \wedge T))] = 0.
\end{aligned}$$

In the above, we note that for each n fixed, since v and $K * G_n$ are globally Lipschitz, $M_2(\rho_n(t)) \in C([0,T]; \mathbb{R})$. We obviously also have for $t \leq T$

$$\mathbb{E}[M_2(\rho_n(t))] \leq (M_2(\rho_0) + (\|v\|_\infty^2 + 4\nu + C_2)t) e^t. \tag{2.12}$$

In summary, a compact containment condition is satisfied, i.e.

$$\lim_{C\to\infty}\sup_n \mathbb{P}(\exists t\in[0,T],\rho_n(t)\notin K_C)=0.$$

Next, we prove that for each $\varphi\in C_c^\infty(\mathbb{R}^2)$, $\{\rho_n^\varphi(t)=\langle\varphi,\rho_n(t)\rangle : n=1,2\ldots\}$ is tight in $C([0,T];\mathbb{R})$. Then by Jakubowski [17], we conclude that $\{\rho_n(\cdot): n=1,2,\ldots\}$ is tight in $C([0,T];\mathcal{P}(\mathbb{R}^2))$. Since $|\langle\varphi,\rho\rangle|\le\|\varphi\|_\infty$, it is sufficient to have the following uniform modulus of continuity estimate

$$\mathbb{E}[|\langle\varphi,\rho_n(t+h)-\rho_n(t)\rangle|]\le Ch^{1/4}. \tag{2.13}$$

For the sufficiency, see Theorems 8.6 and 8.8 in Chapter 3 of Ethier and Kurtz [9]. We use the martingale characterization (2.5) to obtain such an estimate. However, because of the singularity in $u_{J_n*\rho_n}$, we need another energy estimate first.

Notice that $dN_n(X_i(t)-X_i(t))=0$ and that $\|J_n\|_\infty\le C_3\delta_n^{-2}$ hence $|G_n(0)|\le C_3\delta_n^{-2}$. By Ito's formula, we have (recall definition (2.2))

$$\begin{aligned}
(2.14)\quad de_n(\rho_n(t)) &= \langle(\nabla N_n*\rho_n(t))\cdot(\nabla^\perp N_n*\rho_n(t)+v(t)),\rho_n\rangle dt\\
&\qquad +\nu n^{-2}\sum_{i\neq j}\Delta N_n(X_i-X_j)dt+dM_n^e(t)\\
&= \langle(\nabla N_n*\rho_n(t))\cdot v,\rho_n\rangle dt-\nu n^{-2}\sum_{i\neq j}G_n(X_i-X_j)dt+dM_n^e(t)\\
&\le C_{1,\nu}\|J_n*\rho_n\|_2^{3/2}dt-\nu\|J_n*\rho_n\|_2^2dt+C_3\nu n^{-1}\delta_n^{-2}dt+dM_n^e\\
&\le C_{2,\nu}dt-\frac{\nu}{2}\|J_n*\rho_n\|_2^2dt+C_3\nu n^{-1}\delta_n^{-2}dt+dM_n^e.
\end{aligned}$$

In the above, the first inequality follows from

$$\begin{aligned}
(2.15)\qquad \langle(\nabla N_n*\rho_n)\cdot v,\rho_n\rangle &= \langle K*(J_n*\rho_n),J_n*(\rho_n v)\rangle\\
&\le \|(\chi_{\tilde{\mathcal{O}}})K*(J_n*\rho_n)\|_2\|J_n*(\rho_n v)\|_2\le C\|J_n*\rho_n\|_2^{1/2}\|J_n*\rho_n\|_2
\end{aligned}$$

where $\tilde{\mathcal{O}}=\{x:\mathrm{dist}(x,\mathcal{O})<1\}$ is the fattening of a bounded open subset $\mathcal{O}$ containing the support of v (equivalently, p) by the support of all mollifiers J_n, and χ_A is the characteristic function of a set A, and the last step above follows from (7.4).

The martingale

$$M_n^e(t)=\sqrt{2\nu}n^{-2}\sum_{i,j}\int_0^t\nabla N_n(X_i-X_j)dB_i=\sqrt{2\nu}n^{-1}\sum_i\int_0^t(\nabla N_n*\rho_n)(X_i)dB_i,$$

has quadratic variation

$$[M_n^e](t)=2\nu n^{-2}\sum_i\int_0^t|\nabla N_n*\rho_n(X_i(r))|^2dr=2\nu n^{-1}\int_{\mathbb{R}^2}\int_0^t|u_{G_n*\rho_n(r)}(x)|^2\rho_n(r,dx)dr.$$

Therefore, taking expectation in the integral form of (2.14), and combining the result with (2.12) yields

$$\mathbb{E}[e_n(\rho_n(T))+M_2(\rho_n(T)+\frac{\nu}{2}\int_0^T\|J_n*\rho_n(r)\|_2^2dr]\le\mathbb{E}[e_n(\rho_n(0))]+e^TM_2(\rho_0)+C_{\nu,p,T}. \tag{2.16}$$

Since polynomial growth at infinity dominates logarithmic growth, $e_n + M_2 \geq c > -\infty$, and thus we have

$$\frac{\nu}{2}\mathbb{E}[\int_0^T \|J_n * \rho_n(r)\|_2^2 dr] \leq \mathbb{E}[e_n(\rho_n(0))] + e^T M_2(\rho_0) + C_{\nu,p,T}. \tag{2.17}$$

Similarly to (2.15), we have

$$|\langle \nabla\varphi \cdot u_{G_n * \rho_n}, \rho_n\rangle| \leq C\|J_n * \rho_n\|_2^{3/2}.$$

Consequently

$$\mathbb{E}[\int_s^t |\langle \nabla\varphi \cdot u_{G_n * \rho_n}, \rho_n\rangle| dr] \leq C|t-s|^{1/4}.$$

Applying the above estimate to (2.5) we easily obtain (2.13).

As mentioned before, by a combination of Theorems 3.8.6 and 3.8.8 of [9] and by [17], we conclude that the family of probability distributions/laws on $\mathcal{P}(C([0,T];\mathcal{P}(\mathbb{R}^2)))$ for random variables $\{\rho_n(\cdot) : n = 1, 2, \ldots\}$ is tight. By Pohorov theorem, the family of probability laws is relatively compact in the topology of weak convergence of probability measures. We select a convergent subsequence and with a slight abuse of notation, we still label the subsequence by n. By the Skorohod representation theorem (e.g. Theorem 3.1.8 in [9]), we can construct a canonical probability space on which random variables $\rho, \tilde{\rho}_n, n = 1, 2, ...$ are defined, with the property that $\tilde{\rho}_n$ has the same law as ρ_n for $n = 1, 2, ...$ and such that

$$\tilde{\rho}_n \to \rho \quad a.s. \text{ in } C([0,T];\mathcal{P}(\mathbb{R}^2)) \text{ as } n \to +\infty. \tag{2.18}$$

We observe that (2.7), (2.12), (2.17) hold for $\tilde{\rho}_n$, while we cannot replace ρ by $\tilde{\rho}_n$ in (2.5). We will first pass to the limit in (2.12) and (2.17). Since M_2 is lower semi-continuous with respect to weak convergence of probability measures (i.e. narrow convergence), we have by Fatou's lemma $\mathbb{E}[M_2(\rho(t))] \leq \liminf_{n\to\infty} \mathbb{E}[M_2(\tilde{\rho}_n(t))]$ which, together with (2.12), gives us

$$\mathbb{E}[M_2(\rho(t))] \leq (M_2(\rho_0) + (\|v\|_\infty^2 + 4\nu + C_2)t)e^t. \tag{2.19}$$

By (2.4), $\mathbb{E}[e_n(\tilde{\rho}_n(0))] \to e(\rho_0)$. Finally, since for $h \in L^2(\mathbb{R}^2)$

$$\|h\|_2^2 = \sup_{\psi \in C_c(\mathbb{R}^2)} \left[2\int_{\mathbb{R}^2} h\psi\, dx - \|\psi\|_2^2\right],$$

we have for every $\psi \in C_c(\mathbb{R}^2)$

$$\liminf_{n\to\infty} \|J_n * \tilde{\rho}_n(r)\|_2^2 \geq \liminf_{n\to\infty} \left[2\int_{\mathbb{R}^2} J_n * \tilde{\rho}_n(r)\psi\, dx - \|\psi\|_2^2\right].$$

But

$$\lim_{n\to\infty} \int_{\mathbb{R}^2} J_n * \tilde{\rho}_n(r)\psi\, dx = \lim_{n\to\infty} \int_{\mathbb{R}^2} \tilde{\rho}_n(r) J_n * \psi\, dx = \lim_{n\to\infty} \int_{\mathbb{R}^2} \rho(r)\psi\, dx,$$

so we have

$$\|\rho(r)\|_2^2 = \sup_{\psi \in C_c(\mathbb{R}^2)} \left[2\int_{\mathbb{R}^2} \rho(r)\psi\, dx - \|\psi\|_2^2\right] \leq \liminf_{n\to\infty} \|J_n * \tilde{\rho}_n(r)\|_2^2.$$

Combining these facts with (2.17) and using Fatou's lemma, we thus obtain

$$\frac{\nu}{2}\mathbb{E}[\int_0^T \|\rho(r)\|_2^2 dr] \leq e(\rho_0) + e^T M_2(\rho_0) + C_{\nu,p,T}. \tag{2.20}$$

We will now pass to the limit in (2.7) for $\varphi \in C_c^\infty([0,T] \times \mathbb{R}^2)$. Following the proof of Lemma 7.1 we introduce a radial cutoff function $\phi \in C_c^\infty(\mathbb{R}^2)$, $0 \le \phi \le 1$, $\phi(x) = 1$ for $|x| \le 1$ and $\phi(x) = 0$ for $|x| \ge 2$ and define $\phi_r(x) = \phi(x/\tau)$ for $\tau > 0$. Then defining $K_{1,\tau} = -\phi_\tau K^\perp, K_{2,\tau} = -(1-\phi_\tau)K^\perp$,

$$u_{G_n * \tilde{\rho}_n} = -K^\perp * G_n * \tilde{\rho}_n = K_{1,\tau} * G_n * \tilde{\rho}_n + K_{2,\tau} * G_n * \tilde{\rho}_n.$$

We know that there is a set of full measure such that $\tilde{\rho}_n(\cdot)(\omega) \to \rho(\cdot)(\omega)$ in $C([0,T];\mathcal{P}(\mathbb{R}^2))$. For ω in this set we have the following. Since $\|K_{2,\tau} * G_n - K_{2,\tau}\|_\infty \to 0$ as $n \to +\infty$, $\|K_{2,\tau} * G_n * \tilde{\rho}_n(r)(\omega) \to K_{2,\tau} * \tilde{\rho}_n(r)(\omega)\|_\infty \to 0$ as $n \to +\infty$ for every $r \in [0,T]$. Also, since $K_{2,\tau} * \tilde{\rho}_n(r)(\omega)$ are equibounded and equicontinuous, we observe that $K_{2,\tau} * \tilde{(r)}\rho_n(\omega) \to K_{2,\tau} * \rho(r)(\omega)$ locally uniformly as $n \to +\infty$ for every $r \in [0,T]$.

Using these, the inequality $\|\nabla\varphi \cdot K_{2,\tau} * G_n * \rho_n\|_\infty \le C\|K_{2,\tau}\|_\infty$ and the Lebesgue dominated convergence theorem, it is then easy to see that for $0 < r < T$,

$$\int_s^t \langle \nabla\varphi \cdot K_{2,\tau} * G_n * \tilde{\rho}_n, \tilde{\rho}_n(r)\rangle dr \to \int_s^t \langle \nabla\varphi \cdot K_{2,\tau} * \rho, \rho(r)\rangle dr \quad a.s. \tag{2.21}$$

and

$$\mathbb{E}[\int_s^t \langle \nabla\varphi \cdot K_{2,\tau} * G_n * \tilde{\rho}_n, \tilde{\rho}_n(r)\rangle dr] \to \mathbb{E}[\int_s^t \langle \nabla\varphi \cdot K_{2,\tau} * \rho, \rho(r)\rangle dr]. \tag{2.22}$$

By Young's inequality we have $\|\nabla\varphi \cdot K_{1,\tau} * J_n * \tilde{\rho}_n\|_2 \le C\tau\|J_n * \tilde{\rho}_n\|_2$. Thus $|\langle \nabla\varphi \cdot K_{1,\tau} * G_n * \tilde{\rho}_n, \tilde{\rho}_n\rangle| = |\langle K_{1,\tau} * J_n * \tilde{\rho}_n, J_n * (\tilde{\rho}_n \nabla\varphi)\rangle| \le C\tau\|J_n * \tilde{\rho}_n\|_2^2$, which, using (2.17), yields

$$\mathbb{E}[\int_s^t |\langle \nabla\varphi \cdot K_{1,\tau} * G_n * \tilde{\rho}_n, \tilde{\rho}_n(r)\rangle|\, dr] \le C\tau\mathbb{E}[\int_s^t \|J_n * \tilde{\rho}_n(r)\|_2^2 dr] \le C_1\tau. \tag{2.23}$$

Since by the same argument we also have

$$\mathbb{E}[\int_s^t |\langle \nabla\varphi \cdot K_{1,\tau} * \rho, \rho(r)\rangle|\, dr] \le C\tau\mathbb{E}[\int_s^t \|\rho(r)\|_2^2 dr] \le C_1\tau, \tag{2.24}$$

sending $n \to \infty$ and then $\tau \to 0$ gives us

$$\mathbb{E}[\int_s^t \langle \nabla\varphi \cdot u_{G_n * \tilde{\rho}_n}, \tilde{\rho}_n(r)\rangle dr] \to \mathbb{E}[\int_s^t \langle \nabla\varphi \cdot u_\rho, \rho(r)\rangle dr]. \tag{2.25}$$

Moreover, (2.21), (2.23), (2.24) and standard arguments also imply that there is a subsequence n_k such that

$$\int_s^t \langle \nabla\varphi \cdot u_{G_{n_k} * \tilde{\rho}_{n_k}}, \tilde{\rho}_{n_k}(r)\rangle dr \to \int_s^t \langle \nabla\varphi \cdot u_\rho, \rho(r)\rangle dr \quad \text{as } k \to +\infty \ a.s.. \tag{2.26}$$

Similar convergencies of the other terms in (2.7) are easy consequences of the fact that $\tilde{\rho}_n(\cdot) \to \rho(\cdot)$ a.s. in $C([0,T];\mathcal{P}(\mathbb{R}^2))$ and the Lebesgue dominated convergence theorem. In the end we obtain

$$\mathbb{E}[\langle \varphi, \rho(t)\rangle - \langle \varphi, \rho(s)\rangle - \int_s^t [\langle \varphi_t + \nabla\varphi \cdot (u_\rho + v) + \nu\Delta\varphi, \rho(r)\rangle]\, dr] = 0 \tag{2.27}$$

and that for a subsequence n_k

$$\lim_{k\to+\infty}\left[\langle\varphi,\tilde\rho_{n_k}(t)\rangle-\langle\varphi,\tilde\rho_{n_k}(s)\rangle-\int_s^t\left[\langle\varphi_t+\nabla\varphi\cdot(u_{G_{n_k}*\tilde\rho_{n_k}}+v)+\nu\Delta\varphi,\tilde\rho_{n_k}(r)\rangle\right]dr\right]$$

$$=\langle\varphi,\rho(t)\rangle-\langle\varphi,\rho(s)\rangle-\int_s^t\left[\langle\varphi_t+\nabla\varphi\cdot(u_\rho+v)+\nu\Delta\varphi,\rho(r)\rangle\right]dr\quad a.s.. \tag{2.28}$$

By (2.6), (2.7), (2.27), (2.28) and Fatou's lemma we now obtain

$$\begin{aligned}
&\mathrm{Var}\left[\langle\varphi,\rho(t)\rangle-\langle\varphi,\rho(s)\rangle-\int_s^t\left[\langle\varphi_t+\nabla\varphi\cdot(u_\rho+v)+\nu\Delta\varphi,\rho(r)\rangle\right]dr\right]\\
&=\mathrm{Var}\Bigg[\lim_{k\to+\infty}\Big[\langle\varphi,\tilde\rho_{n_k}(t)\rangle-\langle\varphi,\tilde\rho_{n_k}(s)\rangle\\
&\qquad-\int_s^t\left[\langle\varphi_t+\nabla\varphi\cdot(u_{G_{n_k}*\tilde\rho_{n_k}}+v)+\nu\Delta\varphi,\tilde\rho_{n_k}(r)\rangle\right]dr\Big]\Bigg]\\
&\le\liminf_{k\to\infty}\mathrm{Var}\Big[\langle\varphi,\tilde\rho_{n_k}(t)\rangle-\langle\varphi,\tilde\rho_{n_k}(s)\rangle\\
&\qquad-\int_s^t\left[\langle\varphi_t+\nabla\varphi\cdot(u_{G_{n_k}*\tilde\rho_{n_k}}+v)+\nu\Delta\varphi,\tilde\rho_{n_k}(r)\rangle\right]dr\Big]\\
&=\liminf_{k\to\infty}\mathrm{Var}\Big[\langle\varphi,\rho_{n_k}(t)\rangle-\langle\varphi,\rho_{n_k}(s)\rangle\\
&\qquad-\int_s^t\left[\langle\varphi_t+\nabla\varphi\cdot(u_{G_{n_k}*\rho_{n_k}}+v)+\nu\Delta\varphi,\rho_{n_k}(r)\rangle\right]dr\Big]\\
&=\liminf_{k\to\infty}[M^\varphi_{n_k,s}](t)\le\liminf_{k\to\infty}2\nu T\|\nabla\varphi\|^2_\infty {n_k}^{-1}=0.
\end{aligned}\tag{2.29}$$

This, together with (2.27) shows that for every $0\le s<t\le T$ and every $\varphi\in C_c^\infty([0,T]\times\mathbb{R}^2)$,

$$\langle\varphi,\rho(t)\rangle-\langle\varphi,\rho(s)\rangle=\int_s^t\left[\langle\varphi_t+\nabla\varphi\cdot(u_\rho+v)+\nu\Delta\varphi,\rho(r)\rangle\right]dr,\quad a.s.. \tag{2.30}$$

It now remains to use that a.s. $\rho(\cdot)\in C([0,\infty);\mathcal{P}(\mathbb{R}^2))$ and invoke a separability and continuity argument, together with (2.20) and estimates of Lemma 7.5, to conclude that there is a set of full measure such that (2.30) is satisfied for every $0\le s<t\le T$ and every $\varphi\in C_c^\infty([0,T]\times\mathbb{R}^2)$, i.e. that a.s. $(\rho,-\nabla\cdot(\rho\nabla p))$ is a weak solution of (1.1) on $[0,T]$.

Finally, by (2.19) and (2.20), we can choose a trajectory $\rho(\cdot):=\rho(\cdot)(\omega)$ which satisfies (2.8) and and (2.9). The continuity of $M_2(\rho(t))$ at 0 follows from (2.8) and the lower semi-continuity of M_2 with respect to narrow convergence since $\rho(t)\to\rho_0$ narrowly as $t\to 0$. The fact that $\rho(\cdot)\in C([0,\infty);\mathcal{P}_q(\mathbb{R}^2))$ for each $q<2$ is standard. □

Once we know that the weak solution we constructed is in $L^2([0,T],L^2(\mathbb{R}^2))$ one can study its further regularity properties by classical and semigroup techniques (see for instance [13] and the references therein). We do not do it here since we are only interested in estimates and formulas for M_2, s and d^2.

Before we proceed to the next lemma we need to introduce another mollifier. Let $0<\gamma<1$ and $G\in C^\infty(\mathbb{R}^2)$ be a radial, positive function such that $G(x)=ce^{-|x|^\gamma}$ for $|x|>1$ where c

is chosen so that $\int G(x)dx = 1$. We set

$$G_\delta(x) = \delta^{-2} G(\frac{x}{\delta}).$$

We then mollify ρ into $\rho_\delta(t,x) = G_\delta * \rho(t,x) > 0$ for every $t \geq 0$.

Lemma 2.3. *There exists $C_\delta > 0$ depending only on δ and $\sup_{0\leq t\leq T} M_2(\rho(t))$ such that*

$$|\log \rho_\delta(t,x)| \leq C(1+|x|^\gamma). \tag{2.31}$$

Proof. We have by Jensen's inequality

$$\begin{aligned} C_\delta \geq \log \rho_\delta(t,x) &= \log\Big(\int_{\mathbb{R}^2} G_\delta(x-y)\rho(t,y)dy\Big) \\ &\geq \int_{\mathbb{R}^2} (\log G_\delta(x-y))\rho(t,y)dy \\ &\geq -\int_{\mathbb{R}^2} c_\delta(1+|x|^\gamma+|y|^\gamma)\rho(t,y)dy \geq -C_\delta(1+|x|^\gamma). \end{aligned}$$

□

Lemma 2.4. *Let $\rho(\cdot) \in C([0,T];\mathcal{P}(\mathbb{R}^2))$ be such that $(\rho, -\mathrm{div}(\rho v)), v = \nabla p$ solves (1.1-1.3) in the weak sense and let $s(\rho(0)) < +\infty$,*

$$\sup_{0\leq t\leq T} M_2(\rho(t)) + \int_0^T \|\rho\|_2^2 dt < \infty, \tag{2.32}$$

and

$$\lim_{t\to 0} M_2(\rho(t)) = M_2(\rho(0)). \tag{2.33}$$

Set

$$A_t(\rho(\cdot)) = \frac{1}{4\nu}\int_0^t \int_{\mathbb{R}^2} |v(s,x)|^2 \rho(t,x)dxds.$$

Then $\rho \in AC((0,T);\mathcal{P}_2(\mathbb{R}^2))$, and for $0 \leq s < t \leq T$ and some constant $C = C_\nu > 0$,

$$M_2(\rho(t)) \leq (M_2(\rho(0)) + 4\nu A_t(\rho) + 4\nu t)e^t, \tag{2.34}$$

$$s(\rho(t)) + \frac{\nu}{2}\int_0^t I(\rho(r))dr \leq s(\rho(0)) + CA_t(\rho(\cdot)), \tag{2.35}$$

$$M_2(\rho(t)) - M_2(\rho(s)) = \int_s^t \Big(4\nu + 2\int_{\mathbb{R}^2} x\cdot v\rho dx\Big)dr, \tag{2.36}$$

$$s(\rho(t)) - s(\rho(s)) = -\nu\int_s^t I(\rho(r))dr + \int_s^t \Big(\int_{\mathbb{R}^2} v\cdot \frac{\nabla\rho}{\rho} d\rho\Big)dr. \tag{2.37}$$

Proof. Let

$$u_\delta(t,x) = \frac{G_\delta * (\rho u)(t,x)}{\rho_\delta(t,x)}, \quad v_\delta(t,x) = \frac{G_\delta * (\rho v)(t,x)}{\rho_\delta(t,x)}.$$

Then, using the definition of weak solution, it is easy to see that for every $\varphi \in C_c^\infty(\mathbb{R}^2)$ and for $0 < t \le T$,

$$\begin{aligned}\int_{\mathbb{R}^2} \rho_\delta(t,x)\varphi(x)dx - \int_{\mathbb{R}^2} \rho_\delta(0,x)\varphi(x)dx \\ = \int_{\mathbb{R}^2}\int_0^t [\nu\Delta\rho_\delta - \operatorname{div}(\rho_\delta(u_\delta+v_\delta))](r,x)\varphi(x)drdx.\end{aligned} \tag{2.38}$$

Since $\Delta\rho_\delta = (\Delta G_\delta) * \rho, \operatorname{div}(\rho_\delta u_\delta) = (\nabla G_\delta) * (\rho u), \operatorname{div}(\rho_\delta v_\delta) = (\nabla G_\delta) * (\rho v)$, using (7.27), we get that for every $s \in (0,T), x, z \in \mathbb{R}^2$

$$\int_{\mathbb{R}^2} |\Delta\rho_\delta(s,y)|dy \le C_\delta, \quad |\Delta\rho_\delta(s,x) - \Delta\rho_\delta(s,z)| \le C_\delta|x-z|,$$

$$\int_{\mathbb{R}^2} |\operatorname{div}(\rho_\delta u_\delta)(s,y)|dy \le C_\delta \int_{\mathbb{R}^2} |u(s,y)|\rho(s,y)dy \le C_\delta\|\rho_2(s)\|_2,$$

$$|\operatorname{div}(\rho_\delta u_\delta)(s,x) - \operatorname{div}(\rho_\delta u_\delta)(s,z)| \le C_\delta\|\rho_2(s)\|_2|x-z|,$$

$$\int_{\mathbb{R}^2} |\operatorname{div}(\rho_\delta v_\delta)(s,y)|dy \le C_\delta \int_{\mathbb{R}^2} |v(s,y)|^2\rho(s,y)dy \le C_\delta,$$

$$|\operatorname{div}(\rho_\delta v_\delta)(s,x) - \operatorname{div}(\rho_\delta v_\delta)(s,z)| \le C_\delta \int_{\mathbb{R}^2} |v(s,y)|^2\rho(s,y)dy|x-z| \le C_\delta|x-z|.$$

Therefore defining $h_\delta(s,x) := [\nu\Delta\rho_\delta - \operatorname{div}(\rho_\delta(u_\delta+v_\delta))](s,x)$, we see that for every $t \in (0,T), x, z \in \mathbb{R}^2$

$$\left|\int_0^t h_\delta(s,x)ds - \int_0^t h_\delta(s,z)ds\right| \le \tilde{C}_\delta|x-z|$$

and for every $z \in \mathbb{R}^2$, $h_\delta(s,z) \in L^1(0,T)$. Thus (2.38) implies that for every $0 < t \le T, z \in \mathbb{R}^2$

$$\rho_\delta(t,z) - \rho_\delta(0,z) = \int_0^t h_\delta(s,z)ds,$$

i.e. the function $\rho_\delta(\cdot,z)$ is absolutely continuous on $[0,T]$ for every $z \in \mathbb{R}^2$.

Therefore, setting $F(r) = r\log r$, we can write

$$\begin{aligned} s(\rho_\delta(t)) - s(\rho_\delta(0)) &= \int_{\mathbb{R}^2}(F(\rho_\delta(t,x)) - F(\rho_\delta(0,x)))dx \\ &= \int_{\mathbb{R}^2}\int_0^t F'(\rho_\delta(r,x))\Big(\nu\Delta\rho_\delta(r,x) - \operatorname{div}(\rho_\delta(r,x)(u_\delta(r,x)+v_\delta(r,x)))\Big)drdx.\end{aligned} \tag{2.39}$$

We notice that because of (2.32) the integral in (2.39) is finite. We will demonstrate it only for the term $\int\int F'(\rho_\delta(r,x))\operatorname{div}(\rho_\delta(r,x)u_\delta(r,x))drdx$ as the other terms are easy. First we notice that by Lemma 2.3,

$$|F'(\rho_\delta(r,x))| = |\log\rho_\delta(t,x) + 1| \le C(1+|x|^\gamma) \le C(1+|x|) \le C(1+|x-y|+|y|)$$

for every $r \in (0,T), x, y \in \mathbb{R}^2$. Thus, by (2.32) and (7.2),

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^2} |F'(\rho_\delta(r,x))\mathrm{div}(\rho_\delta(r,x)u_\delta(r,x))|drdx \\
&\leq \int_0^T \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} C(1+|x-y|+|y|)|\nabla G_\delta(x-y)|\rho(r,y)|u(r,y))|dydrdx \\
&\leq \int_0^T \int_{\mathbb{R}^2} C_\delta(1+|y|)\rho(r,y)|u(r,y))|dydr \\
&\leq C_\delta \int_0^T \left[\int_{\mathbb{R}^2}(1+|y|)^2\rho(r,y)dy\right]^{\frac{1}{2}} \|\rho(r)\|_2^{\frac{1}{2}}\|u(r)\|_4 ddr \leq C_\delta \int_0^T \|\rho(r)\|_2 < +\infty.
\end{aligned}$$

We want to integrate by parts in (2.39), i.e. we want to write

$$\begin{aligned}
&s(\rho_\delta(t)) - s(\rho_\delta(0)) \\
&= \int_0^t \lim_{R\to\infty}\int_{B_R} F''(\rho_\delta(r,x))\Big(-\nu|\nabla\rho_\delta(r,x)|^2 + \rho_\delta u_\delta\cdot\nabla\rho_\delta + \rho_\delta v_\delta\cdot\nabla\rho_\delta\Big)dxdr \\
&= \int_0^t \lim_{R\to\infty}\int_{B_R} \Big(-\nu\frac{|\nabla\rho_\delta|^2}{\rho_\delta} + (u_\delta+v_\delta)\cdot\nabla\rho_\delta\Big)dxdr \qquad (2.40)\\
&\leq \int_0^t \lim_{R\to\infty}\left(-\nu\left\|\frac{|\nabla\rho_\delta|}{\sqrt{\rho_\delta}}\right\|^2_{L^2(B_R)} + \sqrt{2}\left\|\frac{|\nabla\rho_\delta|}{\sqrt{\rho_\delta}}\right\|_{L^2(B_R)}\left(\int_{\mathbb{R}^2}(|u_\delta|^2+|v_\delta|^2)d\rho_\delta\right)^{\frac{1}{2}}\right)dr \\
&\leq -\frac{\nu}{2}\int_0^t I(\rho_\delta)dr + C_\nu\int_0^t\int_{\mathbb{R}^2}(|u_\delta|^2+|v_\delta|^2)d\rho_\delta dr.
\end{aligned}$$

We notice that the last line of (2.40) is well defined as by Lemma 8.1.9 of [3], together with (2.32) and Lemma 7.5, we have

$$\lim_{\delta\to 0}\int_0^t\int_{\mathbb{R}^2}(|u_\delta|^2+|v_\delta|^2)d\rho_\delta dr = \int_0^t\int_{\mathbb{R}^2}(|u|^2+|v|^2)d\rho dr \leq \int_0^t\left(C\|\rho\|_2^2 + \int_{\mathbb{R}^2}|v|^2d\rho\right)dr. \qquad (2.41)$$

Assuming for a moment that integration by parts was allowed in (2.40), we pass to the limit there as $\delta \to 0$. It is easy to see that $\sqrt{\rho_\delta(r)} \to \sqrt{\rho(r)}$ in $L^2(\mathbb{R}^2)$ as $\delta \to 0$. Thus, using the weak sequential lower semi-continuity of the norm and Fatou's lemma we obtain

$$\begin{aligned}
&\limsup_{\delta\to 0} -\int_0^t I(\rho_\delta(r))dr = -\liminf_{\delta\to 0}\int_0^t \||\nabla\sqrt{\rho_\delta(r)}|\|_2^2 dr \\
&\leq -\int_0^t \liminf_{\delta\to 0}\||\nabla\sqrt{\rho_\delta(r)}|\|_2^2dr \leq -\int_0^t\||\nabla\sqrt{\rho(r)}|\|_2^2dr = -\int_0^t I(\rho(r))dr. \qquad (2.42)
\end{aligned}$$

Now, since $F(\rho(0)) \in L^1(\mathbb{R}^2)$, using Jensen's inequality

$$s(\rho_\delta(0)) = \int_{\mathbb{R}^2} F(G_\delta * \rho(0))dx \leq \int_{\mathbb{R}^2} G_\delta * F(\rho(0))dx,$$

which yields

$$\liminf_{\delta\to 0} -s(\rho_\delta(0)) \geq -\lim_{\delta\to 0}\int_{\mathbb{R}^2} G_\delta * F(\rho(0))dx = -\int_{\mathbb{R}^2}F(\rho(0))dx = -s(\rho(0)). \qquad (2.43)$$

Finally, using Fatou's lemma we have

$$s(\rho(r)) \le \liminf_{\delta\to 0} s(\rho_\delta(r)) \tag{2.44}$$

Therefore letting $\delta \to 0$ in (2.40) and applying (2.41)-(2.44) we obtain

$$s(\rho(t)) + \frac{\nu}{2}\int_s^t I(\rho(r))dr \le s(\rho(0)) + C_{1,\nu}\Big(A_t(\rho(\cdot)) + \int_0^t \|\rho\|_2^2 dxdr\Big). \tag{2.45}$$

Let us now justify that we can integrate by parts in (2.40). We can pass from the first to the second line in (2.40), if we can show that for every s the boundary integral

$$\int_{\partial B_R} (\log(\rho_\delta(s)) + 1)(\nabla\rho_\delta(s) - G_\delta * (\rho u)(s) - G_\delta * (\rho v)(s)) \cdot \mathbf{n} d\sigma(x) \tag{2.46}$$

goes to 0 as $R \to \infty$, where $\mathbf{n}$ is the exterior unit normal vector on ∂B_R. Since $|\partial B_R| = 2\pi R$ and $|\log(\rho_\delta(s)) + 1| \le C(1 + R^\gamma)$ on ∂B_R, this will be achieved if we can show that

$$|\nabla\rho_\delta(s,x)| + |G_\delta * (\rho u)(s,x)| + |G_\delta * (\rho v)(s,x)| \le \frac{C}{|x|^{1+\gamma+\epsilon}} \tag{2.47}$$

for some $\epsilon > 0$ and all $x \in \mathbb{R}^2$, where C is some constant which may depend on s. Let $\tau = 1 + \gamma + \epsilon < 2$. We will use the fact that for every $x, y \in \mathbb{R}^2$

$$|x|^\tau \le C(|x-y|^\tau + |y|^\tau).$$

The C_δs below are generic constants and may differ from line to line.

(1) We have, since $\|v\|_\infty < \infty$,

$$\begin{aligned}
&|x|^\tau(|\nabla\rho_\delta(s,x)| + |G_\delta * (\rho v)(s,x)|)\\
&\le \int_{\mathbb{R}^2} C(|x-y|^\tau + |y|^\tau)(|\nabla G_\delta(x-y)| + G_\delta(x-y)|v(s,y)|)\rho(s,y)dy\\
&\le C_\delta(s) + C_\delta(s)\int_{\mathbb{R}^2} |y|^\tau\rho(s,y)dy \le C_\delta(s)(1 + M_2(\rho(s))) \le C_\delta(s).
\end{aligned}$$

(2) Using generalized Hölder inequality and (7.2)

$$\begin{aligned}
&|x|^\tau|G_\delta * (\rho u)(s,x)| \le \int_{\mathbb{R}^2} C(|x-y|^\tau + |y|^\tau)|G_\delta(x-y)||u(s,y)|\rho(s,y)dy\\
&\le \int_{\mathbb{R}^2} C_\delta(1 + |y|^\tau)|u(s,y)|\rho(s,y)dy \le C_\delta\|\rho(s)\|_2\\
&+C_\delta\left(\int_{\mathbb{R}^2}|y|^2\rho(s,y)dy\right)^{\frac{\tau}{2}}\left(\int_{\mathbb{R}^2}|\rho(s,y)|^{\frac{3}{2}}dy\right)^{\frac{2-\tau}{3}}\|u\|_{\frac{6}{2-\tau}}\\
&\le C_\delta\|\rho(s)\|_2 + C_\delta(1 + \|\rho(s)\|_2^{\frac{2(2-\tau)}{3}}\|\rho(s)\|_2^{1-\frac{2-\tau}{3}}) \le C_\delta(1 + \|\rho(s)\|_2^{1+\frac{2-\tau}{3}}) \le C_\delta(1 + \|\rho(s)\|_2^2).
\end{aligned}$$

This completes the proof of (2.47).

With the above estimate, we can now improve (2.45) by getting rid of the $\int_0^t \|\rho\|_2^2 dr$ term. Since

$$\begin{aligned}
\|\dot\rho\|_{-1,\rho} &\le \|\mathrm{div}(\rho v)\|_{-1,\rho} + \|\nu\Delta\rho\|_{-1,\rho} + \|\mathrm{div}(\rho u)\|_{-1,\rho}\\
&\le (\int_{\mathbb{R}^2}|v|^2 d\rho)^{1/2} + \nu(I(\rho))^{1/2} + (\int_{\mathbb{R}^2}|u|^2 d\rho)^{1/2},
\end{aligned}$$

by Lemma 7.5 we have

$$\int_0^T \|\dot{\rho}\|_{-1,\rho} dr < \infty.$$

By Theorem 8.3.1 of [3], $\rho \in AC((0,T);\mathcal{P}_2(\mathbb{R}^2))$. We can now apply the chain rule of Lemma 7.17 and the convexity results regarding M_2 and s of Lemma 7.18 in Appendix to obtain the following.

Since $I(\rho(r)) < \infty$ for a.e. $r \in (0,T)$, by Lemma 7.6, $\int_{\mathbb{R}^2} u \cdot \nabla \rho dx = 0$ and by Lemma 7.7, $\int_{\mathbb{R}^2} x \cdot u d\rho = 0$. Also, since $M_2(\rho(t))$ is continuous at 0, (2.45) implies that $s(\rho(0))$ is continuous at 0. Therefore we have for $0 \leq s < t \leq T$

$$\begin{aligned}
&M_2(\rho(t)) \\
&= M_2(\rho(s)) + \int_s^t \int_{\mathbb{R}^2} \Big(4\nu + 2x \cdot (u+v)\rho\Big) dx dr = M_2(\rho(s)) + \int_s^t \int_{\mathbb{R}^2} \Big(4\nu + 2x \cdot v\rho\Big) dx dr \\
&\leq M_2(\rho(s)) + 4\nu(t-s) + 2\int_s^t \Big(\sqrt{M_2(\rho(r))}\sqrt{\int_{\mathbb{R}^2} |v|^2 d\rho(r)}\Big) dr \\
&\leq M_2(\rho(s)) + 4\nu(t-s) + \int_s^t \Big(\int_{\mathbb{R}^2} |v|^2 d\rho(r) + M_2(\rho(r))\Big) dr.
\end{aligned}$$

The required estimate now follows from Gronwall's inequality.

On the other hand,

$$\begin{aligned}
s(\rho(t)) &= s(\rho(s)) + \int_s^t \int_{\mathbb{R}^2} \Big(-\nu \frac{|\nabla \rho|^2}{\rho} + (v+u) \cdot \nabla \rho \Big) dx dr \\
&= -\nu \int_s^t I(\rho(r)) dr + \int_s^t \Big(\int_{\mathbb{R}^2} v \cdot \frac{\nabla \rho}{\rho} d\rho \Big) dr \\
&\leq s(\rho(s)) + \int_s^t \Big(-\nu I(\rho(r)) + \sqrt{I(\rho(r))}\sqrt{\int_{\mathbb{R}^2} |v|^2 d\rho(r)} \Big) dr \\
&\leq s(\rho(s)) - \frac{\nu}{2} \int_s^t I(\rho(r)) dr + C_\nu \int_s^t \int_{\mathbb{R}^2} |v|^2 d\rho(r) dr.
\end{aligned}$$

□

Lemma 2.5. *Let $\rho(\cdot), v$ be as in Lemma 2.4. Then*

$$-\frac{1}{2} d^2(\rho(t),\sigma) + \frac{1}{2} d^2(\rho(s),\sigma) = \int_s^t \int_{\mathbb{R}^2} \nabla p_{\rho(r),\sigma}(x) \cdot (\nu \frac{\nabla \rho}{\rho} + u + v) \rho(r,dx) dr \tag{2.48}$$

for all $0 \leq s < t \leq T$ and $\sigma \in \mathcal{P}_2(\mathbb{R}^2)$ with $s(\sigma) < \infty$. In the above, $p_{\rho(r),\sigma}$ is the difference of two convex functions defining Brenier's optimal transport map in Theorem D.25 (Appendix D) [11].

Proof. The result follows from Lemmas 7.17 and 7.18 in Appendix. □

We can now state the following existence result.

Corollary 2.6. *Let $\rho_0 \in \mathcal{P}_2(\mathbb{R}^2)$, be such that $s(\rho_0) < +\infty$. Let $p : [0,\infty) \times \mathbb{R}^2 \to \mathbb{R}$ be such that there exists a sequence $0 = t_0 < t_1 < ... < t_n < ..., t_n \to +\infty$ such that $p \in C_c^\infty([t_n, t_{n+1}] \times \mathbb{R}^2)$ for $n = 0, 1, ...$ Then there exists $\rho(\cdot) \in C([0,+\infty);\mathcal{P}_2(\mathbb{R}^2))$ such* *that* $(\rho, -\nabla\cdot(\rho\nabla p))$ *is a weak solution of (1.1) on* $[0,\infty)$ *and* $\rho(0) = \rho_0$*. Moreover, if* $\rho(\cdot) \in \mathcal{K}_{\rho_0}$ *then* $\rho(\cdot)$ *satisfies (2.34)-(2.37) and (2.48) for every* $0 \le s < t < \infty$*, and*

$$d(\rho(t), \rho_0) \le C_1 \left(C_2 + s(\rho_0) + M_2(\rho_0) + \int_0^t \int_{\mathbb{R}^2} |\nabla p|^2 d\rho dr \right)^{\frac{1}{2}} \sqrt{t} \tag{2.49}$$

for $0 < t \le 1$*.*

Proof. The result follows from Lemmas 2.2, 2.4, and 2.5 and the definition of weak solution which allows us to patch together solutions obtained on intervals $[0,t_1], [t_1,t_2], \ldots$. Estimate (2.49) is the consequence of estimates (2.34), (2.35), (7.26), and the characterization of the 2-Wasserstein distance, see for instance [3], (8.0.3), page 168. ☐

3. The controlled PDE as a mixture of controlled gradient and Hamiltonian flows

It is useful to adapt a geometric point of view of interpreting (1.1). The goal is to discover a gradient and Hamiltonian flow structure of (1.1) and rewrite it as an abstract evolution equatin (3.23) in the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^2), d)$. To achieve this, we need to invoke some results in optimal mass transport theory [3], [29].

3.1. A differential calculus on $\mathcal{P}_2(\mathbb{R}^2)$.

Following Otto [26], we formally view $(\mathcal{P}_2(\mathbb{R}^2), d)$ as an infinite dimensional Riemannian manifold with tangent space $T_\rho\mathcal{P}_2(\mathbb{R}^2)$ at ρ modeled using $H_{-1,\rho}(\mathbb{R}^2)$. Let function $f : \mathcal{P}_2(\mathbb{R}^2) \mapsto \mathbb{R}\cup\{\pm\infty\}$ and $\rho_0 \in E$. This allows to define the notion of gradient.

Definition 3.1 (Gradient)**.** For each $p \in C_c^\infty(\mathbb{R}^2)$, let $\rho^p = \rho^p(t,x), t \ge 0$ be defined according to

$$\partial_t \rho^p(t) + \mathrm{div}(\rho^p \nabla p) = 0, \quad \rho^p(0) = \rho_0.$$

The gradient of f evaluated at ρ_0, written as $\mathrm{grad} f(\rho_0)$, is said to exist, if and only if it can be identified as the unique element in the Schwartz space $\mathcal{D}'(\mathbb{R}^2)$ satisfying

$$\lim_{t\to 0+} \frac{f(\rho^p(t)) - f(\rho_0)}{t} = \langle \mathrm{grad} f(\rho_0), p\rangle, \quad \forall p \in C_c^\infty(\mathbb{R}^2).$$

Following Gangbo, Kim and Pacini [12], Ambrosio and Gangbo [2], we introduce the notion of symplectic gradient.

Definition 3.2 (Symplectic Gradient)**.** For each $p \in C_c^\infty(\mathbb{R}^2)$, let $\rho^p = \rho^p(t,x), t \ge 0$ be defined according to

$$\partial \rho^p(t) + \nabla\cdot(\rho^p(J\nabla p)) = 0, \quad \rho^p(0) = \rho_0.$$

The symplectic gradient of f evaluated at ρ_0, written as J-grad $f(\rho_0)$, is said to exist, if and only if it can be identified as the unique element in the Schwartz space $\mathcal{D}'(\mathbb{R}^2)$ satisfying

$$\lim_{t\to 0+} \frac{f(\rho^p(t)) - f(\rho_0)}{t} = \langle J\text{-}\mathrm{grad} f(\rho_0), p\rangle, \quad \forall p \in C_c^\infty(\mathbb{R}^2).$$

Example 3.3. Let $\varphi_i \in C_c^\infty(\mathbb{R}^2)$, $k = 1, 2, \ldots$. We denote

$$\vec{\varphi} = (\varphi_1, \ldots, \varphi_k), \quad \vec{\psi} = (\psi_1, \ldots, \psi_k).$$

For smooth test functions of the form

$$f(\rho) = h(\langle \rho, \varphi_1\rangle, \ldots, \langle \rho, \varphi_k\rangle) = h(\langle \rho, \vec{\varphi}\rangle), \tag{3.1}$$

where $h \in C^1(\mathbb{R}^k)$, we define the un-constrained first order variational derivative

$$\frac{\delta f}{\delta \rho} = \sum_{i=1}^{k} \partial_i h(\langle \rho, \vec{\varphi} \rangle)\varphi_i.$$

Then

$$\text{grad}_\rho f = -\text{div}(\rho \nabla \frac{\delta f}{\delta \rho}) \tag{3.2}$$

Next, we consider functionals $e, s, d^2(\cdot, \gamma), M_2$ as defined in (1.12), (1.15), (1.10), (1.13). The functionals $M_2, d^2(\cdot, \gamma)$ are continuous and e, s are lower semicontinuous on the metric space $(\mathcal{P}_2(\mathbb{R}^2), d)$. In addition, by Theorem D.28 on page 381 of Feng and Kurtz [11] (also Lemma 10.4.4 of [3]),

$$\text{grad}_\rho\, s(\rho) = -\Delta \rho, \quad \text{if } s(\rho) < \infty, \tag{3.3}$$

$$\text{grad}_\rho\, d^2(\rho, \gamma) = -2\text{div}(\rho \nabla p_{\rho,\gamma}), \quad \text{if } \rho, \gamma \text{ have Lebesgue densities.} \tag{3.4}$$

In the above,

$$p_{\rho,\gamma}(x) := \frac{1}{2}|x|^2 - \varphi_{\rho,\gamma}(x),$$

where $\varphi_{\rho,\gamma}$ is the convex function defining Brenier's optimal transport map as in Kantorovich duality formulation of the 2-Wasserstein metric (e.g. Theorem D.20 and D.25 in [11]). Let $\mathcal{O}$ be the interior of the convex hull of the support of ρ, $\nabla p_{\rho,\gamma}$ is only defined in $\mathcal{O}$. But $\rho(\mathcal{O}) = 1$ and

$$\int_{\mathbb{R}^2} |\nabla p_{\rho,\gamma}(x)|^2 \rho(dx) = \int_{\mathbb{R}^2 \times \mathbb{R}^2} |x-y|^2 \pi_{\rho,\gamma}(dx, dy) = d^2(\rho, \gamma) < \infty. \tag{3.5}$$

So , taking $\rho \nabla p_{\rho,\gamma}$ to be zero off the support of ρ, $\rho \nabla p_{\rho,\gamma} \in L^1(\mathbb{R}^2)$ and $\nabla \cdot (\rho \nabla p_{\rho,\gamma}) \in \mathcal{D}'(\mathbb{R}^2)$. Thus (3.5) means that

$$\|\text{grad}_\rho d^2(\rho, \gamma)\|_{-1,\rho}^2 = 4d^2(\rho, \gamma), \quad \text{if } \rho, \gamma \text{ have Lebesgue densities.} \tag{3.6}$$

Direct calculation reveals that

$$\text{grad}_\rho\, M_2(\rho) = -\text{div}(2x\rho). \tag{3.7}$$

We also refer to a broader class of examples given in Theorem 10.4.13 in [3]. Regarding e we have the following result.

Lemma 3.4. *Let $\rho_0 \in \mathcal{P}_2(\mathbb{R}^2)$ be such that $\|\rho_0\|_2 < \infty$. Then*

$$\text{grad}_{\rho_0}\, e(\rho_0) = -\text{div}(\rho_0 \nabla U_{\rho_0}), \tag{3.8}$$

$$J\text{-grad}_{\rho_0}\, e(\rho_0) = -\text{div}(\rho_0 u_{\rho_0}), \tag{3.9}$$

where $U_{\rho_0} = N * \rho_0$.

Proof. Let $\xi = \nabla p : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be a vector field such that $p \in C_c^\infty(\mathbb{R}^2)$. We define a corresponding flux $\{\Phi(t) : t \in \mathbb{R}\}$ by

$$\partial_t \Phi(t, \cdot) = \xi \circ \Phi(t, \cdot), t \in \mathbb{R}, \quad \Phi(0, x) = x.$$

Let

$$\partial_t \rho + \text{div}(\rho \xi) = 0, \quad \rho(0) = \rho_0.$$

By [3], chapter 8, we know that $\rho \in AC((0,\infty);\mathcal{P}_2(\mathbb{R}^2))$ and $\rho(t) = \Phi(t)_{\#}\rho_0$ in the sense of a change of variable formula that for every non-negative Borel measurable function φ,

$$\int_{\mathbb{R}^2} \varphi(y)\rho(t,dy) = \int_{\mathbb{R}^2} \varphi(\Phi(t,x))\rho_0(dx). \tag{3.10}$$

We notice that since $\Phi(t)$ is smooth and $\Phi(0,x) = x$, hence for $\tau > 0$ small enough and $0 \le t < \tau$

$$\det D\Phi(t,x) \ge \epsilon_0 \text{ for some } \epsilon_0 > 0.$$

Therefore, since $\Phi(t)$ is a diffeomorphism, (3.10) implies that for $0 \le t < \tau$

$$\rho(t,\Phi(t,x)) = \rho_0(x)[\det D\Phi(t,x)]^{-1}.$$

Therefore

$$\frac{1}{\epsilon_0}\|\rho_0\|_2^2 \ge \int_{\mathbb{R}^2} \frac{|\rho_0(x)|^2}{\det D\Phi(t,x)}dx = \int_{\mathbb{R}^2} |\rho(t,\Phi(t,x))|^2 \det D\Phi(t,x)dx = \int_{\mathbb{R}^2} |\rho(t,y)|^2 dy, \tag{3.11}$$

which implies that

$$\sup_{0\le t<\tau} \|\rho(t)\|_2 < \infty.$$

It is also easy to see that

$$\sup_{0\le t<\tau} M_2(\rho(t)) < \infty.$$

Recall that $e_n(\rho) = 1/2\langle N_n * \rho, \rho\rangle$, where $N_n = N * G_n = N * J_n * J_n$ and J_n is a standard mollifier defined in Section 2. We notice that

$$e_n(\rho(t)) = \frac{1}{2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} N_n(\Phi(t,x) - \Phi(t,y))\rho_0(y)\rho_0(x)dydx$$

and thus, by the dominated convergence theorem we easily obtain that $\lim_{t\to 0} e_n(\rho(t)) = e_n(\rho_0)$.

By Lemma 7.18 in Appendix, e_n is λ-convex. By the chain rule in Lemma 7.17,

$$e_n(\rho(t)) = e_n(\rho_0) + \int_0^t \int_{\mathbb{R}^2} (\nabla N_n * \rho(r)) \cdot \xi d\rho(r)dr. \tag{3.12}$$

By Lemma 7.4, $\lim_{n\to\infty} e_n(\rho(t)) = e(\rho(t))$ and $\lim_{n\to\infty} e_n(\rho_0) = e(\rho_0)$. Also, since $\nabla N_n = \nabla N * G_n$, using Hölder inequality and (7.4), we obtain

$$\begin{aligned}
&\left|\int_0^t \int_{\mathbb{R}^2} (\nabla N_n * \rho(r)) \cdot \xi d\rho(r)dr - \int_0^t \int_{\mathbb{R}^2} (\nabla N * \rho(r)) \cdot \xi d\rho(r)dr\right| \\
&\le \int_0^t \|\rho(t)\|_2 \|(\nabla N * (G_n * \rho(t) - \rho(t))) \cdot \xi\|_2 dr \\
&\le C\int_0^t \|G_n * \rho(t) - \rho(t)\|_2^{1/2} dr \to 0 \text{ as } n \to \infty.
\end{aligned}$$

Thus, letting $n \to \infty$ in (3.12) we arrive at

$$e(\rho(t)) = e(\rho_0) + \int_0^t \int_{\mathbb{R}^2} (\nabla N * \rho(r))\xi d\rho(r)dr. \tag{3.13}$$

We now claim that $\int_{\mathbb{R}^2}(\nabla N * \rho(r))\xi d\rho(r)$ is continuous in $r \geq 0$. To see this, we first note that by the change of variables,

$$\int_{\mathbb{R}^2}(\nabla N * \rho(r))\xi d\rho(r) = \frac{1}{2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\nabla N(\Phi(r,x)-\Phi(r,y))(\xi(\Phi(r,x))-\xi(\Phi(r,y)))\rho_0(x)\rho_0(y)dxdy.$$

Secondly, by the Lipschitz continuity of ξ,

$$|(\xi(x)-\xi(y))\nabla N(x-y)| \leq C, \quad \text{for some } C>0.$$

Hence the continuity of $\int_{\mathbb{R}^2}(K * \rho(r))\xi d\rho(r)$ in r follows by the dominated convergence theorem.

Consequently

$$\lim_{t\to 0+} t^{-1}(e(\rho(t)) - e(\rho(0))) = \int_{\mathbb{R}^2}\xi\rho_0(x)(\nabla N * \rho_0)dx$$

which shows (3.8). The proof of (3.9) is virtually the same.

□

We remind that condition $\|\rho_0\|_2 < \infty$ in the above lemma is weaker than $I(\rho_0) < \infty$ since $\|\rho_0\|_2^2 \leq C(1 + I(\rho_0))$.

3.2. **Several useful identities and estimates.** We now give several estimates regarding $\mathrm{grad}_\rho s$, $\mathrm{grad}_\rho M_2$, $\mathrm{grad}_\rho e$, and J-$\mathrm{grad}_\rho e$. For $\rho \in \mathcal{P}_2(\mathbb{R}^2)$ we denote by P_ρ to be the projection operator of $L^2(\rho)$ functions on to subspace

$$L^2_\nabla(\rho) \equiv \overline{\{\nabla\varphi : \varphi \in C_c^\infty(\mathbb{R}^2)\}}^{L^2(\rho)}.$$

Then

$$(L^2_\nabla(\rho))^\perp = \{v \in L^2(\rho) : \mathrm{div}(\rho v) = 0\}.$$

Moreover,

$$\langle -\mathrm{div}(\rho u), -\mathrm{div}(\rho v)\rangle_{-1,\rho} = \int_{\mathbb{R}^2}(P_\rho u)(P_\rho v)d\rho = \int_{\mathbb{R}^2} vP_\rho u d\rho = \int_{\mathbb{R}^2} uP_\rho v d\rho.$$

See Section 8.4 of [3].

Lemma 3.5. *Assume that $\rho \in \mathcal{P}_2(\mathbb{R}^2)$ satisfies $I(\rho) < \infty$ and $\gamma^1, \gamma^2, \ldots$ are as in Definition 4.1. Then:*

$$\|\mathrm{grad}_\rho s\|_{-1,\rho}^2 = I(\rho), \tag{3.14}$$

$$\|J\text{-}\mathrm{grad}_\rho e\|_{-1,\rho}^2 = \int_{\mathbb{R}^2}|P_\rho u_\rho|^2\rho(dx) \leq C(1+I(\rho)),$$

$$\|\mathrm{grad}_\rho M_2(\rho)\|_{-1,\rho}^2 = 4M_2(\rho) \tag{3.15}$$

$$\begin{aligned}\|\mathrm{grad}_\rho \sum_{k=1}^\infty \beta_k d^2(\rho,\gamma^k)\|_{-1,\rho}^2 &= 4\|\sum_{k=1}^\infty \beta_k \nabla\cdot\left(\rho\nabla p_{\rho,\gamma^k}\right)\|_{-1,\rho}^2 \\ &\leq 4\left(\sum_{k=1}^\infty \beta_k\right)\left(\sum_{k=1}^\infty \beta_k\|\nabla\cdot\left(\rho\nabla p_{\rho,\gamma^k}\right)\|_{-1,\rho}^2\right) \\ &= 4\left(\sum_{k=1}^\infty \beta_k\right)\sum_{k=1}^\infty \beta_k d^2(\rho,\gamma^k)\end{aligned} \tag{3.16}$$

and

$$\langle \mathrm{grad}_\rho s(\rho), J\text{-}\mathrm{grad}_\rho e(\rho)\rangle_{-1,\rho} = -\int_{\mathbb{R}^2} \nabla\rho \cdot K^\perp * \rho\, dx = 0 \tag{3.17}$$

$$\langle \mathrm{grad}_\rho s(\rho), \mathrm{grad}_\rho \sum_{k=1}^\infty \beta_k d^2(\rho,\gamma^k)\rangle_{-1,\rho} = 2\sum_{k=1}^\infty \beta_k \int_{\mathbb{R}^2} \nabla\rho \cdot \nabla p_{\rho,\gamma^k} dx \tag{3.18}$$

$$\langle \mathrm{grad}_\rho s(\rho), \mathrm{grad}_\rho M_2(\rho)\rangle_{-1,\rho} = 2\int_{\mathbb{R}^2} \nabla\rho \cdot x\, dx = -4 \tag{3.19}$$

$$\langle J\text{-}\mathrm{grad}_\rho e(\rho), \mathrm{grad}_\rho M_2(\rho)\rangle_{-1,\rho} = -2\int_{\mathbb{R}^2} K^\perp * \rho \cdot x\rho\, dx = 0 \tag{3.20}$$

$$\langle J\text{-}\mathrm{grad}_\rho e(\rho), \mathrm{grad}_\rho \sum_{k=1}^\infty \beta_k d^2(\rho,\gamma^k)\rangle_{-1,\rho} = -2\sum_{k=1}^\infty \beta_k \int_{\mathbb{R}^2} K^\perp * \rho \cdot \nabla p_{\rho,\gamma^k}\rho\, dx \tag{3.21}$$

$$\langle \mathrm{grad}_\rho M_2(\rho), \mathrm{grad}_\rho \sum_{k=1}^\infty \beta_k d^2(\rho,\gamma^k)\rangle_{-1,\rho} = 2\sum_{k=1}^\infty \beta_k \int_{\mathbb{R}^2} x \cdot \nabla p_{\rho,\gamma^k}\rho\, dx \tag{3.22}$$

Proof. (3.14) follows from Theorem D.45 of [11]. (3.15) follows from Lemmas 3.4, 7.5. (3.15), (3.16) follow from (3.7) and (3.4). (3.17) follows from (3.3) and Lemmas 3.4, 7.6. (3.19) follows from (3.3), (3.7) and integration by parts, and (3.20) follows from Lemma 7.7. The remaining formulas are also straightforward using (3.4).

□

We will see in Section 3.3 (equation (3.23)) that the controlled PDE system (1.1-1.3) is just a mixture of a controlled Hamiltonian flow and an anti-gradient flow. Here we provide some estimates for each term of the flow which will be crucial in the proof of the comparison principle. First, an adaptation of the results in Theorem D.50 of [11] (see Remark D.51 there) gives the following monotonicity estimate for $\mathrm{grad}_\rho s$.

Lemma 3.6. *For $\rho,\gamma \in \mathcal{P}_2(\mathbb{R}^2)$ satisfying $I(\gamma)+I(\rho) < +\infty$,*

$$\langle -\mathrm{grad}_\rho s, \mathrm{grad}_\rho d^2\rangle_{-1,\rho} + \langle -\mathrm{grad}_\gamma s, \mathrm{grad}_\gamma d^2\rangle_{-1,\gamma} \le 0.$$

Unfortunately we do not have such a monotonicity property for the Hamiltonian term J-$\mathrm{grad}_\rho e$. However, we have the following approximate result.

Lemma 3.7. *Suppose that $\rho,\gamma \in \mathcal{P}_2(\mathbb{R}^2)$ satisfy $I(\gamma)+I(\rho) < +\infty$. Then for each $\delta > 0$ there exists a constant $C_\delta \ge 0$ such that*

$$\begin{aligned}&\langle J\text{-}\mathrm{grad}_\rho e, \mathrm{grad}_\rho d^2\rangle_{-1,\rho} + \langle J\text{-}\mathrm{grad}_\gamma e, \mathrm{grad}_\gamma d^2\rangle_{-1,\gamma}\\ &\quad\le \delta d(\rho,\gamma)\sqrt{I(\rho)+I(\gamma)}\left((\tilde{C}+s(\rho)+M_2(\rho))^{\frac{3}{2}} + (\tilde{C}+s(\gamma)+M_2(\gamma))^{\frac{3}{2}}\right) + C_\delta d^2(\rho,\gamma),\end{aligned}$$

where $\tilde{C}$ is from Lemma 7.16.

Proof. We denote

$$u(x) = -K^{\perp} * \rho(x), \quad v(y) = -K^{\perp} * \gamma(y).$$

Let $\pi_0 \in \Gamma_o(\rho, \gamma)$, the set of optimal measures. Then by (3.21),

$$\begin{aligned} & \langle J\text{-grad}_\rho e, \text{grad}_\rho d^2 \rangle_{-1,\rho} + \langle J\text{-grad}_\gamma e, \text{grad}_\gamma d^2 \rangle_{-1,\gamma} \\ = & \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} (x-y)(u(x) - v(y)) \pi_0(dx, dy) \end{aligned}$$

Let $\hat{J}_\delta$ be the kernel approximating $-K^{\perp}$ from Lemma 7.16, i.e. $\hat{J}_\delta$ is equal to $\tilde{J}_k$ from Lemma 7.15 for big enough k, and let

$$u_\delta = \hat{J}_\delta * \rho, \quad v_\delta = \hat{J}_\delta * \gamma$$

Since $\hat{J}_\delta$ is Lipschitz we have

$$(z - z')(\hat{J}_\delta(z) - \hat{J}_\delta(z')) \leq C_\delta |z - z'|^2.$$

for all $z, z' \in \mathbb{R}^2$. It now follows that

$$\begin{aligned} & \int (x-y)(u_\delta(x) - v_\delta(y)) \pi_0(dx, dy) \\ = & \int_{x,y} \int_{x',y'} (x-y) \Big(\hat{J}_\delta(x - x') - \hat{J}_\delta(y - y') \Big) \pi_0(dx', dy') \pi_0(dx, dy) \\ = & \frac{1}{2} \int_{x,y} \int_{x',y'} \Big(x - x' - (y - y') \Big) \Big(\hat{J}_\delta(x - x') - \hat{J}_\delta(y - y') \Big) \pi_0(dx', dy') \pi_0(dx, dy) \\ \leq & \frac{1}{2} C_\delta \int |x-y|^2 \pi_0(dx, dy) = \frac{1}{2} C_\delta d^2(\rho, \gamma). \end{aligned}$$

The second equality above follows from the fact that $\hat{J}_\delta(-z) = -\hat{J}_\delta(z)$ and symmetry of the integral.

On the other hand, by Lemma 7.16,

$$\begin{aligned} & \int (x-y) \Big((u(x) - u_\delta(x)) + (v(y) - v_\delta(y)) \Big) \pi_0(dx, dy) \\ \leq & \sqrt{\int\int |x-y|^2 d\pi_0(dx, dy)} \Big(\sqrt{\int\int |u - u_\delta|^2 \rho(dx)} + \sqrt{\int\int |v - v_\delta|^2 \gamma(dy)} \Big) \\ \leq & d(\rho, \gamma) \sqrt{\delta} \sqrt{I(\rho) + I(\gamma)} \left((\tilde{C} + s(\rho) + M_2(\rho))^{\frac{3}{2}} + (\tilde{C} + s(\gamma) + M_2(\gamma))^{\frac{3}{2}} \right) \end{aligned}$$

which completes the proof.

□

3.3. **Controlled gradient-Hamiltonian flow.** Using (3.3), Lemma 3.4, and Corollary 2.6 we can interpret (1.1) as a controlled gradient-Hamiltonian system.

Lemma 3.8. *Let $\rho(\cdot) \in \mathcal{K}_{\rho_0}$, where $s(\rho(0)) < +\infty$. Then the weak solution of the controlled PDE (1.1) can be re-written as a mixture of a Hamiltonian flow, negative gradient flow, and control variable in the form*

$$\partial_t \rho = J\text{-grad}_\rho e - \nu \text{grad}_\rho s + m, \tag{3.23}$$

where $m = -\nabla \cdot (\rho \nabla p)$.

4. Viscosity solutions

In this section we introduce the notion of viscosity solution. We recall that $E = \mathcal{P}_2(\mathbb{R}^2)$. For $\rho \in E$ such that $I(\rho) < +\infty$ and $\tilde{m} \in H_{-1,\rho}$ we define

$$
\begin{aligned}
H(\rho, \tilde{m}) &= \sup_{m \in H_{-1,\rho}} \left[\langle J\text{-grad}_\rho e - \nu \operatorname{grad}_\rho s, \tilde{m} \rangle_{-1,\rho} + \langle m, \tilde{m} \rangle_{-1,\rho} - \frac{1}{4\nu} \|m\|^2_{-1,\rho} \right] \\
&= \langle J\text{-grad}_\rho e - \nu \operatorname{grad}_\rho s, \tilde{m} \rangle + \nu \|\tilde{m}\|^2_{-1,\rho}.
\end{aligned} \tag{4.1}
$$

For $p \in C_c^\infty(\mathbb{R}^2)$ we denote $m_p := -\nabla \cdot (\rho \nabla p) \in H_{-1,\rho}$. We then define

$$
H_p(\rho, \tilde{m}) = \langle J\text{-grad}_\rho e - \nu \operatorname{grad}_\rho s, \tilde{m} \rangle_{-1,\rho} + \langle m_p, \tilde{m} \rangle_{-1,\rho} - \frac{1}{4\nu} \|m_p\|^2_{-1,\rho}.
$$

We notice that

$$
H(\rho, \tilde{m}) = \sup_{p \in C_c^\infty(\mathbb{R}^2)} H_p(\rho, \tilde{m}). \tag{4.2}
$$

Let $\alpha > 0$ and $h \in C_b(E)$. We want to show that the value function

$$
f(\rho_0) = \sup \left\{ \int_0^\infty e^{-\alpha^{-1} t} \Big(\alpha^{-1} h(\rho(t)) - L(\rho(t), \dot{\rho}(t)) \Big) dt : \rho \in \mathcal{K}_{\rho_0} \right\} \tag{4.3}
$$

is the unique viscosity solution of the HJB equation

$$
f - \alpha H(\rho, \operatorname{grad}_\rho f) = h(\rho). \tag{4.4}
$$

We need to define the notion of viscosity solution.

Definition 4.1. We say that $\psi : E \to \mathbb{R} \cup \{+\infty\}$ is a test function if

$$
\psi(\rho) = \theta s(\rho) + \delta M_2(\rho) + \sum_{k=1}^\infty \beta_k d^2(\rho, \gamma^k) + c,
$$

where $0 < \theta < 1, \delta > 0, c \in \mathbb{R}$, $\beta_k \geq 0$ for $k \geq 1$, $\sum_{k=1}^\infty \beta_k < +\infty$, the set $\{\gamma^k : k \geq 1\}$ is bounded and every γ^k has Lebesgue density.

Obviously $\sum_{k=1}^\infty \beta_k d^2(\rho, \gamma^k)$ is continuous in E. Since convergence in E implies convergence of second moments also $M_2(\rho)$ is continuous in E. It is also well known that $s(\rho)$ is lower semicontinuous in E, for instance using the lower semicontinuity of relative entropy (see for instance Lemma 1.4.3 of [8]) and continuity of $M_2(\rho)$. In fact every test function ψ is narrowly lower semicontinuous, and its level sets $\{\rho : \psi(\rho) \leq r\}$ are compact in the narrow topology of E.

We recall that if $I(\rho) < +\infty$ then $\rho \in D(\psi)$ and $\operatorname{grad}_\rho \psi(\rho) \in H_{-1,\rho}$ and then $H(\rho, \operatorname{grad}_\rho \psi(\rho))$ is well defined. Otherwise we set

$$
H(\rho, \operatorname{grad}_\rho \psi(\rho)) = -\infty \quad \text{if } I(\rho) = +\infty
$$

and for every $p \in C_c^\infty(\mathbb{R}^2)$

$$
H_p(\rho, -\operatorname{grad}_\rho \psi(\rho)) = H(\rho, -\operatorname{grad}_\rho \psi(\rho)) = +\infty \quad \text{if } I(\rho) = +\infty.
$$

Definition 4.2. We say that $g : E \to \mathbb{R}$ is a viscosity subsolution of (4.4) if whenever $(g - \psi)^*$ has a local maximum at ρ_0 for a test function ψ, we have

$$
(g - \psi)^*(\rho_0) + \psi(\rho_0) - \alpha H(\rho_0, \operatorname{grad}_{\rho_0} \psi(\rho_0)) \leq h(\rho_0).
$$

We say that $g : E \to \mathbb{R}$ is a viscosity supersolution of (4.4) if whenever $(g+\psi)_*$ has a local minimum at ρ_0 for a test function ψ, we have

$$(g+\psi)_*(\rho_0) - \psi(\rho_0) - \alpha H(\rho_0, -\mathrm{grad}_{\rho_0}\psi(\rho_0)) \geq h(\rho_0).$$

A function $g : E \to \mathbb{R}$ is a viscosity solution of (4.4) if it is both a viscosity subsolution and a viscosity supersolution of (4.4).

Above, $(g-\psi)^*$ (respectively, $(g-\psi)_*$) means the upper (respectively, lower) semi-continuous envelope of $g-\psi$.

The reader may wonder what roles the various parts of test functions play. The part $\theta s(\rho) + \delta M_2(\rho)$ plays a role of a cutoff function and a function which will produce coercive terms in the equation. The functions of the form $\sum_{k=1}^{\infty} \beta_k d^2(\rho, \gamma^k)$ contain doubling functions and functions which allow to perform perturbed optimization (see the next section). We also mention that the first definition of discontinuous viscosity solution, for equations with unbounded terms in a Hilbert space, was introduced by Ishii in [16].

Remark 4.3. We notice that the above definition of viscosity solution implies that ρ_0 in the subsolution and the supersolution parts necessarily satisfies $I(\rho_0) < +\infty$ so in fact $H(\rho_0, \pm\mathrm{grad}_{\rho^0}\psi(\rho_0))$ is always finite and we don't have to use the extensions of H introduced before and it is enough to use the definition of H given by (4.1).

5. The comparison principle

In this section we will show a comparison result. We first recall a variational principle of Borwein-Preiss (see [6], Theorem 2.6 and Remark 2.7) which is stated here in the form adapted to our case and which can be easily deduced from the proof of Theorem 2.6 of [6].

Lemma 5.1. *Let $g : E \times E \to [-\infty, +\infty)$ be upper semicontinuous and such that $g(\rho, \gamma) = -\infty$ if either ρ or γ does not have Legesgue density. Let for $n \geq 1$, (ρ^0, γ^0) be such that*

$$g(\rho^0, \gamma^0) > \sup_{E\times E} g - \frac{1}{n}.$$

Then there exist sequences $(\tilde{\rho}^k), (\tilde{\gamma}^k)$ of measures that have Lebesgue densities such that $d(\tilde{\rho}^k, \rho^0) \leq 1, d(\tilde{\gamma}^k, \gamma^0) \leq 1, k \geq 1$, $\tilde{\rho}^k \to \rho^n, \tilde{\gamma}^k \to \gamma^n$ for some $\rho^n, \gamma^n \in E$, and sequences of nonnegative numbers $(\beta_k^1), (\beta_k^2)$ such that $\sum_{k=1}^{+\infty} \beta_k^i = 1, i = 1, 2$, such that

$$g(\rho^n, \gamma^n) > \sup_{E\times E} g - \frac{1}{n}$$

and

$$g(\rho^n, \gamma^n) - \frac{1}{n}\sum_{k=1}^{+\infty} \beta_k^1 d^2(\rho^n, \tilde{\rho}^k) - \frac{1}{n}\sum_{k=1}^{+\infty} \beta_k^2 d^2(\gamma^n, \tilde{\gamma}^k)$$
$$\geq g(\rho, \gamma) - \frac{1}{n}\sum_{k=1}^{+\infty} \beta_k^1 d^2(\rho, \tilde{\rho}^k) - \frac{1}{n}\sum_{k=1}^{+\infty} \beta_k^2 d^2(\gamma, \tilde{\gamma}^k).$$

for all $(\rho, \gamma) \in E \times E$.

Theorem 5.2. *Let $\alpha > 0$ and $h_i \in C_b(E), i = 1, 2$. Let $\overline{f}$ and $\underline{f}$ be two bounded functions such that $\overline{f}$ is a viscosity subsolution of*

$$f - \alpha H f = h_1 \tag{5.1}$$

and $\underline{f}$ is a viscosity supersolution of

$$f - \alpha Hf = h_2. \tag{5.2}$$

Then

$$\lim_{R\to+\infty} \limsup_{r\to 0} \big\{ \overline{f}(\rho) - \underline{f}(\gamma) : (s(\rho) + M_2(\rho)) + (s(\gamma) + M_2(\gamma)) \le R, d(\rho,\gamma) \le r \big\} \tag{5.3}$$
$$\le \lim_{R\to+\infty} \limsup_{r\to 0} \big\{ h_1(\rho) - h_2(\gamma) : (s(\rho) + M_2(\rho)) + (s(\gamma) + M_2(\gamma)) \le R, d(\rho,\gamma) \le r \big\}.$$

Proof. If (5.3) does not hold then there exists $\delta > 0$ such that if $0 < \beta - 1$ is sufficiently small we have

$$\begin{aligned} k_\beta &:= \lim_{R\to+\infty} \limsup_{r\to 0} \big\{ \beta\overline{f}(\rho) - \underline{f}(\gamma) : (s(\rho) + M_2(\rho)) + (s(\gamma) + M_2(\gamma)) \le R, d(\rho,\gamma) \le r \big\} \\ &> \delta + \lim_{R\to+\infty} \limsup_{r\to 0} \big\{ \beta h_1(\rho) - h_2(\gamma) : (s(\rho) + M_2(\rho)) + (s(\gamma) + M_2(\gamma)) \le R, \\ &\qquad d(\rho,\gamma) \le r \big\} = \delta + \tilde{k}_\beta. \end{aligned} \tag{5.4}$$

Set for $0 < \theta < 1$

$$\begin{aligned} k_{\beta,\theta} &:= \limsup_{r\to 0} \big\{ \beta\overline{f}(\rho) - \theta(s(\rho) + M_2(\rho)) - \underline{f}(\gamma) - \theta(s(\gamma) + M_2(\gamma)) : d(\rho,\gamma) \le r \big\} \\ &= \limsup_{r\to 0} \big\{ (\beta\overline{f} - \theta(s + M_2))^*(\rho) - (\underline{f} + \theta(s + M_2))_*(\gamma) : d(\rho,\gamma) \le r \big\}, \end{aligned}$$

$$k_{\beta,\theta,\epsilon} := \sup \big\{ (\beta\overline{f} - \theta(s + M_2))^*(\rho) - (\underline{f} + \theta(s + M_2))_*(\gamma) - \frac{1}{2\epsilon} d^2(\rho,\gamma) \big\}.$$

We remind that

$$s(\rho) + M_2(\rho) \ge C_1 \quad \text{for all } \rho \in E \tag{5.5}$$

for some constant C_1, and then it is easy to see that

$$k_\beta = \lim_{\theta\to 0} k_{\beta,\theta}, \tag{5.6}$$

$$k_{\beta,\theta} = \lim_{\kappa\to 0} k_{\beta,\theta,\epsilon}, \tag{5.7}$$

Denote

$$\overline{f}_{\beta,\theta}(\rho) = (\beta\overline{f} - \theta(s + M_2))^*(\rho), \quad \underline{f}_{\beta,\theta}(\gamma) = (\underline{f} + \theta(s + M_2))_*(\gamma).$$

We notice that if $\overline{f}_{\beta,\theta}(\rho) > -\infty$ then $s(\rho) < +\infty$ and so ρ must be absolutely continuous with respect to the Lebesgue measure, and if $\underline{f}_{\beta,\theta}(\gamma) < +\infty$ then $s(\gamma) < +\infty$ and thus γ has Lebesgue density. Therefore we can apply Lemma 5.1 to obtain for $n \ge 1$ sequences $(\tilde{\rho}^k), (\tilde{\gamma}^k)$ of measures that have Lebesgue densities such that $\tilde{\rho}^k \to \rho^n, \tilde{\gamma}^k \to \gamma^n$ for some $\rho^n, \gamma^n \in E$, where $d(\tilde{\rho}^k, \rho^n) \le 1, d(\tilde{\gamma}^k, \gamma^n) \le 1, k \ge 1$ and sequences of nonnegative numbers $(\beta_k^1), (\beta_k^2)$ satisfying $\sum_{k=1}^{+\infty} \beta_k^i = 1, i = 1, 2$, such that

$$\begin{aligned} &\overline{f}_{\beta,\theta}(\rho^n) - \underline{f}_{\beta,\theta}(\gamma^n) - \frac{1}{2\epsilon} d^2(\rho^n,\gamma^n) \\ &\quad > \sup_{E\times E} \left\{ \overline{f}_{\beta,\theta}(\rho) - \underline{f}_{\beta,\theta}(\gamma) - \frac{1}{2\epsilon} d^2(\rho,\gamma) \right\} - \frac{1}{n} = k_{\beta,\theta,\epsilon} - \frac{1}{n} \end{aligned} \tag{5.8}$$

and

$$\overline{f}_{\beta,\theta}(\rho^n) - \underline{f}_{\beta,\theta}(\gamma^n) - \frac{1}{2\epsilon}d^2(\rho^n,\gamma^n) - \frac{1}{n}\sum_{k=1}^{+\infty}\beta_k^1 d^2(\rho^n,\tilde{\rho}^k) - \frac{1}{n}\sum_{k=1}^{+\infty}\beta_k^2 d^2(\gamma^n,\tilde{\gamma}^k)$$
$$\geq \overline{f}_{\beta,\theta}(\rho) - \underline{f}_{\beta,\theta}(\gamma) - \frac{1}{2\epsilon}d^2(\rho,\gamma) - \frac{1}{n}\sum_{k=1}^{+\infty}\beta_k^1 d^2(\rho,\tilde{\rho}^k) - \frac{1}{n}\sum_{k=1}^{+\infty}\beta_k^2 d^2(\gamma,\tilde{\gamma}^k).$$

for all $(\rho,\gamma) \in E \times E$. Moreover it follows from (5.5) and (5.8) that

$$s(\rho^n) + M_2(\rho^n) \leq C(\theta), \quad s(\gamma^n) + M_2(\gamma^n) \leq C(\theta) \tag{5.9}$$

for some constant $C(\theta)$ independent of n. It is also a rather straightforward observation (see [16] for similar arguments) that

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{2\epsilon}d^2(\rho^n,\gamma^n) = 0 \quad \text{for every } \theta, \tag{5.10}$$

$$\lim_{\theta\to 0}\limsup_{\epsilon\to 0}\limsup_{n\to\infty}\left(\theta(s(\rho^n)+M_2(\rho^n)) + \theta(s(\gamma^n)+M_2(\gamma^n))\right) = 0. \tag{5.11}$$

To see these, let $\rho_j^n \to \rho^n, \gamma_j^n \to \gamma^n$ be such that

$$\overline{f}_{\beta,\theta}(\rho^n) = \lim_{j\to+\infty}(\beta\overline{f} - \theta(s+M_2))(\rho_j^n),$$
$$\underline{f}_{\beta,\theta}(\gamma^n) = \lim_{j\to+\infty}(\underline{f} + \theta(s+M_2))(\gamma_j^n).$$

We than have from (5.8) that for big j

$$k_{\beta,\theta,\epsilon} < (\beta\overline{f} - \theta(s+M_2))(\rho_j^n) - (\underline{f} + \theta(s+M_2))(\gamma_j^n) - \frac{1}{2\epsilon}d^2(\rho_j^n,\gamma_j^n) + \frac{1}{n}.$$

Therefore

$$k_{\beta,\theta,\epsilon} + \frac{1}{4\epsilon}d^2(\rho_j^n,\gamma_j^n) < (\beta\overline{f} - \theta(s+M_2))(\rho_j^n)$$
$$- (\underline{f} + \theta(s+M_2))(\gamma_j^n) - \frac{1}{4\epsilon}d^2(\rho_j^n,\gamma_j^n) + \frac{1}{n} \leq k_{\beta,\theta,2\epsilon} + \frac{1}{n},$$

which implies

$$k_{\beta,\theta,\epsilon} + \frac{1}{4\epsilon}d^2(\rho^n,\gamma^n) \leq k_{\beta,\theta,2\epsilon} + \frac{1}{n}. \tag{5.12}$$

Likewise we obtain

$$k_{\beta,\theta,\epsilon} + \frac{1}{4\epsilon}d^2(\rho_j^n,\gamma_j^n) + \frac{\theta}{2}((s+M_2)(\rho_j^n) + (s+M_2)(\gamma_j^n)) < k_{\beta,\frac{\theta}{2},2\epsilon} + \frac{1}{n},$$

which, using the lower semicontinuity of $s + M_2$, gives

$$k_{\beta,\theta,\epsilon} + \frac{1}{4\epsilon}d^2(\rho^n,\gamma^n) + \frac{\theta}{2}((s+M_2)(\rho^n) + (s+M_2)(\gamma^n)) \leq k_{\beta,\frac{\theta}{2},2\epsilon} + \frac{1}{n}. \tag{5.13}$$

It now remains to take $\lim_{\epsilon\to 0}\limsup_{n\to+\infty}$ in (5.12) and use (5.7), then take the limits $\lim_{\theta\to 0}\limsup_{\epsilon\to 0}\limsup_{n\to+\infty}$ in (5.13) and use (5.6) to produce (5.10) and (5.11) respectively.

We also observe that (5.6),(5.7), (5.8) and (5.10), imply

$$\lim_{\theta\to 0}\lim_{\epsilon\to 0}\limsup_{n\to\infty}(\overline{f}_{\beta,\theta}(\rho^n) - \underline{f}_{\beta,\theta}(\gamma^n)) = k_\beta. \tag{5.14}$$

Finally we notice that since $s + M_2/2 > -c_1$ for some constant $c_1 > 0$, we have $\theta M_2 \leq 2\theta(s + M_2 + c_1)$ and this, together with (5.11) yields

$$\lim_{\theta\to 0}\limsup_{\epsilon\to 0}\limsup_{n\to\infty} \theta(M_2(\rho^n) + M_2(\gamma^n)) = 0. \tag{5.15}$$

Set

$$\varphi(\rho) = \sum_{k=1}^{+\infty} \beta_k^1 d^2(\rho, \tilde\rho^k), \quad \psi(\gamma) = \sum_{k=1}^{+\infty} \beta_k^2 d^2(\gamma, \tilde\gamma^k).$$

We can now use the definition of viscosity solution. We recall that we have $I(\rho^n)+I(\gamma^n) < +\infty$. For $\overline{f}$ we obtain

$$\begin{aligned}
&\overline{f}_{\beta,\theta,\kappa}(\rho^n) - \alpha\Big[\langle J\text{-grad}_{\rho^n} e(\rho^n) - \nu\,\text{grad}_{\rho^n} s(\rho^n), \theta\text{grad}_{\rho^n} s(\rho^n)\\
&+ \theta\text{grad}_{\rho^n} M_2(\rho^n) + \frac{1}{n}\text{grad}_{\rho^n}\varphi_n(\rho^n) + \frac{1}{2\epsilon}\text{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\rho^n}\\
&+ \frac{\nu}{\beta}\left\|\theta\text{grad}_{\rho^n} s(\rho^n) + \theta\text{grad}_{\rho^n} M_2(\rho^n) + \frac{1}{n}\text{grad}_{\rho^n}\varphi_n(\rho^n) + \frac{1}{2\epsilon}\text{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\right\|_{-1,\rho^n}^2\Big]\\
&\qquad - \beta h_1(\rho^n) \leq C_1\theta,
\end{aligned} \tag{5.16}$$

where C_1 is the constant from (5.5). We now estimate, using (3.14)-(3.16), (3.17)- (3.22), Lemma 7.5, and standard Schwarz's and Young's inequalities,

$$\begin{aligned}
&\alpha\langle J\text{-grad}_{\rho^n} e(\rho^n), \theta\text{grad}_{\rho^n} s(\rho^n)\\
&\qquad + \theta\text{grad}_{\rho^n} M_2(\rho^n) + \frac{1}{n}\text{grad}_{\rho^n}\varphi_n(\rho^n) + \frac{1}{2\epsilon}\text{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\rho^n}\\
&\leq \alpha\big[\langle J\text{-grad}_{\rho^n} e(\rho^n), \frac{1}{2\epsilon}\text{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\rho^n}\\
&\qquad + \frac{1}{n}\|J\text{-grad}_{\rho^n} e(\rho^n)\|_{-1,\rho^n}^2 + \frac{C}{n}\|\text{grad}_{\rho^n}\varphi_n(\rho^n)\|_{-1,\rho^n}^2\big]\\
&\leq \alpha\langle J\text{-grad}_{\rho^n} e(\rho^n), \frac{1}{2\epsilon}\text{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\rho^n} + \frac{C_\theta}{n}(1 + I(\rho^n)) + \frac{C}{n},
\end{aligned} \tag{5.17}$$

and

$$\begin{aligned}
&-\alpha\nu\langle \text{grad}_{\rho^n} s(\rho^n), \theta\text{grad}_{\rho^n} s(\rho^n)\\
&\qquad + \theta\text{grad}_{\rho^n} M_2(\rho^n) + \frac{1}{n}\text{grad}_{\rho^n}\varphi_n(\rho^n) + \frac{1}{2\epsilon}\text{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\rho^n}\\
&\leq -\alpha\nu\theta I(\rho^n) + 4\alpha\nu\theta + \frac{1}{n}I(\rho^n) + \frac{C}{n}\|\text{grad}_{\rho^n}\varphi_n(\rho^n)\|_{-1,\rho^n}^2\\
&\qquad - \alpha\nu\langle \text{grad}_{\rho^n} s(\rho^n), \frac{1}{2\epsilon}\text{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\rho^n}\\
&\leq -\left(\alpha\nu\theta - \frac{1}{n}\right) I(\rho^n) + 4\alpha\nu\theta + \frac{C}{n} - \alpha\nu\langle \text{grad}_{\rho^n} s(\rho^n), \frac{1}{2\epsilon}\text{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\rho^n},
\end{aligned} \tag{5.18}$$

and

$$\frac{\alpha\nu}{\beta}\Big\|\theta \mathrm{grad}_{\rho^n} s(\rho^n) + \theta \mathrm{grad}_{\rho^n} M_2(\rho^n) + \frac{1}{n}\mathrm{grad}_{\rho^n}\varphi_n(\rho^n) + \frac{1}{2\epsilon}\mathrm{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\Big\|^2_{-1,\rho_i^n}$$

$$\le \frac{\alpha\nu}{\beta}\left[\left(2\theta^2 + \frac{\theta\beta}{4}\right) I(\rho^n) + \left(2\theta^2+\theta\right) M_2(\rho^n) + C\left(\theta^2 + \frac{1}{n^2} + \frac{1}{n}\right) + \frac{1}{\epsilon^2}\left(1 + C_2\left(\theta + \frac{1}{n}\right)\right) d^2(\rho^n,\gamma^n)\right], \tag{5.19}$$

where we have estimated

$$\langle \theta \mathrm{grad}_{\rho^n} s(\rho^n) + \theta \mathrm{grad}_{\rho^n} M_2(\rho^n) + \frac{1}{n}\mathrm{grad}_{\rho^n}\varphi_n(\rho^n), \frac{1}{2\epsilon}\mathrm{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\rho^n}$$
$$\le \frac{\theta\beta}{4} I(\rho^n) + \theta\Psi_2(\rho^n) + \frac{1}{n}\|\mathrm{grad}_{\rho^n}\varphi_n(\rho^n)\|^2_{-1,\rho^n} + \frac{1}{\epsilon^2}C_2\left(\theta + \frac{1}{n}\right)\|\mathrm{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\|_{-1,\rho^n}.$$

Let now $\tilde\delta > 0$ be such that

$$\tilde\delta < (1-\tilde\delta) - \frac{1+\tilde\delta}{\beta}. \tag{5.20}$$

Then, if

$$\frac{2\alpha\nu\theta^2}{\beta} < \frac{\alpha\nu\theta}{8}, \quad \frac{C_\theta + 1}{n} < \frac{\alpha\nu\theta}{8}, \quad C_2\left(\theta + \frac{1}{n}\right) < \tilde\delta,$$

we see from (5.16)-(5.19) and (5.15) that

$$\overline{f}_{\beta,\theta,\kappa}(\rho^n) - \alpha\langle J\text{-}\mathrm{grad}_{\rho^n} e(\rho^n) - \nu\,\mathrm{grad}_{\rho^n} s(\rho^n), \frac{1}{2\epsilon}\mathrm{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\rho^n} + \frac{\alpha\nu\theta}{2} I(\rho^n) - \frac{\alpha\nu(1+\tilde\delta)}{\beta}\frac{1}{\epsilon^2} d^2(\rho^n,\gamma^n) - \beta h_1(\rho^n) \le \omega(\theta,\epsilon,n), \tag{5.21}$$

where $\lim_{\theta\to 0}\limsup_{\epsilon\to 0}\limsup_{n\to\infty}\omega(\theta,\epsilon,n) = 0$.

Repeating similar arguments we also obtain

$$\underline{f}_{\beta,\theta,\kappa}(\gamma^n) + \alpha\langle J\text{-}\mathrm{grad}_{\gamma^n} e(\gamma^n) - \nu\,\mathrm{grad}_{\gamma^n} s(\gamma^n), \frac{1}{2\epsilon}\mathrm{grad}_{\gamma^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\gamma^n} - \frac{\alpha\nu\theta}{2} I(\gamma^n) - \alpha\nu(1-\tilde\delta)\frac{1}{\epsilon^2} d^2(\rho^n,\gamma^n) - h_1(\gamma^n) \ge -\omega(\theta,\epsilon,n). \tag{5.22}$$

Combining (5.21) with (5.22) and using (5.20) yields

$$\begin{aligned} &\overline{f}_{\beta,\theta,\kappa}(\rho^n) - \underline{f}_{\beta,\theta,\kappa}(\gamma^n) + \frac{\alpha\nu\theta}{2}(I(\rho^n) + I(\gamma^n)) + \alpha\nu\tilde\delta\frac{1}{\epsilon^2} d^2(\rho^n,\gamma^n) \\ &- \alpha\langle J\text{-}\mathrm{grad}_{\rho^n} e(\rho^n) - \nu\,\mathrm{grad}_{\rho^n} s(\rho^n), \frac{1}{2\epsilon}\mathrm{grad}_{\rho^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\rho^n} \\ &- \alpha\langle J\text{-}\mathrm{grad}_{\gamma^n} e(\gamma^n) - \nu\,\mathrm{grad}_{\gamma^n} s(\gamma^n), \frac{1}{2\epsilon}\mathrm{grad}_{\gamma^n} d^2(\rho^n,\gamma^n)\rangle_{-1,\gamma^n} \\ &\le \beta h_1(\rho^n) - h_1(\gamma^n) + \omega(\theta,\epsilon,n). \end{aligned} \tag{5.23}$$

It now follows from Lemmas 3.6 and 3.7, together with (5.9), applied with

$$\delta = \frac{2}{(\tilde C + C(\theta))^{\frac{3}{2}}}\sqrt{\frac{\nu\theta}{2}}\sqrt{\nu\tilde\delta},$$

that

$$
\begin{aligned}
&\overline{f}_{\beta,\theta,\kappa}(\rho^n) - \underline{f}_{\beta,\theta,\kappa}(\gamma^n) + \frac{\alpha\nu\theta}{2}(I(\rho^n)+I(\gamma^n)) + \alpha\nu\tilde{\delta}\frac{1}{\epsilon^2}d^2(\rho^n,\gamma^n) \\
&- 2\alpha\sqrt{\frac{\nu\theta}{2}}\sqrt{\nu\tilde{\delta}}\frac{1}{\epsilon}d(\rho^n,\gamma^n)\sqrt{I(\rho^n)+I(\gamma^n)} - C_\theta\frac{1}{\epsilon}d^2(\rho^n,\gamma^n) \\
&\leq \beta h_1(\rho^n) - h_1(\gamma^n) + \omega(\theta,\epsilon,n)
\end{aligned} \tag{5.24}
$$

which, together with (5.10), obviously implies

$$
\overline{f}_{\beta,\theta,\kappa}(\rho^n) - \underline{f}_{\beta,\theta,\kappa}(\gamma^n) \leq \beta h_1(\rho^n) - h_1(\gamma^n) + \omega(\theta,\epsilon,n). \tag{5.25}
$$

It now remains to apply $\lim_{\theta\to 0}\limsup_{\epsilon\to 0}\limsup_{n\to\infty}$ to both sides of (5.25) and invoke (5.10), (5.11) and (5.14) to obtain

$$
k_\beta \leq \tilde{k}_\beta
$$

which contradicts (5.4).

□

Corollary 5.3. *Let $\alpha > 0$ and $h \in C_b(E)$ and be uniformly continuous on finite level sets of $s + M_2$. Then the equation $f - \alpha H f = h$ has at most one bounded viscosity solution. Such a solution is uniformly continuous on finite level sets of $s + M_2$.*

6. Existence of viscosity solutions

We begin with a preliminary lemma.

Lemma 6.1. *Let ψ be a test function, $p \in C_c^\infty(\mathbb{R}^2)$. Then:*

(i) $(\lambda,\rho) \to \frac{1}{\lambda}H(\rho, \lambda\,\mathrm{grad}_\rho\psi(\rho))$ defined on $(0,+\infty)\times E$ is upper semicontinuous at $(1,\rho)$ for every ρ.

(ii) $(\lambda,\rho) \to \frac{1}{\lambda}H_p(\rho, -\lambda\,\mathrm{grad}_\rho\psi(\rho))$ defined on $(0,+\infty)\times E$ is lower semicontinuous at $(1,\rho)$. In particular, $(\lambda,\rho) \to \frac{1}{\lambda}H(\rho, -\lambda\,\mathrm{grad}_\rho\psi(\rho))$ is also lower semicontinuous at $(1,\rho)$.

Proof. We will only show (i) as the proof of (ii) is similar. We recall that if $I(\rho) < +\infty$ then

$$
\begin{aligned}
&\mathrm{grad}_\rho s(\rho) = -\nabla\cdot\left(\rho\frac{\nabla\rho}{\rho}\right) \\
&J\text{-}\mathrm{grad}_\rho e(\rho) = -\nabla\cdot\big(\rho(-K^\perp * \rho)\big) \\
&\mathrm{grad}_\rho M_2(\rho) = -\nabla\cdot(2x\rho) \\
&\mathrm{grad}_\rho\sum_{k=1}^\infty \beta_k d^2(\rho,\gamma^k) = -2\sum_{k=1}^\infty \beta_k\nabla\cdot\big(\rho\nabla p_{\rho,\gamma^k}\big).
\end{aligned} \tag{6.1}
$$

Therefore, using (3.14)-(3.22) and elementary Schwarz's and Young's inequalities, if λ is close to 1 so that $\theta/2 < \lambda\theta < (1+\theta)/2$ we obtain that there exists a constant $C_{\theta,\delta}$ and for every $\epsilon > 0$ a constant C_ϵ such that

$$
\begin{aligned}
H(\rho,\lambda\,\mathrm{grad}_\rho\psi(\rho)) \leq& -\frac{\nu\theta(1-\theta)}{4}I(\rho) + \epsilon\int_{\mathbb{R}^2}|u_\rho|^2\rho\,dx + \epsilon I(\rho) \\
&+ C_\epsilon\sum_{k=1}^\infty\beta_k d^2(\rho,\gamma^k) + C_{\theta,\delta}\left(1 + M_2(\rho) + \sum_{k=1}^\infty \beta_k d^2(\rho,\gamma^k)\right).
\end{aligned} \tag{6.2}
$$

Since by Lemma 7.5

$$\int_{\mathbb{R}^2} |u_\rho|^2 \rho \, dx \le C(1 + I(\rho)),$$

we therefore obtain that if ϵ is small enough then

$$H(\rho, \lambda \,\mathrm{grad}_\rho \psi(\rho)) \le -\frac{\nu\theta(1-\theta)}{8} I(\rho) + C_{\theta,\delta}\left(1 + M_2(\rho) + \sum_{k=1}^{\infty} \beta_k d^2(\rho, \gamma^k)\right). \tag{6.3}$$

This obviously shows that if $\rho^n \to \rho_0$ and $I(\rho^n) \to +\infty$ then $H(\rho, \lambda \,\mathrm{grad}_\rho \psi(\rho))$ is upper semicontinuous at $(1, \rho_0)$.

Therefore it remains to show the upper semicontinuity in the case when $\rho^n \to \rho_0$ and $\sup_n I(\rho^n) = R < +\infty$. In this case all the terms of the Hamiltonian are bounded so we can eliminate λ from our considerations and just set $\lambda = 1$.

We will be using a result proved in [11] (Lemma D.48) that in our case

$$\nabla\sqrt{\rho^n} = \frac{\nabla\rho^n}{2\sqrt{\rho^n}} \rightharpoonup \frac{\nabla\rho_0}{2\sqrt{\rho_0}} = \nabla\sqrt{\rho_0} \quad \text{weakly in } L^2(\mathbb{R}^2), \tag{6.4}$$

$$\sqrt{\rho^n} \to \sqrt{\rho_0} \quad \text{in } L^2(\mathbb{R}^2), \tag{6.5}$$

$$\sqrt{\rho^n}\nabla p_{\rho^n,\gamma^k} \to \sqrt{\rho_0}\nabla p_{\rho_0,\gamma^k} \quad \text{in } L^2(\mathbb{R}^2) \tag{6.6}$$

for every $k \ge 1$.

Since $\rho^n \to \rho_0$ in E, ρ^n have uniformly integrable 2nd moments. This, together with (6.5), implies that

$$x\sqrt{\rho^n} \to x\sqrt{\rho_0} \quad \text{in } L^2(\mathbb{R}^2). \tag{6.7}$$

We now notice that, by (3.14), $-\|\mathrm{grad}_\rho s(\rho)\|^2_{-1,\rho} = -I(\rho)$ is upper semicontinuous and, by (3.15) and (3.16), $\|\mathrm{grad}_\rho M_2(\rho)\|^2_{-1,\rho}$ and $\|\mathrm{grad}_\rho \sum_{k=1}^\infty \beta_k d^2(\rho,\gamma^k)\|^2_{-1,\rho}$ are continuous.

We will now show that

$$u_{\rho^n}\sqrt{\rho^n} \to u_{\rho_0}\sqrt{\rho_0} \quad \text{in } L^2(\mathbb{R}^2). \tag{6.8}$$

Let $\phi \in C_c^\infty(\mathbb{R}^2)$ be radial, $0 \le \phi \le 1$, $\phi(x) = 1$ for $|x| \le 1/2$ and $\phi(x) = 0$ for $|x| \ge 1$. We set for $r > 0$, $\phi_r(x) = \phi(x/r)$. We decompose $-K^\perp = -\phi_r K^\perp - (1-\phi_r)K^\perp =: K_{1,r} + K_{2,r}$ which induces the decomposition of u_ρ into $u_\rho = u_\rho^{1,r} + u_\rho^{2,r} = K_{1,r} * \rho + K_{2,r} * \rho$. Now

$$\begin{aligned}\|u_{\rho^n}\sqrt{\rho^n} - u_{\rho_0}\sqrt{\rho_0}\|_2 \le \|(u^{1,r}_{\rho^n} - u^{1,r}_{\rho_0})\sqrt{\rho^n}\|_2 + \|(u^{2,r}_{\rho^n} - u^{2,r}_{\rho_0})\sqrt{\rho^n}\|_2 \\ + \|u_{\rho_0}(\sqrt{\rho^n} - \sqrt{\rho_0})\|_2.\end{aligned}$$

Using Hölder inequality and Young's inequality for convolutions we obtain (see also (7.10))

$$\|(u^{1,r}_{\rho^n} - u^{1,r}_{\rho_0})\sqrt{\rho^n}\|_2 \le \|u^{1,r}_{\rho^n} - u^{1,r}_{\rho_0}\|_4 \|\rho^n\|_2^{1/2} \le \|K_{1,r}\|_{\frac{4}{3}} \|\rho_n - \rho_0\|_2 \|\rho^n\|_2^{1/2} \le C r^{1/2}, \tag{6.9}$$

$$\|(u^{2,r}_{\rho^n} - u^{2,r}_{\rho_0})\sqrt{\rho^n}\|_2 \le \|u^{2,r}_{\rho^n} - u^{2,r}_{\rho_0}\|_4 \|\rho^n\|_2^{1/2} \le \|K_{2,r}\|_4 \|\rho_n - \rho_0\|_1 \|\rho^n\|_2^{1/2} \to 0 \text{ as } n \to \infty, \tag{6.10}$$

and, by (7.2), (7.28) and (6.5),

$$\|u_{\rho_0}(\sqrt{\rho^n} - \sqrt{\rho_0})\|_2 \le \|u_{\rho_0}\|_8 \|\sqrt{\rho^n} - \sqrt{\rho_0}\|_4^{1/2} \|\sqrt{\rho^n} - \sqrt{\rho_0}\|_2^{1/2} \le C\|\sqrt{\rho^n} - \sqrt{\rho_0}\|_2^{1/2} \to 0 \text{ as } n \to \infty. \tag{6.11}$$

Thus (6.8) follows after we let $r \to 0$.

We now notice that

$$
\begin{aligned}
&|2\sum_{k=m}^{\infty}\beta_k\int_{\mathbb{R}^2}\nabla\rho^n\cdot\nabla p_{\rho^n,\gamma^k}dx|\le 2\sum_{k=m}^{\infty}\beta_k\sqrt{I(\rho^n)}\left(\int_{\mathbb{R}^2}|\nabla p_{\rho^n,\gamma^k}|^2\rho^n\,dx\right)^{\frac12}\\
&\le 2\sum_{k=m}^{\infty}\beta_k\sqrt{I(\rho^n)}\,d(\rho^n,\gamma^k)\le 2\left(\sum_{k=m}^{\infty}\beta_k\right)^{\frac12}\sqrt{I(\rho^n)}\left(\sum_{k=m}^{\infty}\beta_k d^2(\rho^n,\gamma^k)\right)^{\frac12}\\
&\le 2\left(\sum_{k=m}^{\infty}\beta_k\right)^{\frac12}\sqrt{R}\left(\sum_{k=m}^{\infty}\beta_k d^2(\rho^n,\gamma^k)\right)^{\frac12}\le\omega(\frac{1}{m})
\end{aligned}
$$

for some modulus ω independent of n, and the same estimate is also true for ρ^0.

Therefore continuity of the term in (3.21) it is enough to show the continuity of the partial sums

$$
\sum_{k=1}^{m}\beta_k\int_{\mathbb{R}^2}K^{\perp}*\rho^n\cdot\nabla p_{\rho^n,\gamma^k}dx
$$

for every $m\ge 1$. This is however clear by (6.6) and (6.8).

The continuity of the term in (3.18) (respectively, (3.22)) follows by the same argument if we use (6.4) and (6.6) (respectively, (6.6), (6.7)). ☐

To prove that the value function f is a viscosity solution of (4.4) we will use the principle of dynamic programming. Its proof follows standard arguments (see for instance [4]) since our definitions of weak solution of (1.1) and of admissible trajectories $\mathcal{K}_{\rho_0}$ allow us to concatenate admissible trajectories, i.e. if $\rho_1(\cdot)\in\mathcal{K}_{\rho_0}$, $\rho_2(\cdot)\in\mathcal{K}_{\rho_1(t)}$ and we define

$$
\rho(\tau)=\begin{cases}\rho_1(\tau) & \text{if } \tau\in(0,t]\\ \rho_2(\tau-t) & \text{if } \tau>t,\end{cases}
$$

then $\rho(\cdot)\in\mathcal{K}_{\rho_0}$.

Lemma 6.2. *The value function f satisfies the dynamic programming principle, i.e. for every $\rho_0\in E$ and $t>0$*

$$
f(\rho_0)=\sup_{\rho\in\mathcal{K}_{\rho_0}}\left\{\int_0^t e^{-\alpha^{-1}\tau}\left(\alpha^{-1}h(\rho(\tau))-L(\rho(\tau),\dot\rho(\tau))\right)d\tau+e^{-\alpha^{-1}t}f(\rho(t))\right\}. \tag{6.12}
$$

Theorem 6.3. *Let $\alpha>0$ and $h\in C_b(E)$. Then the value function f defined by (4.3) is a viscosity solution of (4.4).*

Proof. We will first show that f is a viscosity subsolution.

Let $(f-\psi)^*$ have a local maximum at ρ_0. Let $\rho^n\to\rho_0$ be such that

$$
f(\rho^n)-\psi(\rho^n)\ge(f-\psi)^*(\rho_0)-\frac{1}{n^2}.
$$

This implies that

$$
f(\rho^n)-\psi(\rho^n)\to(f-\psi)^*(\rho_0) \tag{6.13}
$$

and

$$
f(\rho)-\psi(\rho)\le f(\rho^n)-\psi(\rho^n)+\frac{1}{n^2} \tag{6.14}
$$

for ρ in some neighborhood of ρ_0. This obviously implies that

$$\sup_n s(\rho^n) = C_0 < +\infty. \tag{6.15}$$

By the dynamic programming principle for every $n \geq 1$ there exists $\tilde{\rho}^n \in \mathcal{K}_{\rho^n}$ and a corresponding m_{p^n} as in the definition of $\mathcal{K}_{\rho^n}$ such that

$$f(\rho^n) \leq \int_0^{\frac{1}{n}} e^{-\frac{\tau}{\alpha}} \left(\alpha^{-1} h(\tilde{\rho}^n(\tau)) - \frac{1}{4\nu} \|m_{p^n}(\tau)\|^2_{-1,\tilde{\rho}^n(\tau)} \right) d\tau + e^{-\frac{1}{\alpha n}} f(\tilde{\rho}^n(\frac{1}{n})) + \frac{1}{n^2}. \tag{6.16}$$

We notice for future use that a particular consequence of (6.16) and the boundedness of f and h is that

$$f(\rho^n) \leq \frac{R_1}{n} + f(\tilde{\rho}^n(\frac{1}{n})) \tag{6.17}$$

for some constant R_1 independent of n. Moreover, it also follows that

$$\sup_n \int_0^{\frac{1}{n}} \|m_{p^n}(\tau)\|^2_{-1,\tilde{\rho}^n(\tau)} d\tau = R_2 < +\infty \tag{6.18}$$

for some $R_2 \geq 0$. Therefore, (6.15), (6.18) and (2.49) of Corollary 2.6 imply

$$d(\tilde{\rho}^n(\tau), \tilde{\rho}^n(0)) \leq C\sqrt{\tau} \tag{6.19}$$

for some constant C independent of n.

Now, it follows from (6.14) and (6.16) that

$$\begin{aligned} & f(\tilde{\rho}^n(\frac{1}{n}))(1 - e^{-\frac{1}{\alpha n}}) - \frac{2}{n^2} \\ & \quad \leq \psi(\tilde{\rho}^n(\frac{1}{n})) - \psi(\tilde{\rho}^n(0)) + \int_0^{\frac{1}{n}} e^{-\frac{\tau}{\alpha}} \left(\alpha^{-1} h(\tilde{\rho}^n(\tau)) - \frac{1}{4\nu} \|m_{p^n}(\tau)\|^2_{-1,\tilde{\rho}^n(\tau)} \right) d\tau. \end{aligned}$$

Therefore, using (2.36), (2.37), (2.48) and writing the terms using the abstract gradient notation (6.1) we obtain

$$\begin{aligned} & \frac{1}{n\alpha} f(\tilde{\rho}^n(\frac{1}{n})) + o(\frac{1}{n}) \\ & \quad \leq \int_0^{\frac{1}{n}} \Bigg(\Bigg[\langle J\text{-grad}_{\tilde{\rho}^n(\tau)} e(\tilde{\rho}^n(\tau)) - \nu \, \text{grad}_{\tilde{\rho}^n(\tau)} s(\tilde{\rho}^n(\tau)) \\ & + m_{p^n}(\tau), \text{grad}_{\tilde{\rho}^n(\tau)} \psi(\tilde{\rho}^n(\tau)) \rangle_{-1,\tilde{\rho}^n(\tau)} - e^{-\frac{\tau}{\alpha}} \frac{1}{4\nu} \|m_{p^n}(\tau)\|^2_{-1,\tilde{\rho}^n(\tau)} \Bigg] + \alpha^{-1} e^{-\frac{\tau}{\alpha}} h(\tilde{\rho}^n(\tau)) \Bigg) d\tau \\ & \leq \int_0^{\frac{1}{n}} \left(e^{-\frac{\tau}{\alpha}} H\left(\tilde{\rho}^n(\tau), e^{\frac{\tau}{\alpha}} \text{grad}_{\tilde{\rho}^n(\tau)} \psi(\tilde{\rho}^n(\tau))\right) + \alpha^{-1} e^{-\frac{\tau}{\alpha}} h(\tilde{\rho}^n(\tau)) \right) d\tau. \end{aligned}$$

Multiplying the above inequality by n/α we therefore obtain that there exists $t_n \in (0, 1/n)$ such that

$$f(\tilde{\rho}^n(\frac{1}{n})) + no(\frac{1}{n}) \leq \alpha e^{-\frac{t_n}{\alpha}} H\left(\tilde{\rho}^n(t_n), e^{\frac{t_n}{\alpha}} \text{grad}_{\tilde{\rho}^n(t_n)} \psi(\tilde{\rho}^n(t_n))\right) + e^{-\frac{t_n}{\alpha}} h(\tilde{\rho}^n(t_n))$$

which, upon using (6.17), gives

$$(6.20)\qquad \begin{aligned}&(f(\rho^n)-\psi(\rho^n))+\psi(\rho^n)-\alpha e^{-\frac{t_n}{\alpha}}H\left(\tilde\rho^n(t_n),e^{\frac{t_n}{\alpha}}\mathrm{grad}_{\tilde\rho^n(t_n)}\psi(\tilde\rho^n(t_n))\right)\\&\qquad\le e^{-\frac{t_n}{\alpha}}h(\tilde\rho^n(t_n))+no(\frac{1}{n}).\end{aligned}$$

Therefore we can now pass to $\liminf_{n\to\infty}$ in (6.20) using $\rho^n\to\rho_0$, Lemma 6.1(i), (6.13), (6.19), and the lower semicontinuity of ψ to conclude that

$$(f-\psi)^*(\rho_0)+\psi(\rho_0)-\alpha H(\rho_0,\mathrm{grad}_{\rho_0}\psi(\rho_0))\le h(\rho_0).$$

To show the supersolution property let $(f+\psi)_*$ have a local minimum at ρ^0 and let $\rho^n\to\rho_0$ be such that

$$f(\rho^n)+\psi(\rho^n)\le(f+\psi)_*(\rho_0)+\frac{1}{n^2}.$$

As before we have

$$f(\rho^n)+\psi(\rho^n)\to(f+\psi)_*(\rho_0)$$

and

$$f(\rho)+\psi(\rho)\ge f(\rho^n)+\psi(\rho^n)-\frac{1}{n^2}$$

for ρ in some neighborhood of ρ_0.

Let $p\in C_c^\infty(\mathbb{R}^2)$ and let $m_p=-\nabla(\rho\nabla p)$. By the dynamic programming principle if $\tilde\rho^n(\cdot)\in\mathcal{K}_{\rho^n}$ corresponds to the control m_p then

$$f(\rho^n)\ge\int_0^{\frac{1}{n}}e^{-\frac{\tau}{\alpha}}\left(\alpha^{-1}h(\tilde\rho^n(\tau))-\frac{1}{4\nu}\|m_p^n(\tau)\|^2_{-1,\tilde\rho^n(\tau)}\right)d\tau+e^{-\frac{1}{\alpha n}}f(\tilde\rho^n(\frac{1}{n})).$$

This in particular implies that

$$(6.21)\qquad f(\rho^n)\ge f(\tilde\rho^n(\frac{1}{n}))+\omega(\frac{1}{n}),$$

where $\omega(\frac{1}{n})\to0$ as $n\to+\infty$ which now replaces (6.17). We can now repeat the steps of the proof of the subsolution property to obtain

$$f(\tilde\rho^n(\frac{1}{n}))+no(\frac{1}{n})\ge n\int_0^{\frac{1}{n}}\left(\alpha e^{-\frac{\tau}{\alpha}}H_p\left(\tilde\rho^n(\tau),e^{\frac{\tau}{\alpha}}\mathrm{grad}_{\tilde\rho^n(\tau)}\psi(\tilde\rho^n(\tau))\right)+e^{-\frac{\tau}{\alpha}}h(\tilde\rho^n(\tau))\right)d\tau$$

which, together with (6.21), implies

$$f(\tilde\rho^n(\frac{1}{n}))+no(\frac{1}{n})\ge\alpha e^{-\frac{t_n}{\alpha}}H_p\left(\tilde\rho^n(t_n),e^{\frac{t_n}{\alpha}}\mathrm{grad}_{\tilde\rho^n(t_n)}\psi(\tilde\rho^n(t_n))\right)+e^{-\frac{t_n}{\alpha}}h(\tilde\rho^n(t_n))$$

for some $t_n\in(0,1/n)$. Thus again we can send $n\to+\infty$, use Lemma 6.1(ii), and the lower semicontinuity of ψ to get

$$(f+\psi)_*(\rho_0)-\psi(\rho_0)-\alpha H_p(\rho_0,\mathrm{grad}_{\rho_0}\psi(\rho_0))\ge h(\rho_0)$$

for every $p\in C_c^\infty(\mathbb{R}^2)$. It remains to invoke (4.2) to conclude the proof. □

Remark 6.4. The main differences and difficulties (compared to [11, 10]) in the proof of the comparison theorem come from the terms involving J-$\mathrm{grad}_\rho e$. They are a result of the singular nature of the kernel $K^\perp$ and they required several new technical results. In particular the most difficult to control terms in (5.23) were estimated with the help of the key Lemma 3.7. Lemma 3.7 in turn, is proved using a technical result about an approximation of the velocity vector field u (Lemma 7.16), which requires harmonic analysis tools of Appendix

B. The terms involving J-$\text{grad}_\rho e$ also needed delicate treatment in the proof of existence of viscosity solutions, to show the semicontinuity of the Hamiltonian evaluated on test functions.

7. Appendix

7.1. Appendix A: Estimates regarding 2-D Newtonian and related potentials.

Recall that the kernels N and K are defined in (1.3),

$$U = N * \rho, \quad u = -K^{\perp} * \rho,$$

and the 2×2 matrix $Du = (\nabla u_1, \nabla u_2)$.

Lemma 7.1. *Let $\rho \in \mathcal{P}_2(\mathbb{R}^2)$. Then:*

(1) *$U \in L^p_{loc}(\mathbb{R}^2)$, for $p \in (1, \infty)$.*

(2) *$u \in L^p_{loc}(\mathbb{R}^2)$ for $p \in (0, 2)$; indeed, $u \in L^{2,\infty}(\mathbb{R}^2)$. Furthermore, $\nabla U = K * \rho$ and $u = -J\nabla U$.*

(3) *Let the singular integral operator P be defined through principal value integral as*

$$P\rho(x) = \text{P.V.-}\int_{\mathbb{R}^2} P(x-y)\rho(dy) = \lim_{\epsilon \to 0+} \int_{|x-y|\geq \epsilon} P(x-y)\rho(dy),$$

with

$$P(z) = \frac{1}{2\pi}\frac{\sigma(z)}{|z|^2}, \quad \sigma(z) = \frac{1}{|z|^2}\begin{bmatrix} -2z_1z_2 & z_1^2 - z_2^2 \\ z_1^2 - z_2^2 & 2z_1z_2 \end{bmatrix}.$$

Then as a Schwartz distribution

$$Du(x) = (P\rho)(x) + \frac{\rho(x)}{2}J. \tag{7.1}$$

In particular, $\text{div} u = 0$.

(4) *If $\|\rho\|_2 < +\infty$ then $u \in L^p(\mathbb{R}^2)$ for every $p > 2$, and*

$$\|u\|_p \leq C\|\rho\|_2^{1-2/p}, \tag{7.2}$$

which implies that for every $p > 2$ and $\epsilon > 0$

$$\|u\|_p \leq \epsilon\|\rho\|_2 + C_{\epsilon,p}. \tag{7.3}$$

Moreover $u \in L^2_{loc}(\mathbb{R}^2)$ and for every $R > 0$ we have

$$\|u\|^2_{L^2(B_R)} \leq C_R\|\rho\|_2. \tag{7.4}$$

Estimates (7.2)-(7.4) are also true for $u = K * \rho$.

(5) *For each $p \in (1, \infty)$, there exists a finite $C_p > 0$,*

$$\|Du\|_p \leq C_p\|\rho\|_p. \tag{7.5}$$

Proof. Let $\phi \in C_c^\infty(\mathbb{R}^2)$ be a radial function such that $0 \leq \phi \leq 1$, $\phi(x) = 1$ for $|x| \leq 1$ and $\phi(x) = 0$ for $|x| \geq 2$. First, we introduce a smooth cut-off function $\phi_B(x) = \phi(x/B)$, $B > 0$. Then

$$U = N * \rho = U_{1,B} + U_{2,B} \equiv (\phi_B N) * \rho + [(1-\phi_B)N] * \rho. \tag{7.6}$$

By Young's inequality for convolution and direct integration,

$$\|U_{1,B}\|_q \leq \|\phi_B N\|_q \|\rho\|_1 = C_{B,q}\|\rho\|_1 < \infty, \quad 1 \leq q < \infty. \tag{7.7}$$

Furthermore, because $\log r \le r$ for $r \ge 1$,

$$|U_{2,B}(x)| \le C_1 + \int_{\mathbb{R}^2} |x-y||\rho(y)|dy \le C_2(1 + \int_{\mathbb{R}^2} |y||\rho(y)|dy + |x|). \tag{7.8}$$

Therefore, $U \in L^p_{loc}(\mathbb{R}^2)$, $p \in (1,\infty)$.

Similarly,

$$u = -JK * \rho = u_{1,B} + u_{2,B} \equiv -J(\phi_B K) * \rho - J((1-\phi_B)K) * \rho. \tag{7.9}$$

Hence by Young's inequality

$$\|u_{1,B}\|_p \le \|\phi_B K\|_p \|\rho\|_1 < \infty, \quad \|u_{2,B}\|_{p'} \le \|(1-\phi_B)K\|_{p'} \|\rho\|_1 < \infty, \quad 1 \le p < 2 < p' \le \infty.$$

Hence $u \in L^p_{loc}(\mathbb{R}^2)$ for $p \in (0,2)$. Indeed, assuming $\rho(dx) = \rho(x)dx$, $|(K*\rho)(x)| \le (2\pi)^{-1}\int_{\mathbb{R}^2} \frac{1}{|y|}\rho(x-y)dy$, by an inequality regarding Riesz potential (e.g. Theorem 6.1.3 of Grafakos [15]),

$$\|K * \rho\|_{L^{2,\infty}(\mathbb{R}^2)} \le C\|\rho\|_{L^1(\mathbb{R}^2)}.$$

The general case of ρ without density follows by standard procedure of mollifying with a smooth convolution kernel.

The conclusion $\nabla U = K * \rho$ is standard and follows from Fubini's theorem and integration by parts against test functions $\varphi \in C_c^\infty(\mathbb{R}^2)$ to show the equality in the distribution sense.

The expression for ∇u in (7.1) is the second part of Proposition 2.20 in [20], which follows from Proposition 2.17 in [20] and Fubini theorem (noting $K \in L^1_{loc}(\mathbb{R}^2)$).

Regarding (7.2)-(7.4), using Young's inequality we have for $p \ge 2$

$$\|u_{1,B}\|_p \le \|\phi_B K\|_{\frac{2p}{p+2}} \|\rho\|_2 \le CB^{\frac{2}{p}} \|\rho\|_2 \tag{7.10}$$

and for $2 < p \le \infty$

$$\|u_{2,B}\|_p \le \|(1-\phi_B)K\|_p \|\rho\|_1 \le CB^{\frac{2-p}{p}} \tag{7.11}$$

which gives (7.2) if $B = \|\rho\|_2^{-1}$. In particular

$$\|u_{2,B}\|_\infty \le \frac{C}{B} \tag{7.12}$$

for some constant C, so taking $B = \|\rho\|_2^{-1/2}$ in (7.10) with $p=2$ gives (7.4). (7.3) follows from (7.2) because for any $\alpha \in (0,1)$ and $\epsilon > 0$, $r^\alpha \le \epsilon r + C_\epsilon$ for $r \ge 0$.

The estimate $\|Du\|_p \le C_p \|\rho\|_p$ follows from Calderon-Zygmund inequality $\|Pf\|_p \le c_p \|f\|_p$, $p \in (1,\infty)$ (or see [13], page 225). □

Using Green's formula, we can identify that as Schwartz distribution,

$$(-\Delta_x) N(x-y) = \delta_y(dx).$$

Therefore, as long as $(-\Delta)^{-1}$ can be reasonably defined (see next lemma), $(-\Delta)^{-1}\rho = N * \rho$. The following is Lemma 1.12 of [20]):

Lemma 7.2. *Let $g \in L^1(\mathbb{R}^d) \cap C^1(\mathbb{R}^2)$, $\int_{|x|\ge 1} (\log|x|)|g(x)|dx < \infty$. Then $U = N * g \in C^2(\mathbb{R}^2)$ and $-\Delta U = g$.*

Invoking estimate (7.7) for $p \in (1, +\infty]$ and (7.8), for each $B > 1$, there exists $C_B > 0$ such that

$$e(\rho) = \frac{1}{2} \int_{\mathbb{R}^2} U(x)\rho(dx) = \frac{1}{2}\langle U_{1,B} + U_{2,B}, \rho\rangle \le C_B\Big(\|\rho\|_p + 1 + M_1(\rho)\Big). \tag{7.13}$$

Finally, if we use the decomposition $u = u_{1,B} + u_{2,B} = u_1 + u_2$ as in (7.9), then, for $i = 1, 2$,

$$\|\partial_i u_2\|_2 = \|\partial_i((1 - \phi_B)K^{\perp}) * \rho\|_2 \le \|\partial_i((1 - \phi_B)K^{\perp})\|_2 \|\rho\|_1 \le C_{\phi,B},$$

which, together with (7.5), gives

$$\|Du_1\|_2 \le C(1 + \|\rho\|_2). \tag{7.14}$$

We next recall the following Sobolev inequalities in 2 space dimensions.

Lemma 7.3. *Let $q \in [2, \infty)$, $q' \in [1, 2]$. There exists constants $C_q, C_{q'} > 0$ such that*

$$\|f\|_q \le C_q \|f\|_{W^{1,2}(\mathbb{R}^2)}, \tag{7.15}$$

$$\|f\|_{q'} \le C_{q'} \|f\|_{W^{1,1}(\mathbb{R}^2)}. \tag{7.16}$$

There exists a constant $a > 0$ such that

$$\int_{\mathbb{R}^2} (e^{a \frac{|f(x)|^2}{\|f\|_2^2 + \|\nabla f\|_2^2}} - 1)dx \le 1. \tag{7.17}$$

Proof. The first two are standard Sobolev inequalities (see for instance Cases B and C of Theorem 4.12 of Adams and Fournier [1]), the last one is a version of Trudinger-Moser's inequality (e.g. Theorem 8.27 and Section 8.29 in Adams and Fournier [1]). □

Lemma 7.4. *Let $M_2(\rho)$ and $s(\rho)$ be finite. Then $e(\rho)$ is well defined and*

$$|e(\rho)| \le C_1 + C_2(M_2(\rho) + s(\rho)). \tag{7.18}$$

Moreover if $N_n := G_n * N$, where $G_n = J_n * J_n$ and the J_n are defined at the beginning of section 2, and $e_n(\rho) := 1/2\langle N_n * \rho, \rho\rangle$, then

$$\lim_{n\to\infty} e_n(\rho) = e(\rho). \tag{7.19}$$

(Convergence (7.19) is also true if the G_n are replaced by any family of standard mollifiers with compact support.)

Proof. To prove (7.18) we will show a stronger statement that if $\rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R}^2)$ then

$$\int_{\mathbb{R}^2} \rho_1(x) \int_{\mathbb{R}^2} |\log|x - y||\rho_2(y)dydx \le C_1 + C_2 \sum_{i=1}^{2}(M_2(\rho_i) + s(\rho_i))). \tag{7.20}$$

Let $\phi \in C_c^\infty(\mathbb{R}^2)$ be radial, $0 \le \phi \le 1$, $\phi(x) = 1$ for $|x| \le 1/2$ and $\phi(x) = 0$ for $|x| \ge 1$. Denote $N_1(x) = \phi(|x|)|\log|x||$, $N_2(x) = (1 - \phi(|x|))|\log|x||$. Then

$$\begin{aligned}\int_{\mathbb{R}^2} \rho_1(x) \int_{\mathbb{R}^2} N_2(x - y)\rho_2(y)dydx &\le C \int_{\mathbb{R}^2} \rho_1(x) \int_{\mathbb{R}^2} (|x| + |y|)\rho_2(y)dydx \\ &\le C \int_{\mathbb{R}^2} |x|\rho_1(x)dx + C \int_{\mathbb{R}^2} |y|\rho_2(y)dy \le C(1 + \sum_{i=1}^{2} M_2(\rho_i)).\end{aligned} \tag{7.21}$$

We now notice that if $r, s \ge 0$ then

$$rs \le e^r + (s \log s - s).$$

Therefore

$$N_1(x-y)\rho_2(y) \le e^{N_1(x-y)} + (\rho_2(y)\log\rho_2(y) - \rho_2(y))$$

and thus

$$\begin{aligned}
&\int_{\mathbb{R}^2} \rho_1(x) \int_{\mathbb{R}^2} N_1(x-y)\rho_2(y)dydx \le \|\rho_1\|_1 \sup_{x\in\mathbb{R}^2} \int_{\mathbb{R}^2} N_1(x-y)\rho_2(y)dy \\
&\quad\le \sup_{x\in\mathbb{R}^2} \left(\int_{B_1(x)} e^{N_1(x-y)}dy + \int_{B_1(x)} (\rho_2(y)\log\rho_2(y) - \rho_2(y))dy \right) \\
&\quad\le \int_{B_1} \frac{1}{|y|}dy + \int_{\{\rho_2(y)>1\}} \rho_2(y)\log\rho_2(y)dy \\
&\qquad\le C + s(\rho_2) + M_2(\rho_2),
\end{aligned} \tag{7.22}$$

where the last inequality follows from (1.16). This gives (7.18).

To show (7.19) we now define $N_{1,\epsilon}(x) = -1/(2\pi)\phi(x/\epsilon)\log|x|, N_{2,\epsilon}(x) = -1/(2\pi)(1-\phi(x/\epsilon))\log|x|$. Then

$$e_n(\rho) = \int\int (N_{1,\epsilon} * G_n)(x-y)\rho(y)\rho(x)dydx + \int\int (N_{2,\epsilon} * G_n)(x-y)\rho(y)\rho(x)dydx.$$

Since $N_{2,\epsilon} * G_n$ converges pointwise to $N_{2,\epsilon}$ as $n \to \infty$ and $|N_{2,\epsilon} * G_n(z)| \le C_\epsilon(1+|z|)$ on $\mathbb{R}^2$, by Lebesgee dominated convergence theorem we get

$$\lim_{n\to\infty} \int\int (N_{2,\epsilon} * G_n)(x-y)\rho(y)\rho(x)dydx = \int\int N_{2,\epsilon}(x-y)\rho(y)\rho(x)dydx. \tag{7.23}$$

Now arguing as in the proof of (7.20), we obtain for n large enough so that $\delta_n < \epsilon$,

$$\begin{aligned}
&\left|\int\int (N_{1,\epsilon} * G_n)(x-y)\rho(y)\rho(x)dydx\right| = \left|\int\int N_{1,\epsilon}(y)(G_n * \rho)(x-y)\rho(x)dydx\right| \\
&\le \frac{\|\rho\|_1}{2\pi} \sup_{x\in\mathbb{R}^2} \int_{B_\epsilon} |N(y)|(G_n * \rho)(x-y)dy \\
&\le \frac{1}{2\pi} \sup_{x\in\mathbb{R}^2} \left(\int_{B_\epsilon} \frac{1}{|y|}dy + \int_{B_\epsilon(x)} (G_n * \rho)(z)\log(G_n * \rho)(z)dz \right) \\
&\le \epsilon + \frac{1}{2\pi} \sup_{x\in\mathbb{R}^2} \int_{B_\epsilon(x)} \int_{\mathbb{R}^2} G_n(z-w)(\rho(w)\log\rho(w))_+ dwdz \\
&\quad\le \epsilon + \frac{1}{2\pi} \sup_{x\in\mathbb{R}^2} \int_{B_{3\epsilon}(x)} (\rho(w)\log\rho(w))_+ dw := \omega(\epsilon),
\end{aligned} \tag{7.24}$$

where $\omega(\epsilon) \to 0$ as $\epsilon \to 0$. Above we used Jensen's inequality to obtain that $(G_n * \rho)(z)\log(G_n * \rho)(z) \le G_n * (\rho\log\rho)(z)$. Since by the same argument we also have

$$\left|\int\int N_{1,\epsilon}(x-y)\rho(y)\rho(x)dydx\right| \le \omega(\epsilon), \tag{7.25}$$

(7.19) follows from (7.23), (7.24) and (7.25) if we let $\epsilon \to 0$.

□

Lemma 7.5. *There exists a constant $C > 0$ such that*

$$\int_{\mathbb{R}^2} |u|^2 d\rho \le C\|\rho\|_2^2 \le C(1 + I(\rho)) \tag{7.26}$$

and

$$\int_{\mathbb{R}^2} |u| d\rho \leq C\|\rho\|_2 \tag{7.27}$$

Proof. By (7.2)

$$\int_{\mathbb{R}^2} |u|^2 \rho dx \leq \|u\|_4^2 \|\rho\|_2 \leq C\|\rho\|_2^2$$

and the conclusion now follows since

$$\|\rho\|_2 \leq C(\|\rho\|_1 + \|\nabla\rho\|_1) \leq C(1 + \left\|\frac{\nabla\rho}{\sqrt{\rho}}\right\|_2 \|\sqrt{\rho}\|_2) = C(1 + \sqrt{I(\rho)}). \tag{7.28}$$

Now

$$\int_{\mathbb{R}^2} |u| d\rho \leq \left(\int_{\mathbb{R}^2} |u|^2 d\rho\right)^{\frac{1}{2}} \|\rho\|_1^{\frac{1}{2}} \leq C\|\rho\|_2.$$

□

Lemma 7.6. *Suppose $I(\rho) < \infty$. Then*

$$\int_{\mathbb{R}^2} |u(x)||\nabla\rho|(x) dx < \infty$$

and

$$\int_{\mathbb{R}^2} u(x) \cdot \nabla\rho(x) dx = 0.$$

Proof. By Lemma 7.5,

$$\int_{\mathbb{R}^2} |u||\nabla\rho| dx \leq \sqrt{\int_{\mathbb{R}^2} |u|^2 d\rho} \sqrt{\int_{\mathbb{R}^2} \frac{|\nabla\rho|^2}{\rho} dx} \leq C(1 + I(\rho)).$$

Let J_δ be a standard mollifier with compact support. Set $u_\delta := J_\delta * u = -K^\perp * (J_\delta * \rho)$. Then $u_\delta \in C^\infty(\mathbb{R}^2)$ and, by Hölder's inequality, (7.2) and $\|J_\delta * \rho\|_2 \leq \|\rho\|_2$,

$$\int_{\mathbb{R}^2} |u_\delta||\nabla\rho| dx \leq \|u_\delta\|_4 \|\rho\|_2^{1/2} (I(\rho))^{1/2} \leq C\|\rho\|_2 (I(\rho))^{1/2} \leq C(1 + I(\rho)),$$

and similarly

$$\int_{\mathbb{R}^2} |u_\delta| \rho dx \leq C\|\rho\|_2 \leq C(1 + I(\rho))^{1/2}$$

Also $\mathrm{div} u_\delta = J_\delta * \mathrm{div} u = 0$. Moreover,

$$\lim_{\delta\to 0} \int_{\mathbb{R}^2} u_\delta \cdot \nabla\rho dx = \int_{\mathbb{R}^2} u \cdot \nabla\rho dx,$$

since

$$\int_{\mathbb{R}^2} |u_\delta - u||\nabla\rho| dx \leq \|u_\delta - u\|_4 \|\rho\|_2^{1/2} (I(\rho))^{1/2} \to 0 \text{ as } \delta \to 0.$$

Now, using the functions ϕ_B from the proof of Lemma 7.1, we have,

$$\left|\int_{\mathbb{R}^2} u_\delta \cdot \nabla\rho dx\right| = \left|\lim_{B\to+\infty} \int_{\mathbb{R}^2} \phi_B u_\delta \cdot \nabla\rho dx\right|$$

$$= \left|\lim_{B\to+\infty} \int_{\mathbb{R}^2} \left[\nabla\phi_B \cdot u_\delta + \phi_B \mathrm{div} u_\delta\right] \rho dx\right| \leq \lim_{B\to+\infty} \frac{C}{B} \int_{\mathbb{R}^2} |u_\delta| \rho dx \to 0 \text{ as } \delta \to 0.$$

∎

Lemma 7.7. *Assume that $I(\rho) < \infty$, then*

$$\int_{\mathbb{R}^2} x \cdot u_\rho(x)\rho(dx) = 0.$$

Proof. The integral is well defined since by Lemma 7.5, $\int_{\mathbb{R}^2} |u|^2\rho(dx) < \infty$. The conclusion now follows because

$$\begin{aligned}
\int_{\mathbb{R}^2} x \cdot u(x)\rho(dx) &= -\int_{\mathbb{R}^2} x \cdot (K^\perp * \rho)(x)\rho(dx) \\
&= -\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} ((x-y)+y) \cdot K^\perp(x-y)\rho(dx)\rho(dy) \\
&= -\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} y \cdot K^\perp(x-y)\rho(dx)\rho(dy) = -\int_{\mathbb{R}^2} y \cdot u(y)\rho(dy).
\end{aligned}$$

∎

7.2. **Appendix B: A useful inequality regarding product of functions.** In this section several technical estimates are proved which are key in proving Lemma 7.16, which is essential in establishing (3.7).

The Fourier transform, denoted by $\mathcal{F}$ (acting on Schwartz functions) is defined via

$$\hat{f}(\xi) := \int_{\mathbb{R}^d} f(x)e^{-2\pi i x\cdot\xi}dx,$$

whereas the inverse transform $\mathcal{F}^{-1}$ is given by

$$f(x) = \int_{\mathbb{R}^d} \hat{f}(\xi)e^{2\pi i x\cdot\xi}d\xi,$$

Let us build the following partition of unity. Take first an even Schwartz function $\psi(x) \in C_c^\infty(\mathbb{R}^1)$, so that

$$\psi(\xi) = \begin{cases} 1 & |\xi| \le 1 \\ 0 & |\xi| > 2. \end{cases}$$

Introduce a function $\chi : \chi(\xi) = \psi(\xi) - \psi(2\xi)$. Clearly, for all $\xi \neq 0$, we have

$$\sum_{k=-\infty}^{\infty} \chi(2^{-k}\xi) = 1.$$

We define the Littlewood-Paley operators $P_k, P_{\le k}, P_{\sim k}$ via the formulas

$$\begin{aligned}
&\widehat{P_k f}(\xi) = \chi(2^{-k}\xi)\hat{f}(\xi) \\
&\widehat{P_{\le k} f}(\xi) = \psi(2^{-k}\xi)\hat{f}(\xi) \\
&P_{\sim k} := P_{k-2} + \ldots + P_{k+2}.
\end{aligned}$$

Occasionally, we denote for simplicity $f_k := P_k f$, $f_{\le k} := P_{\le k} f$. We collect the properties of the Littlewood-Paley operators in the following

Lemma 7.8. *We have the integral representation formulas for* $P_k, P_{\leq k}$

$$P_k f(x) = 2^{kd} \int_{\mathbb{R}^d} \hat{\chi}(2^k(x-y)) f(y) dy$$
$$P_{\leq k} f(x) = 2^{kd} \int_{\mathbb{R}^d} \hat{\psi}(2^k(x-y)) f(y) dy.$$

Moreover, we have the following decomposition formula

$$P_k[fg] = \sum_{l \geq k+3} P_k[f_{\sim l} g_l] + P_k[f_{\sim k} g_{\leq k+3}] + P_k[f_{\leq k+3} g_{\sim k}]. \tag{7.29}$$

The former sum corresponds to what is commonly called high-high interaction term, while the latter two sums represent are the high-low interaction term and low-high interaction term respectively.

Proof. The kernel representation formulas for P_k and $P_{\leq k}$ are due to the following simple properties of the (inverse) Fourier transform

$$P_k f = \mathcal{F}^{-1}(\chi(2^{-k}\cdot)\hat{f}(\cdot)) = \mathcal{F}^{-1}[\chi(2^{-k}\cdot)] * f = 2^{kd}\hat{\chi}(2^k\cdot) * f$$

and similar for $P_{\leq k} f$.

Regarding (7.29), we first make the simple observation
$supp(\widehat{fg}) = supp[\hat{f} * \hat{g}] \subseteq supp\ \hat{f} + supp\ \hat{g}$. Write

$$P_k[fg] = P_k[fg_{\sim k}] + P_k[f_{\sim k} g] + P_k[(f_{\leq k-3} + f_{\geq k+3})(g_{\leq k-3} + g_{\geq k+3})]$$

In addition, there is the observation that $P_k[f_{>k+3} g_{\sim k}] = 0$ by the observation above and hence

$$P_k[fg_{\sim k}] = P_k[f_{\leq k+3} g_{\sim k}]$$
$$P_k[f_{\sim k} g] = P_k[f_{\sim k} g_{\leq k+3}]$$

Also $P_k[f_{\leq k-3} g_{\leq k-3}] = 0 = P_k[f_{\leq k-3} g_{\geq k+3})]$. Finally,

$$P_k[f_{\geq k+3} g_{\geq k+3}] = \sum_{l \geq k+3} P_k[f_{\geq k+3} g_l] = \sum_{l \geq k+3} P_k[f_{\sim l} g_l].$$

where the first identity is by definition and the second identity follows due to
$P_k[f_{\leq l-3} g_l] = 0 = P_k[f_{\geq l+3} g_l]$, whenever $l \geq k+3$.

Putting all terms together yields (7.29). □

Note that, since χ, ψ are even functions, $\hat{\chi}, \hat{\psi}$ are real-valued functions.

Definition 7.9. Let $\zeta : \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing, convex function, $\zeta(0) = 0, \lim_{r\to\infty} \zeta(r) = +\infty$. Then, the Orlicz function space X_ζ is a Banach space with a norm given by

$$\|u\|_{X_\zeta} := \inf\{\lambda > 0 : \int_{\mathbb{R}^d} \zeta(\frac{|u(x)|}{\lambda}) dx \leq 1\}.$$

Let ζ^* be the Legendre transformation of ζ. Then X_{ζ^*} is an Orlicz space too. We would like to mention the following property (e.g. Proposition 1 on page 58 of Rao and Ren [27])

$$\sup_{x \in R^d} |\int u(x-y) v(y) dy| \leq C \|u\|_{X_{\zeta^*}} \|v\|_{X_\zeta}. \tag{7.30}$$

Note that the version in [27] requires $\zeta(1)+\zeta^*(1) \leq 1$. We do not require this here, hence we have a constant C on the right hand side of the inequality. In the next lemma, we compute the dual Orlicz function to $\varphi(r) = (r^2+1)\log(r^2+1)$, which gives rise to the Orlicz space $L^2Log(L)(\mathbb{R}^n)$, which is defined as X_φ, with $\varphi(r) = (r^2+1)\log(r^2+1)$. Note that according to the definition $\|\cdot\|_{L^2} \leq C\|\cdot\|_{L^2Log(L)}$.

Lemma 7.10. *Let $F : \mathbb{R} \mapsto \mathbb{R}, h : \mathbb{R}^d \mapsto \mathbb{R}$ be two convex functions with respective Legendre-Fenchel transform F^* and h^*. Suppose that F is non-decreasing, $\lim_{t\to\infty} F(t) = \infty$ and that h is lower semicontinuous. Then*

$$(F \circ h)^*(y) = \inf\{F^*(\alpha) + \alpha h^*(\beta) : \alpha \in \mathbb{R}, \beta \in \mathbb{R}^d; \alpha\beta = y\}.$$

Proof. This is Theorem D.3.5 in Dupuis and Ellis [8]. □

Lemma 7.11.

$$\varphi^*(s) \leq \frac{2s^2}{\max(1, \log(s))}$$

Proof. Take

$$F(p) = (p+1)\log(p+1),$$

then

$$F^*(\alpha) = e^{\alpha-1} - \alpha.$$

Let $h(x) = |x|^2/2$, then $h^*(\beta) = \beta^2/2$. Consequently

$$\begin{aligned}
(\varphi)^*(s) &= (F \circ h)^*(s) \\
&= \inf\{e^{\alpha-1} - \alpha + \alpha\frac{\beta^2}{2} : \alpha\beta = s, \alpha, \beta \in \mathbb{R}\} \\
&= \inf_{\alpha\in\mathbb{R}}\{e^{\alpha-1} - \alpha + \frac{s^2}{2\alpha}\} \\
&= e^{\alpha(s)-1} - \alpha(s) + \frac{s^2}{2\alpha(s)}
\end{aligned}$$

where $\alpha(s)$ satisfies

$$e^{\alpha(s)-1} - 1 = \frac{s^2}{2\alpha^2(s)}, \quad \text{implying } \alpha(s) > 1,$$

For $\alpha = \alpha(s)$,

$$e^{2\alpha} > 2\alpha^2(e^{\alpha-1} - 1) = s^2 \Rightarrow s^2 < e^{2\alpha} \Rightarrow \log(s) \leq \alpha,$$

$$e^{\alpha-1} - \alpha + \frac{s^2}{2\alpha} = \frac{s^2}{\alpha} + \frac{s^2}{\alpha^2} + 1 - \alpha \leq \frac{2s^2}{\max(1, \log(s))}.$$

□

Proposition 7.12. *Let Q_k be either P_k or $P_{\leq k}$. Then*

$$\|Q_k f\|_{L^\infty(\mathbb{R}^d)} \leq C\frac{2^{dk/2}}{\sqrt{k}}\|f\|_{L^2Log(L)(\mathbb{R}^d)}. \tag{7.31}$$

There is the bilinear estimate

$$\|P_{>j}[uv]\|_{L^2(\mathbb{R}^2)} \leq \frac{C}{j^{1/4}}(\|\nabla u\|_{L^2(\mathbb{R}^2)} + \|\nabla v\|_{L^2(\mathbb{R}^2)})(\|u\|_{L^2Log(L)(\mathbb{R}^2)} + \|v\|_{L^2Log(L)(\mathbb{R}^2)}) \tag{7.32}$$

Proof. We start with the proof of (7.31), after which we will be able to deduce (7.32) as a consequence of it and (7.29). We will give a proof of this only for P_k, the other one being similar. We have

$$\|P_k f\|_{L^\infty} = \sup_x 2^{kd} | \int_{\mathbb{R}^d} \hat{\chi}(2^k(x-y))f(y)dy| \leq C2^{kd}\|\hat{\chi}(2^k\cdot)\|_{X_{\varphi^*}}\|f\|_{L^2Log(L)(\mathbb{R}^d)}.$$

It remains to show $\mu = \|\hat{\varphi}(2^k\cdot)\|_{X_{\varphi^*}} \leq C2^{-dk/2}k^{-1/2}$. From the defining equation of $\|\cdot\|_{X_{\varphi^*}}$,

$$\int_{\mathbb{R}^d} \varphi^*(\hat{\chi}(2^k y)/\mu)dy = 1.$$

We have by Lemma 7.8

$$\frac{2}{\mu^2}\int_{\mathbb{R}^d} \frac{|\hat{\chi}(2^k y)|^2}{\max(1, \ln(|\hat{\chi}(2^k y)|/\mu))}dy \geq 1.$$

A change of variables $2^k y = z$ yields the inequality

$$\int_{\mathbb{R}^d} \frac{|\hat{\chi}(z)|^2}{\max(1, \ln(|\hat{\chi}(z)|/\mu))}dz \geq C2^{kd}\mu^2 \tag{7.33}$$

From (7.33), since the left-hand side is bounded by $\int_{\mathbb{R}^d} |\hat{\chi}(z)|^2 \leq C$, we immediately get $\mu \leq C2^{-kd/2}$. If we feed this estimate back in the left-hand side of (7.33), we get

$$C2^{kd}\mu^2 \leq C\int_{z:|\hat{\chi}(z)|\leq 2^{-kd/4}} |\hat{\chi}(z)|^2 + \frac{C}{k}\int_{\mathbb{R}^d} |\hat{\chi}(z)|^2 dz \leq C2^{-kd/4} + \frac{C}{k}$$

It follows that $\mu \leq C2^{-kd/2}k^{-1/2}$.

For the proof of (7.32), we write first

$$P_{>j}[uv] = \sum_{k:k>j} P_k[uv]$$

Next, for each term $P_k[uv]$, we employ the decomposition (7.29). We have

$$P_k(uv) = \sum_{l\geq k+3} P_k[u_{\sim l}v_l] + P_k[u_{\sim k}v_{\leq k+3}] + P_k[u_{\leq k+3}v_{\sim k}]$$

Since the second and third sum are symmetric in u, v, we consider only one of them, say $P_k[u_{\sim k}v_{\leq k+3}]$. We have by Hölder's inequality

$$\|P_k[u_{\sim k}v_{\leq k+3}]\|_{L^2(\mathbb{R}^2)} \leq C\|u_{\sim k}\|_{L^2(\mathbb{R}^2)}\|v_{\leq k+3}\|_{L^\infty(\mathbb{R}^2)}$$

By (7.31), we can estimate $\|v_{\leq k+3}\|_{L^\infty(\mathbb{R}^2)} \leq C2^k k^{-1/2}\|v\|_{L^2Log(L)(\mathbb{R}^2)}$, whence we obtain the estimate

$$\begin{aligned}\|P_k[u_{\sim k}v_{\leq k+3}]\|_{L^2(\mathbb{R}^2)} &\leq C2^k k^{-1/2}\|u_{\sim k}\|_{L^2(\mathbb{R}^2)}\|v\|_{L^2Log(L)(\mathbb{R}^2)} \leq \\ &\leq Ck^{-1/2}\|\nabla u_{\sim k}\|_{L^2(\mathbb{R}^2)}\|v\|_{L^2Log(L)(\mathbb{R}^2)}.\end{aligned}$$

Consequently, we have

$$\begin{aligned}&\|\sum_{k:k>j} P_k[u_{\sim k}v_{\leq k+3}]\|_{L^2}^2 \leq \sum_{k:k>j}\|P_k[u_{\sim k}v_{\leq k+3}]\|_{L^2}^2 \leq \\ &\leq C\|v\|_{L^2Log(L)(\mathbb{R}^2)}^2 \sum_{k:k>j}\frac{1}{k}\|\nabla u_{\sim k}\|_{L^2(\mathbb{R}^2)}^2 \leq \frac{C}{j}\|v\|_{L^2Log(L)(\mathbb{R}^2)}^2\|\nabla u\|_{L^2(\mathbb{R}^2)}^2,\end{aligned}$$

which shows (7.32) for the high-low terms.

For the high-high interaction term, we estimate by Sobolev embedding and Hölder's

$$\|P_k[u_{\sim l}v_l\|_{L^2(\mathbb{R}^2)} \leq C2^{k/2}\|u_{\sim l}v_l\|_{L^{4/3}(\mathbb{R}^2)} \leq$$
$$\leq C2^{k/2}\|u_{\sim l}\|_{L^2(\mathbb{R}^2)}\|v_l\|_{L^4(\mathbb{R}^2)}.$$

We use the Gagliardo-Nirenberg's inequality and (7.31) to estimate

$$\|v_l\|_{L^4(\mathbb{R}^2)} \leq \|v_l\|^{1/2}_{L^\infty(\mathbb{R}^2)}\|v_l\|^{1/2}_{L^2(\mathbb{R}^2)} \leq$$
$$\leq C2^{l/2}l^{-1/4}\|v\|^{1/2}_{L^2Log(L)(\mathbb{R}^2)}\|v\|^{1/2}_{L^2(\mathbb{R}^2)}$$

Putting all these estimates together, we obtain

$$\|\sum_{k>j}\sum_{l\geq k+3} P_k(u_{\sim l}v_l)\|^2_{L^2(\mathbb{R}^2)} \leq C\sum_{k>j}\|\sum_{l\geq k+3} P_k(u_{\sim l}v_l)\|^2_{L^2(\mathbb{R}^2)} \leq$$
$$\leq C\|v\|_{L^2Log(L)(\mathbb{R}^2)}\|v\|_{L^2(\mathbb{R}^2)}\sum_{k>j}\left(\sum_{l\geq k+3}\frac{2^{l/2+k/2}}{l^{1/4}}\|u_{\sim l}\|_{L^2(\mathbb{R}^2)}\right)^2 \leq$$
$$\leq \frac{C}{\sqrt{j}}\|v\|_{L^2Log(L)(\mathbb{R}^2)}\|v\|_{L^2(\mathbb{R}^2)}\sum_{k>j}\sum_{l_1,l_2\geq k+3} 2^{l_1/2+l_2/2+k}\|u_{\sim l_1}\|_{L^2(\mathbb{R}^2)}\|u_{\sim l_2}\|_{L^2(\mathbb{R}^2)}.$$

Now,

$$\sum_{k>j}\sum_{l_1,l_2\geq k+3} 2^{l_1/2+l_2/2+k}\|u_{\sim l_1}\|_{L^2(\mathbb{R}^2)}\|u_{\sim l_2}\|_{L^2(\mathbb{R}^2)} \leq$$
$$\leq C\sum_{l_1,l_2} 2^{l_1/2+l_2/2+\min(l_1,l_2)}\|u_{\sim l_1}\|_{L^2(\mathbb{R}^2)}\|u_{\sim l_2}\|_{L^2(\mathbb{R}^2)}.$$

By symmetry, it will suffice to estimate the sum above with $l_1 \leq l_2$. We have by Cauchy-Schwartz

$$\sum_{l_1,l_2:l_1\leq l_2} 2^{(3l_1/2+l_2/2)}\|u_{\sim l_1}\|_{L^2(R^2)}\|u_{\sim l_2}\|_{L^2(\mathbb{R}^2)} \leq$$
$$\leq (\sum_{l_1,l_2:l_1\leq l_2} 2^{2l_1}\|u_{\sim l_1}\|^2_{L^2(\mathbb{R}^2)}2^{-(l_2-l_1)})^{1/2} \times (\sum_{l_1,l_2:l_1\leq l_2} 2^{2l_2}\|u_{\sim l_2}\|^2_{L^2(R^2)}2^{-(l_2-l_1)})^{1/2} \leq C\|\nabla u\|^2_{L^2(\mathbb{R}^2)}.$$

Thus,

$$\|\sum_{k>j}\sum_{l\geq k+3} P_k(u_{\sim l}v_l)\|^2_{L^2(\mathbb{R}^2)} \leq \frac{C}{\sqrt{j}}\|v\|_{L^2Log(L)(\mathbb{R}^2)}\|v\|_{L^2(\mathbb{R}^2)}\|\nabla u\|^2_{L^2(\mathbb{R}^2)}.$$

Recalling $\|\cdot\|_{L^2} \leq C\|\cdot\|_{L^2Log(L)}$, we conclude (7.32). □

The following estimate will be useful for approximating velocity field u.

Lemma 7.13. *Let $\varphi \in C^\infty_c$ and radial, so that supp$\varphi \subset \{x : |x| < 1\}$. Define*

$$g(x) = \nabla[\varphi(x)\frac{x_1}{|x|^2}]$$

Then

$$\|\hat{g}\|_{L^\infty} \leq C.$$

Proof. We have

$$\hat{g}(\xi) = -2\pi i \xi_1 \int_{\mathbb{R}^2} \varphi(x) \frac{x_1}{|x|^2} e^{-2\pi i x\cdot\xi} dx.$$

If $|\xi_1| < 100$, we have

$$|\hat{g}(\xi)| \le 200\pi \int_{\mathbb{R}^2} |\varphi(x)| \frac{1}{|x|} dx \le 200\pi \times 2\pi \int_0^1 |\varphi(\rho)| d\rho < \infty.$$

Let $|\xi_1| \sim 2^l, l >> 1$. Then, since $\psi(2^l x) + \sum_{k=-5}^{l-1} \chi(2^k x) = 1$ on the support of φ, we have

$$|\hat{g}(\xi)| \le C 2^l \Big(\int |\varphi(x)| \psi(2^l x) \frac{1}{|x|} dx + \sum_{k=-5}^{l-1} |\int \varphi(x) \chi(2^k x) \frac{x_1}{|x|^2} e^{-2\pi i x\cdot\xi} dx| \Big)$$

The first integral is estimated easily

$$\int |\varphi(x)| \psi(2^l x) \frac{1}{|x|} dx \le \int_0^{2^{-l+1}} \frac{1}{\rho} \rho dx \le C 2^{-l}.$$

For the second integral, integrate by parts twice in the variable x_1. We have

$$\int \varphi(x) \chi(2^k x) \frac{x_1}{|x|^2} e^{-2\pi i x\cdot\xi} dx = \frac{-1}{4\pi^2 \xi_1^2} \int \frac{d^2}{dx_1^2} [\varphi(x) \chi(2^k x) \frac{x_1}{|x|^2}] e^{-2\pi i x\cdot\xi} dx$$

Observe that

$$|\frac{-1}{4\pi^2 \xi_1^2}| \le C 2^{-2l},$$

while (recall $\text{supp}(\chi) \subset \{1/2 < |x| < 2\}$)

$$|\frac{d^2}{dx_1^2} [\varphi(x) \chi(2^k x) \frac{x_1}{|x|^2}]| \le C 2^{3k} (|\chi(2^k x)| + |\chi'(2^k x)| + |\chi''(2^k x)|).$$

Thus, we have

$$|\frac{-1}{4\pi^2 \xi_1^2} \int \varphi(x) \chi(2^k x) \frac{x_1}{|x|^2} e^{-2\pi i x\cdot\xi} dx| \le C 2^{-2l} 2^{3k} \int_{2^{-k-1}}^{2^{-k+1}} \rho d\rho = C 2^{k-2l}.$$

Thus,

$$|\hat{g}(\xi)| \le C 2^l (2^{-l} + \sum_{k=-5}^{l-1} 2^{k-2l}) \le C.$$

□

7.3. **Appendix C: Approximation of the velocity field u.** First, we have the following estimate.

Lemma 7.14. *There exist constants $C_1, C_2 > 0$ such that*

$$\|P_{>k}\rho\|_2^2 \le \frac{C_1}{k^{1/2}} I(\rho) \Big(C_2 + S(\rho) + M_2(\rho) \Big)^2.$$

Proof. Without loss of generality, we assume that the right hand side is finite. We take $u = v = \sqrt{\rho}$ in (7.32) and notice that, for $\varphi(r) = (1 + r^2) \log(1 + r^2)$,

$$\|\sqrt{\rho}\|_{L^2 Log(L)} \le 1 + \int_{\mathbb{R}^2} \varphi(\sqrt{\rho}) dx \le 3 + \int ((\rho \log \rho) \vee 0) dx \le 3 + \int_{\rho(x) \ge 1} \rho \log \rho dx,$$

where the first inequality follows from properties of Orlicz norm (e.g. Theorem 13 on page 69 of Rao and Ren [27]), and the second inequality follows from

$$\varphi(r) \le r^2 + r^2\log(1+r^2) \le r^2(1+2\log 2) + (r^2\log r^2)\vee 0. \tag{7.34}$$

Noting that $r|\log r| \le \sqrt{r}$ for $0 \le r \le 1$, we get

$$\int_{0\le\rho(x)<1} \rho|\log\rho|dx \le \int_{0\le\rho(x)\le e^{-|x|}} + \int_{e^{-|x|}<\rho(x)<1} \le \int e^{-|x|/2}dx + \int |x|\rho(x)dx. \le C + M_2(\rho).$$

Thus we conclude that

$$\|\sqrt{\rho}\|_{L^2Log(L)} \le C + s(\rho) + M_2(\rho).$$

Since

$$\|\nabla\sqrt{\rho}\|_2^2 \le I(\rho),$$

the lemma now follows from (7.32). □

Recall that $u = u_{1,B} + u_{2,B}$ in (7.9) for each $B > 0$. We approximate $u_1 = u_{1,B}$ by

$$u_{\le k,1} = P_{\le k}u_1 = P_{\le k}\Big(\,(-\phi_B K^\perp) * \rho\Big) = J_{\le k,1} * \rho$$

where

$$J_{\le k,1}(x) = \mathcal{F}^{-1}\Big((\widehat{-\phi_B K^\perp})\psi(2^{-2k}\cdot)\Big)(x) = \int_{\xi\in\mathbb{R}^2} (\widehat{-\phi_B K^\perp})(\xi)\psi(2^{-2k}\xi)e^{i\xi\cdot x}d\xi.$$

We now approximate u by

$$u_{\le k,1} + u_2 = \Big(J_{\le k,1} - (1-\phi_B)K^\perp\Big) * \rho =: \tilde{J}_k * \rho.$$

Then

Lemma 7.15. *For each k, $B > 0$ fixed, $\tilde{J}_k = J_{\le k,1} - (1-\phi_B)K^\perp$ is Lipschitz continuous, bounded, and $\tilde{J}_k(-z) = -\tilde{J}_k(z)$ for all $z \in \mathbb{R}^2$.*

Proof. The kernel $(1-\phi_B(z))K^\perp(z)$ is Lipschitz. Therefore we only need to verify the same property for $J_{\le k,1}$:

$$\begin{aligned} |J_{\le k,1}(x) - J_{\le k,1}(y)| &\le \int_{\mathbb{R}^2} |(\widehat{-\phi_B K^\perp})(\xi)|\psi(2^{-2k}\xi)|\xi||x-y|d\xi \\ &\le \Big(\int_{|\xi|\le 2^{2k+1},\xi\in\mathbb{R}^2} |\xi||\widehat{\phi_B K}(\xi)|d\xi\Big)|x-y| \le C_{k,B}|x-y|. \end{aligned}$$

where the last inequality follows because for every

$$\|\widehat{\phi_B K}\|_\infty \le \int_{\mathbb{R}^2} |\phi_B(z)K(z)|dz < +\infty.$$

The boundedness and antisymmetry of $\tilde{J}_k$ are straightforward from its definition. □

We notice that

$$\begin{aligned} P_{>k}u_1 &= -P_{>k}(\phi_B K^\perp) * \rho = -(\phi_B K^\perp) * (P_{>k}\rho) = -J(\phi_B K) * (P_{>k}\rho) \\ \nabla P_{>k}u_1 &= -P_{>k}\nabla((\phi_B K^\perp) * \rho) \\ &= -P_{>k}\Big((\nabla(\phi_B K^\perp)) * \rho\Big) = -J(\nabla(\phi_B K)) * (P_{>k}\rho). \end{aligned}$$

Therefore

$$\|P_{>k}u_1\|_2 \le \|\phi_B K\|_1 \|P_{>k}\rho\|_2$$

In addition, by Lemma 7.13, since ϕ_B is radial,

$$\|\nabla\widehat{(\phi_B K)}\|_\infty < \infty$$

and

$$\|\nabla P_{>k}u_1\|_2 = \|(\nabla(\phi_B K)) * (P_{>k}\rho)\|_2 \le \|\nabla\widehat{(\phi_B K)}\|_\infty \|P_{>k}\rho\|_2 \le C\|P_{>k}\rho\|_2$$

In summary,

$$\|u_{>k,1}\|_2 + \|\nabla u_{>k,1}\|_2 \le C\|P_{>k}\rho\|_2. \tag{7.35}$$

Lemma 7.16. *For every $\delta > 0$ there exists a bounded, Lipschitz continuous and antisymmetric function $\hat{J}_\delta : \mathbb{R}^2 \mapsto \mathbb{R}$ such that for every $\rho \in \mathcal{P}_2(\mathbb{R}^2)$, the function $u_\delta = \hat{J}_\delta * \rho$ satisfies*

$$\int_{\mathbb{R}^2} |u(x) - u_\delta(x)|^2 \rho(x)dx \le \delta I(\rho)\Big(\tilde{C} + s(\rho) + M_2(\rho)\Big)^3 \tag{7.36}$$

for some absolute constant $\tilde{C}$.

Proof. We only need to prove the case when $I(\rho) < \infty$. We set $\hat{J}_\delta := \tilde{J}_k$, where B is fixed and k will be chosen later. Then

$$u_\delta = u_{\le k,1} + u_{2,B} = -J\Big(P_{\le k}(\phi_B K) + (1-\phi_B)K\Big) * \rho$$

and

$$u - u_\delta = u_1 - u_{\le k,1} = u_{>k,1}.$$

Let $a > 0$ be the constant in Moser's inequality (7.17). It then follows that

$$\begin{aligned}
&\int_{\mathbb{R}^2} |u - u_\delta|^2 \rho(x)dx = \int_{\mathbb{R}^2} |u_{>k,1}|^2 \rho(x)dx \\
= \;& a^{-1}(\|u_{>k,1}\|_2^2 + \|\nabla u_{>k,1}\|_2^2) \int_{\{x:\rho(x)\le 1\}\cup\{x:\rho(x)>1\}} \frac{a|u_{>k,1}(x)|^2}{\|u_{>k,1}\|_2^2 + \|\nabla u_{>k,1}\|_2^2}\rho(x)dx \\
\le \;& a^{-1}C_1\|P_{>k}\rho\|_2^2\Big(a + \int_{\{x:\rho(x)>1\}} \frac{a|u_{>k,1}(x)|^2}{\|u_{>k,1}\|_2^2 + \|\nabla u_{>k,1}\|_2^2}\rho(x)dx\Big) \\
\le \;& a^{-1}C_1\|P_{>k}\rho\|_2^2\Big(a + \int_{\mathbb{R}^2} (e^{\frac{a|u_{>k,1}(x)|^2}{\|u_{>k,1}\|_2^2 + \|\nabla u_{>k,1}\|_2^2}} - 1)dx + \int_{\{x:\rho(x)>1\}} \rho\log\rho dx\Big) \\
\le \;& C_2\|P_{>k}\rho\|_2^2(C_3 + s(\rho) + M_2(\rho)) \\
\le \;& \frac{C_3}{k^{1/2}} I(\rho)(\tilde{C} + s(\rho) + M_2(\rho))^3.
\end{aligned}$$

In the above, the first inequality follows from (7.35), the second from the Legendre transform identity

$$(e^x - 1)^* = \sup_y(xy - e^y + 1) = x\log x - x + 1,$$

the third follows from (7.17), the fourth from an elementary calculation as in the proof of Lemma 7.14, and the last from Lemma 7.14. It now remains to take k sufficiently big. $\square$

7.4. **Appendix D: The chain rule.** The following is an adaptation of Proposition 10.3.18 of Ambrosio, Gigli and Savaré [3] to special cases which cover our applications. We restrict our attention to λ-convex functionals on $\mathcal{P}_2(\mathbb{R}^d)$ defined in Definition 9.1.1 of [3]. This is a notion of convexity along geodesics in the mass transport sense.

Lemma 7.17. *Let $f : \mathcal{P}_2(\mathbb{R}^d) \mapsto (-\infty, +\infty]$ be lower semicontinuous and λ-convex. Assume that $D(f) \neq \emptyset$ and that there exists $\rho_0 \in D(f)$ and $r_0 > 0$ such that*

$$\inf\{f(\rho) : d(\rho, \rho_0) \leq r_0\} > -\infty.$$

Let $\rho(\cdot) \in AC((a,b); \mathcal{P}_2(\mathbb{R}^d))$ with $\rho(t) \in D(f)$, and moreover let

$$\partial_t \rho + \operatorname{div}(\rho v) = 0, \quad a < t < b,$$

hold for some v such that $v(r, \cdot) \in L^2_\nabla(\rho(t))$ and $\int_a^b \int_{\mathbb{R}^d} |v(r,x)|^2 \rho(r, dx) dr < \infty$.

Suppose that

$$\operatorname{grad} f(\rho(t)) = -\operatorname{div}(\rho(t) u(t)), \quad u(t, \cdot) \in L^2_\nabla(\rho(t))$$

for a.e. $t \in (a,b)$.

Then for every $(s,t) \subset (a,b)$,

$$f(\rho(t)) - f(\rho(s)) = \int_s^t \int_{\mathbb{R}^d} v(r,x) \cdot u(r,x) \rho(r, dx) dr.$$

Proof. All references (to theorems, lemmas, etc.) below refer to [3].

By Lemma 10.1.3, λ-convexity implies that f is regular in the sense of Definition 10.1.4. By Theorem 4.1.2, λ-convexity in this context also implies that (10.3.1b) is satisfied. The conclusion follows from Theorem 10.3.18. □

Lemma 7.18. *The functions $-d^2(\cdot, \gamma)$ defined in* (1.10) *(with $\gamma \in \mathcal{P}_2(\mathbb{R}^d)$ fixed), $s(\cdot)$ in* (1.15)*, $e_n(\cdot)$ in* (2.2) *and $M_p(\cdot)$ in* (1.13) *when $p \geq 1$, are all λ-convex, for some $\lambda \in \mathbb{R}$.*

Proof. Again, all the theorems, propositions, etc., in the proof refer to [3].

The case of $-d^2(\cdot, \gamma)$ follows from Theorem 7.3.2; the case of $s(\cdot)$ follows from Proposition 9.3.9 and Remark 9.3.10; the case of $M_p(\cdot)$ follows from Proposition 9.3.2. The function $e_n(\cdot)$ is a special case of Example 9.3.4 and Proposition 9.3.5. Note that $N_n(x) + C_n|x|^2$ is convex for some $C_n \geq 0$. See also examples in Chapters 16 and 17 of [29]. □

References

1. R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, Second edition, Pure and Applied Mathematics, Academic Press [An imprint of Elsevier Science], New York-London, 2003.
2. L. Ambrosio and W. Gangbo, Hamiltonian ODEs in the Wasserstein space of probability measures *Comm. Pure Appl. Math.* 61 (2008), no. 1, 18–53.
3. L. Ambrosio, N. Gigli and G. Savaré. *Gradient flows*, Lectures in Mathematics, ETH Zurich, 2005, Birkhäuser.
4. M. Bardi and I. Capuzzo-Dolcetta, *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*, with appendices by Maurizio Falcone and Pierpaolo Soravia, Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, 1997.
5. A. L. Bertozzi and T. Laurent, The behavior of solutions of multidimensional aggregation equations with mildly singular interaction kernels, *Chin. Ann. Math. Ser. B* 30 (2009), no. 5, 463–482.
6. J. M. Borwein and D. Preiss, A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions, *Trans. Amer. Math. Soc.* 303 (1987), no. 2, 517–527.
7. A. J. Chorin, *Vorticity and turbulence*, Applied Mathematical Sciences, 103, Springer-Verlag, New York, 1994, Corrected Second Printing, 1998.

8. P. Dupuis and R. Ellis, *A Weak convergence approach to the theory of large deviations*, Wiley and Sons, New York, 1997.
9. S. N. Ethier and T. G. Kurtz, *Markov Processes, characterization and convergence*, Wiley and Sons, New York, 1986.
10. J. Feng and M. Katsoulakis, A comparison principle for Hamilton-Jacobi equations related to controlled gradient flows in infinite dimensions. *Archive for Rational Mechanics and Analysis* 192 (2009), 275–310.
11. J. Feng and T. G. Kurtz, *Large deviation for stochastic processes*, Mathematical Surveys and Monographs, Vol 131, American Mathematical Society, 2006.
12. W. Gangbo, H. K. Kim and T. Pacini, Differential forms on Wasserstein space and infinite dimensional Hamiltonian systems, *Mem. Amer. Math. Soc.* 211 (2011), no. 993, 77 pp.
13. M. H. Giga, Y. Giga and J. Saal, *Nonlinear partial differential equations: Asymptotic behavior of solutions and self-similar solutions*, Progress in Nonlinear Differential Equations and Their Applications, Vol. 79, Birkhauser, 2010.
14. Y. Giga, T. Miyakawa and H. Osada, Two-dimensional Navier-Stokes flow with measures as initial vorticity, *Arch. Rational Mech. Anal.* 104 (1988), no. 3, 223–250.
15. L. Grafakos, *Classical and modern Fourier analysis*, Prentice Hall, 2003.
16. H. Ishii, Viscosity solutions for a class of Hamilton-Jacobi equations in Hilbert spaces, *J. Funct. Anal.* 105 (1992), no. 2, 301–341.
17. A. Jakubowski, On the Skorohod topology, *Ann. Inst. H. Poincaré Probab. Statist.* 22 (1986), 263–285.
18. R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* 29 (1998), no. 1, 1–17.
19. P. L. Lions, *On Euler equations and statistical physics*, Cattedra Galileiana, [Galileo Chair] Scuola Normale Superiore, Classe di Scienze, Pisa, 1998.
20. A. Majda and A. L. Bertozzi, *Vorticity and incompressible flow*, Cambridge Texts in Applied Mathematics, 27, Cambridge University Press, Cambridge, 2002.
21. C. Marchioro and M. Pulvirenti, Hydrodynamics in two dimensions and vortex theory, *Comm. Math. Phys.* 84 (1982), 483–503.
22. S. Méléard, A trajectorial proof of the vortex method for the two-dimensional Navier-Stokes equation, *Ann. Appl. Probab.* 10 (2000), no. 4, 1197–1211.
23. S. Méléard, Monte-Carlo approximations for 2d Navier-Stokes equations with measure initial data, *Probab. Theory Related Fields* 121 (2001), no. 3, 367–388.
24. S. Méléard, Stochastic particle approximations for two-dimensional Navier-Stokes equations, Dynamics and randomness II, 147–197, *Nonlinear Phenom. Complex Systems*, 10, Kluwer Acad. Publ., Dordrecht, 2004.
25. H. Osada, Propagation of chaos for the two-dimensional Navier-Stokes equation, Probabilistic methods in mathematical physics (Katata/Kyoto, 1985), 303–334, Academic Press, Boston, MA, 1987.
26. F. Otto, The geometry of dissipative evolution equations: the porous medium equation, *Comm. Partial Differential Equations* 26, 1-2 (2001), 101–174.
27. M. M. Rao and Z. D. Ren, *Theory of Orlicz spaces*, Monographs and Textbooks in Pure and Applied Mathematics, 146. Marcel Dekker, Inc., New York, 1991.
28. C. M. Topaz and A. L. Bertozzi, Swarming patterns in a two-dimensional kinematic model for biological groups, *SIAM J. Appl. Math.* 65 (2004), no. 1, 152–174.
29. C. Villani, *Optimal transport. Old and new*, Comprehensive Studies in Mathematics, Vol. 338, Springer-Verlag, New York, 2009.

Department of Mathematics, University of Kansas, Lawrence, KS 66045, USA
E-mail address: jfeng@math.ku.edu

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA
E-mail address: swiech@math.gatech.edu

Department of Mathematics, University of Kansas, Lawrence, KS 66045, USA
E-mail address: stefanov@math.ku.edu